Boehringer Ingelheim

Protocol for observational studies based on existing data

Document Number:	**c03270726-08**
BI Study Number:	1245.96
BI Investigational Product(s):	Jardiance® (empagliflozin) Synjardy® (empagliflozin/metformin)
Title:	Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated with empagliflozin compared to patients treated with DPP-4 inhibitors
Brief lay title:	Post-authorisation safety study in patients with type 2 diabetes to assess the risk of liver injury, kidney injury, urinary tract and genital infections, and diabetic ketoacidosis in patients treated with empagliflozin, compared to DPP-4 inhibitors
Protocol version identifier:	8.0
Date of last version of protocol:	13 Jun 2019 (version 7.0)
PASS:	Yes
EU PAS Register number:	ENCEPP/SDPP/13413; http://www.encepp.eu/encepp/viewResource.htm?id=13414
Active substance:	A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin
Medicinal product:	Jardiance Synjardy
Product reference:	EMEA/H/C/002677 EMEA/H/C/003770
Procedure number:	EMEA/H/C/002677/MEA EMEA/H/C/003770/MEA
Joint PASS:	No

Research question and objectives:	To estimate, among patients with type 2 diabetes mellitus (T2D), the risk of acute liver injury (ALI), the risk of acute kidney injury (AKI) and chronic kidney disease (CKD), the risk of severe complications of urinary tract infection (UTI), the risk of genital infections (GI), and the risk of diabetic ketoacidosis (DKA) among patients treated with empagliflozin compared with patients treated with dipeptidyl peptidase-4 (DPP-4) inhibitors
Country(-ies) of study:	United Kingdom, Denmark, United States
Authors:	[redacted]
Marketing authorisation holder(s):	Boehringer Ingelheim International GmbH Binger Straße 173 55216 Ingelheim am Rhein Germany
MAH contact person:	[redacted]
EU-QPPV:	[redacted]
Signature of EU-QPPV:	*The signature of the EU-QPPV is provided electronically*
Date:	19 July 2021 (version 8.0)

1. TABLE OF CONTENTS

2. LIST OF ABBREVIATIONS

ADQI	Acute Dialysis Quality Initiative
AKI	Acute Kidney Injury
ALI	Acute Liver Injury
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical
BI	Boehringer Ingelheim International GmbH
BMI	Body Mass Index
CI	Confidence Interval
CKD	Chronic Kidney Disease
CKD-EPI	CKD Epidemiology Collaboration
CPR	Central Personal Registration (number), Denmark
CPRD	Clinical Practice Research Datalink
CPRD Aurum	Database of De-Identified Coded Primary Care Records for Use in Public Health Research (of the CPRD)
CPRD GOLD	General Practitioner Online Database (of the CPRD)
CPT	Current Procedural Terminology (codes)
DKA	Diabetic Ketoacidosis
DPP-4	Dipeptidyl Peptidase-4
ED	Emergency Department
eGFR	Estimated Glomerular Filtration Rate
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
FDA	United States Food and Drug Administration
GI	Genital Infection
GLDs	Glucose-Lowering Drugs
GLP-1	Glucagon-Like Peptide-1
GP	General Practitioner

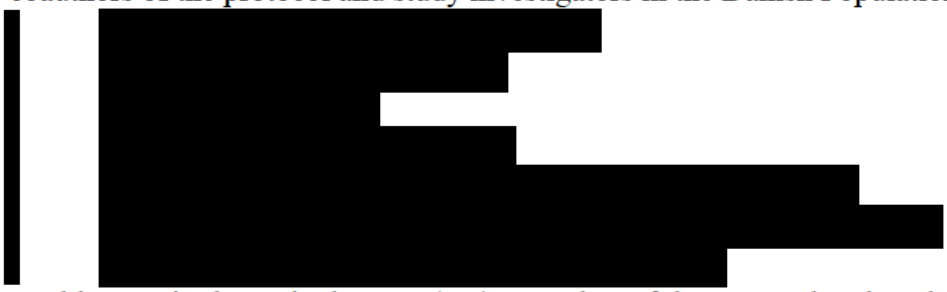
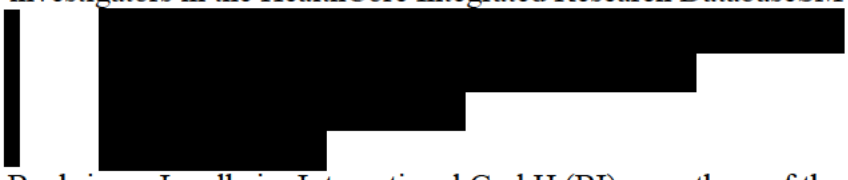
HCPCS	Healthcare Common Procedure Coding System
HES	Hospital Episode Statistics
IACS	Instituto Aragonés de Ciencias de la Salud (Aragon Institute of Health Sciences) in Aragon, Spain
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th Revision
ICD-9	International Classification of Diseases, 9th Revision
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
IRB	Institutional Review Board
IRR	Incidence Rate Ratio
ISAC	Independent Scientific Advisory Committee
MedDRA	Medical Dictionary for Drug Regulatory Activities
NICE	National Institute for Health and Care Excellence
OQA	Office of Quality Assurance (RTI Health Solutions)
PaCO2	Partial Pressure of Carbon Dioxide
PASS	Post-Authorisation Safety Study
PPV	Positive Predictive Value
RR	Relative Risk
RTI-HS	RTI Health Solutions
SGLT2	Sodium-Glucose Cotransporter 2
SIDIAP	Sistema d'Informació per el Desenvolupament de la Investigació en Atenció Primària (Information System for the Advancement of Research in Primary Care), Catalonia, Spain
SIRS	Systemic Inflammatory Response Syndrome
STD	Sexually Transmitted Disease
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
T1D	Type 1 Diabetes Mellitus
T2D	Type 2 Diabetes Mellitus
THIN	The Health Improvement Network
UK	United Kingdom
UKPDS	UK Prospective Diabetes Study
ULN	Upper Limit of the Normal Range
US	United States

UTI	Urinary Tract Infection

3. RESPONSIBLE PARTIES

The responsible parties, study investigators, and protocol authors are as follows:

- RTI Health Solutions (RTI-HS): coauthors of the protocol, study investigators responsible for the study in the Clinical Practice Research Datalink (CPRD) database in the United Kingdom (UK), and study coordinator

- Department of Clinical Epidemiology from the Aarhus University Hospital, Denmark: coauthors of the protocol and study investigators in the Danish Population Registries

- HealthCore, in the United States (US): coauthor of the protocol and study investigators in the HealthCore Integrated Research DatabaseSM (HIRD)

- Boehringer Ingelheim International GmbH (BI): coauthors of the protocol and sponsors of the study


The study investigators share responsibility with 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) about the study protocol, the progress of the study, and study results.

4. ABSTRACT

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021
Title of study:	Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated with empagliflozin compared to patients treated with DPP-4 inhibitors		
Rationale and background:	As part of the risk management plan, Boehringer Ingelheim International GmbH (BI) has committed to perform a post-authorisation safety study (PASS) to evaluate the liver and renal safety of empagliflozin. The study will also evaluate the risks of severe complications of urinary tract infection (UTI) and genital infections (GI). Diabetic ketoacidosis (DKA) was included in the protocol amendment (version 4.0) as an additional safety topic in line with BI's commitment within the Article 20 referral, started by the European Medicines Agency (EMA) in June 2015. Two data sources and an extension of the study period to meet the required sample size to fulfil the study objectives were added in protocol amendment (version 5.0). This protocol also proposed refinements to the study design to ensure that the analysis will be adequate to address the different data sources and the complexity of a multinational database study. Among these refinements was a streamlining of the comparisons to a single comparator group. Version 6.0 clarified some of the changes proposed in version 5.0, to keep severe GI as a secondary endpoint and to report the number of patients with elevated liver enzymes. Version, 7.0, further clarifies some aspects of the data analysis. The current version, 8.0, adjusts the study milestones that have been affected by the COVID-19 pandemic, leading to a delay in the reception of data for the full study period in Denmark and in the start of validation activities in the UK CPRD.		
Research question and objectives:	To estimate, among patients with type 2 diabetes mellitus (T2D), the risk of acute liver injury (ALI), the risk of acute kidney injury (AKI) and chronic kidney disease (CKD), the risk of severe complications of UTI, the risk of GI, and the risk of DKA among patients treated with empagliflozin compared with patients treated with dipeptidyl peptidase-4 (DPP-4) inhibitors.		

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021
Study design:	This will be a non-interventional cohort study using existing data (records from routine medical care). The study will use a new-user design and compare new users of empagliflozin with new users of DPP-4 inhibitors. Propensity scores based on information before or at the index date will be used to account for potential confounding. The index date will be defined as the date on which each identified new user receives the index prescription for empagliflozin or a DPP-4 inhibitor.		
Population:	The study population will include all eligible patients with T2D initiating treatment with empagliflozin or with a DPP-4 inhibitor. Eligible patients will be included if they are aged 18 or more years and have at least 12 months of continuous registration in the data source. The data sources included in the study will be the Clinical Practice Research Datalink (CPRD) in the United Kingdom (UK). For the evaluation of the rarest outcomes, the Danish Population Registries in Denmark, and the HealthCore Integrated Research Database^{SM} (HIRD) in the United States (US) will be added. • Each member of the empagliflozin-exposed population must have at least one prescription/dispensing for empagliflozin, with or without other glucose-lowering drugs (GLDs), and no prior prescription/dispensing of empagliflozin, other SGLT2 inhibitors, or DPP-4 inhibitors during the 12 months before or at the index date. • Each member of the population exposed to a DPP-4 inhibitor must have at least one prescription/dispensing for a DPP-4 inhibitor, with or without other GLDs, and no prior prescription/dispensing of a DPP-4 inhibitor, empagliflozin, or other SGLT2 inhibitor during the 12 months before or at the index date. New users of the study drugs would usually be (1) switching from monotherapy with another GLD to monotherapy with a study drug, (2) switching from dual or triple therapy with another GLD to dual or triple therapy with a study drug and other GLDs, (3) adding a study drug to therapy with one or two other GLDs to become patients on dual or triple therapy or (4) naive to GLD treatment and start the study drugs as first-line therapy. Treatment complexity will be considered in the analysis. Prior or current use of insulin is allowed, and an additional stratified analysis by insulin use is planned. Only patients with T2D will be included in the study. Patients with type 1		

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021

	diabetes (T1D) will be excluded. Algorithms to identify T2D and T1D will be adapted to the type and availability of data in each data source. Different exclusion criteria will be applied according to each of the outcomes of interest (e.g., patients with CKD will be excluded from the analysis of AKI), which will result in different cohorts (see Section 9.2.8.1). Follow-up will start the day after the index date and, for each specified outcome, will continue until the occurrence of the study outcome, end of study data, the date during follow-up that specific exclusion criteria are met, the end date of the first continuous treatment of the index drug (empagliflozin or DPP-4 inhibitor) plus a defined grace period (30 days after the end of the last prescription's days' supply in main analyses), or the date on which a new treatment episode starts with the other index drugs or other SGLT2 inhibitors.
Variables:	*Primary outcomes*: • Hospitalisation, emergency department (ED) visit, or specialist visit for ALI in patients without predisposing conditions • Hospitalisation, ED visit, or specialist visit for AKI • Hospitalisation or ED visit due to DKA • Severe complications of UTI (inpatient and outpatient)—only in the CPRD • Genital infections (inpatient and outpatient)—only in the CPRD *Secondary outcomes*: • Hospitalisation, ED visit, or specialist visit for ALI in patients with and without predisposing conditions • Chronic kidney disease (inpatient and outpatient)—only in the CPRD • Severe GI—only in the CPRD A sensitivity analysis will be performed by adding outpatient primary care cases of ALI and AKI to the ALI and AKI primary outcomes. This analysis will be performed only in the CPRD and HIRD, where primary care data are available. Validation of identified cases will be implemented for all outcomes in the three data sources. In addition, a random sample of 100 additional outpatient/primary cases of ALI and AKI will also be validated in the CPRD and HIRD. If the number of cases identified for any of the outcomes is 100 or

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021

	fewer, all cases will be validated for those outcomes. Otherwise, a random sample of cases will be selected. The number of selected cases for validation will be based on the proportion of patients that can be validated in each database, and the target is to validate 100 cases. Cases identified in the CPRD and Hospital Episode Statistics (HES) will be validated through questionnaires sent to general practitioners (GPs). Cases identified in the Danish Population Registries and in the HIRD will be validated through medical record data abstraction and/or laboratory test results. *Exposures (index drugs)*: • Empagliflozin (and fixed-dose combinations with metformin) • DPP-4 inhibitors: sitagliptin, saxagliptin, linagliptin, vildagliptin, alogliptin (and fixed-dose combinations of these drugs with metformin) Fixed-dose combinations of SGLT2 inhibitors with DPP-4 inhibitors will not be included in the study. Current use of the index drugs will be defined from the date of prescription of empagliflozin or DPP-4 inhibitor to the end of supply for that prescription plus a period of 30 days. Recent use will be defined from the end of current use (30 days after end of supply) through 90 days later (which is 120 days after end of supply). End of day's supply will be estimated according to prescription instructions in the CPRD or based on available information on the duration of dispensings (e.g., number of packages bought, strength, and number of pills) in Denmark and the HIRD.
Data sources:	The data sources included in the study will be the CPRD in the UK, and for the evaluation of the rarest outcomes, also the Danish Population Registries in Denmark and the HIRD in the US. The **CPRD** contains diagnostic and prescribing information recorded by GPs as part of their routine clinical practice in the UK. The database currently contains data for over 16.7 million patients in the CPRD GOLD database and 22.7 million patients in the CPRD Aurum database, representative of the UK population in terms of age and sex, with research-quality data from 790 UK practices in CPRD GOLD and 873 practices in CPRD Aurum; 2.6 million of these patients in CPRD GOLD and 7.3 million patients in CPRD Aurum are active. Detailed information on prescriptions written by the GPs, including

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021

	prescribed dose and duration, is routinely recorded in the database. Additional diagnostic and treatment information can be found in letters from specialists, hospitals, and other sources. Linkage to the HES and cause-of-death information is currently available for approximately 54% of patients in CPRD GOLD and 93% of patients in CPRD Aurum [R19-1734]. The **Danish Population Registries.** The Danish health care system provides universal coverage to all Danish residents (5.7 million inhabitants). The proposed registries are the Danish National Patient Register, which includes data on all hospital admissions since 1 January 1977 and on specialist outpatient hospital clinic and ED visits since 1995, and the National Health Services Prescription Database, which encompasses the reimbursement records of all reimbursed drugs sold in community pharmacies and hospital-based outpatient pharmacies in Denmark since 2004. The national health registries can be linked to all other national databases through the unique Civil Personal Registration Number. No primary care diagnosis data are available for research purposes, but the new nationwide Register of Laboratory Results for Research (LAB_F) tracks all laboratory test results from both primary and secondary care. The **HIRD** contains geographically diverse longitudinal medical and pharmacy claims data from approximately 40 million health plan members across the US. Member enrolment, medical care (professional and facility claims), outpatient prescription drug use, outpatient laboratory test result data (available for 30% of the patients), and health care utilisation may be tracked for health plan members in the database dating back to January 2006. For approximately 50% of members, data in the HIRD can be linked to inpatient and outpatient medical records (source records for validation) and to national vital records (date and cause of death).
Study size:	The study size is driven by the uptake of empagliflozin after marketing approval in the UK, Denmark, and US. With a comparator:empagliflozin ratio of 10:1 and a power of 80%, the study size required to detect an incidence rate ratio (IRR) of 3 (for the comparison of empagliflozin to comparator) is between 18,000 and 30,000 person-years of empagliflozin use (among new users) for ALI, approximately 8,000 person-years for DKA, and 1,400 to 3,200 person-years of empagliflozin use for AKI. For all other outcomes, the number of empagliflozin new-user person-years required to detect an IRR of 3

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021

	would be less than 1,400. The number of empagliflozin new users and crude incidence rates of the events of interest have been assessed annually starting June 2016. Based on data from the second (2017) and third (2018) interim reports, the number of expected empagliflozin users by the end of the study period in the UK was projected to be insufficient to answer the scientific questions of interest for the outcomes of ALI, AKI, and DKA. The addition of data sources from Denmark and the US for these outcomes and the extension of the study period from 3 to 5 years after launch for all the outcomes should allow the target number of users to accrue.
Data analysis:	The following estimates and comparisons will be generated: • Crude and adjusted incidence rates of each of the outcomes among empagliflozin new users and DPP-4 inhibitor new users. Incidence rates will be reported as point estimates (in cases per 1,000 person-years) and 95% confidence intervals (CIs). • Summary IRRs, after adjusting for propensity score deciles, among empagliflozin new users vs. DPP-4 inhibitor new users. The adjusted IRRs for each of the primary outcomes will be the main effect estimates of interest. Adjusted incidence rates and IRRs will be calculated using analytic techniques involving stratification by categories of propensity scores. An additional analysis will further stratify the IRRs by categories of insulin use at the index date. Sensitivity analyses will be performed to evaluate the potential for other sources of bias and confounding. Meta-analytic methods will be used to combine the IRRs obtained from the main analysis performed by all the data sources.
Milestones:	The start of data collection (data extraction) is planned to occur between September 2019 and December 2019, depending on the time lag in each data source. Annual progress reports, including monitoring of users in each data source, and project status were sent to EMA in June 2019 and June 2020. The end of data collection was planned to occur between July 2020 and October 2020, once validation was finalised and all data were available to perform the planned analysis. However, validation could not be performed in time due to the COVID-19 pandemic. There were delays in receiving the final datasets in Denmark, and the start of data validation activities were also delayed in the

Name of company: Boehringer Ingelheim International GmbH			
Name of finished medicinal product: Jardiance Synjardy			
Name of active ingredient: A10BK03 Empagliflozin A10BD20 Empagliflozin/metformin			
Protocol date: 13 Jun 2019	**Study number:** 1245.96	**Version/Revision:** 8.0	**Version/Revision date:** 19 July 2021
	UK. Therefore, the end of data collection will occur between July 2020 and March 2022 and the final report will be produced after completion of validation activities and analysis in all three data sources included in this study and is expected to be finalized by December 2022.		

5. AMENDMENTS AND UPDATES

Protocol version 5.0 was to ensure enough new users of empagliflozin meet the study objectives by adding additional data sources to evaluate the rarest outcomes and extending the study period for all outcomes. Considering the complexity of study with two comparator groups and multiple outcomes, and the inclusion of multiple databases, the study will be streamlined to focus on the core study objectives and facilitate the interpretation of the results. The exploratory analysis of one of the comparator groups (i.e., other SGLT2 inhibitors) was deleted. The definition of the rarest outcomes has also been modified to have primary outcomes that are evaluable in all three data sources in a consistent manner. Finally, given the low number of cases for some of the outcomes, the plan for secondary and sensitivity analyses has been revised to avoid issues with sparse data. Protocol version 6.0 clarified some of the changes proposed in version 5.0, to keep severe GI as a secondary endpoint, and to report the number of patients with elevated liver enzymes. Protocol version 7.0 further clarifies aspects of the data analysis. In Protocol version 8.0, the study milestone for the final report is updated to account for delayed validation activities due to the COVID-19 pandemic.

Version number	Protocol date	Comment
1.0	05 Feb 2015	Draft protocol v1: submitted to PRAC for review
2.0	23 Jun 2015	Draft protocol v2: submitted to PRAC for review
3.0	21 Oct 2015	PRAC approved protocol
4.0	10 Jun 2016	PRAC approved protocol amendment no 1
5.0	17 May 2018	Draft protocol amendment no 2 (v1): submitted to PRAC for review
6.0	03 Dec 2018	Draft protocol amendment no 2 (v2): submitted to PRAC for review
7.0	03 Jun 2019	Draft protocol amendment no 2 (v3): submitted to PRAC for review
8.0	12 July 2021	Draft protocol amendment no 3: submitted to PRAC for review

A detailed overview of changes implemented since the PRAC approved protocol (version 3.0) is presented below.

Version number	Date	Section of study protocol	Amendment or update	Reason
8.0	12 July 2021	Section 3	Updated responsible parties	To reflect changes of responsible parties
8.0	12 July 2021	Section 4	Updated abstract	To reflect changes in the protocol milestones that have been delayed due to the COVID-19 pandemic.
8.0	12 July 2021	Section 6	Updated milestones, with a final report in December 2022	Completion of the final study report is delayed due to the COVID-19 pandemic.
7.0	03 Jun 2019	Section 4	Updated abstract	To reflect changes in the protocol
7.0	03 Jun 2019	Section 9.3	Updated list of DPP-4 inhibitors	As per PRAC request
7.0	03 Jun 2019	Section 9.3.2.1, 9.3.2.3, and 9.3.2.8	Clarified that no ED visit information is available in HES; ED visits will be identified from the CPRD primary care records	As per PRAC request, to clarify in which outcomes ED visits will be used, and to describe ED visits ascertainment in CPRD
7.0	03 Jun 2019	Section 9.4.1 and 9.5	Confirmed the use of CPRD Aurum and updated information about CPRD GOLD and CPRD Aurum databases, and on monitoring of users in CPRD Aurum	To increase study size in the UK. To describe the number of users expected to increase the study size in the UK, based on the monitoring of users of empagliflozin as of April 2019
7.0	03 Jun 2019	Section 9.7.5	Added information on the mechanism of missing not at random	To clarify that for variables that are missing not at random, the missingness itself, by definition, is an unmeasured confounder. Sensitivity analyses evaluating the potential influence of unmeasured confounders have already been included.
7.0	03 Jun 2019	Section 9.7.6	Added information on sensitivity analyses	To document fully in one place all of the sensitivity analyses to be performed. Furthermore, an additional sensitivity analysis is described.
7.0	03 Jun 2019	Section 9.9.3.3	To clarify that the proportion of patients aged 65 years or	As per PRAC request

Version number	Date	Section of study protocol	Amendment or update	Reason
			older will be described for all data sources	
7.0	03 Jun 2019	Annex 4, Table 12, and Table 14	Codes related to male complications of genital infections have been moved from Table 12 to Table 14	To correct prior typographical error
6.0	03 Dec 2018	Section 4	Update abstract	To reflect changes in the protocol
6.0	03 Dec 2018	Section 8	Updated study objectives	To formulate a study objective for each primary and secondary outcome and retain severe genital infection as a secondary outcome
6.0	03 Dec 2018	Section 9.2.1	Updated text to provide further details	To further clarify case ascertainment and identification of T2D in each data source
6.0	03 Dec 2018	Section 9.2.6	Clarified lookback period	To clarify that lookback period will be the same for most of the variables within each database, except for a small number of specific covariables
6.0	03 Dec 2018	Section 9.2.7	Updated wording of inclusion criteria	To clarify the inclusion criteria, per PRAC request
6.0	03 Dec 2018	Section 9.3.1	Updated list of DPP-4 inhibitors products	To include medications approved since previous versions of the protocol
6.0	03 Dec 2018	Section 9.3.2	Updated outcomes definition	To include ED visits in addition to those outcomes that require hospitalisation
6.0	03 Dec 2018	Section 9.3.2	Updated validation of ALI cases	To describe cases of ALI that have ALT and/or AST ≥ 3 × ULN but < 5 ULN and therefore did not fulfil the Aithal et al. (2011)[R14-1933] criteria
6.0	03 Dec 2018	Section 9.3.2	Reintroduced the genital infection secondary outcome	Per PRAC request, retained severe genital infection as an outcome and added a description of complications of genital infection
6.0	03 Dec 2018	Section 9.3.3	Added description of cases of elevated	To gain more knowledge about the potential hepatotoxicity of

Version number	Date	Section of study protocol	Amendment or update	Reason
			liver enzymes identified through laboratory results	empagliflozin
6.0	03 Dec 2018	Section 9.4.1	Added description of CPRD Aurum	Per PRAC request
6.0	03 Dec 2018	Section 9.7.1	Provided additional details on potential methods to deal with propensity score deciles with zero events	To suggest alternative methods of estimating the adjusted treatment effect from propensity scores (in the situation of zero margins in propensity score strata)
6.0	03 Dec 2018	Section 9.7.5	Additional clarifications have been added	To account for situations where multiple imputation would not be applicable
6.0	03 Dec 2018	Section 9.7.7	Provided more detail on meta-analysis techniques	To describe in more detail the meta-analytic techniques and clarify that a random-effects model will be used as the main analysis and the fixed-effects model will be used in a sensitivity analysis
6.0	03 Dec 2018	Section 9.9	Added limitations	Per PRAC request. To provide details on limitations by data source, limitations due to current new user definition and to study size using all data sources, and generalisability of the results
6.0	03 Dec 2018	Annex 4	Reintroduced Read codes to identify study outcomes	Per PRAC request
5.0	17 May 2018	Section 3	Added responsible parties from the new data sources added to the study	To incorporate new responsible parties
5.0	17 May 2018	Section 4	Updated abstract	To reflect all changes to version 5.0
5.0	17 May 2018	Section 6	Updated milestones, with a final report in 2021	Changed due to extension of the study period and considering the time lag between the end of the study period and data extraction and the time needed for analysis and validation of the outcomes in each data source

Version number	Date	Section of study protocol	Amendment or update	Reason
5.0	17 May 2018	Section 7	Updated epidemiology data for DKA	A study of interest was published after the approval of protocol version 4.0
5.0	17 May 2018	Section 8	Deleted the comparison between empagliflozin and other SGLT2 inhibitors from the study objectives	The comparator group "other SGLT2 inhibitors" was qualified as exploratory in previous reviews of the protocol done by the EMA. The study is evaluating class effects, so it is not expected to find differences in the risk of the outcomes of interest between different SGLT2 inhibitors; if differences exist, they will be small, and this study will not have the power needed to detect them. MAHs of other SGLT2 inhibitors are doing similar analyses for their products in similar or the same data sources.
5.0	17 May 2018	Section 8	Updated study objectives	To reflect changes in comparator, clarify that incidence rates of secondary outcomes will also be estimated, and streamline stratified analyses
5.0	17 May 2018	Section 9.1	Update study design to add data sources and extended the study period	Per EMA feedback to the second interim report
5.0	17 May 2018	Section 9.2	Updated setting to add data sources and extended the study period	Per EMA feedback to the second interim report
5.0	17 May 2018	Section 9.2	Updated study population to modify the definition of new users to not have a prescription or dispensing of the study medications during the 12 months prior to the index date (previously defined as no prior use ever)	This will allow inclusion of more users of empagliflozin in the study, the main limitation of which is the accrual of patients exposed to the drug of interest. This definition has been used in other EMA-approved PASS, and it is considered appropriate because the comparators are indicated for the same target population and same stage of the disease, and a 12-month wash-out period is considered enough for the acute

Version number	Date	Section of study protocol	Amendment or update	Reason
				outcomes being evaluated.
5.0	17 May 2018	Section 9.3.1	Updated exposures to delete other SGLT2 inhibitors	Discussed above (see amendments to Section 8); exposure identified by prescriptions
5.0	17 May 2018	Section 9.3.2	Deleted mention of Read codes and referred to codes in the CPRD as "primary care" codes	To refer to all potential codes that may appear in the CPRD primary care databases: CPRD GOLD uses Read codes, but CPRD Aurum, which may be needed in the future, uses Read codes, SNOMED codes, and EMIS local codes
5.0	17 May 2018	Section 9.3.2	Modified the ALI primary outcome definitions to be hospitalisation or referral to a specialist for ALI in patients with no predisposing conditions	To have a primary outcome that can be assessed in the three data sources included in the study. Outpatient ALI cannot be assessed in Denmark due to lack of primary care data but will be assessed in a sensitivity analysis in the CPRD and HIRD
5.0	17 May 2018	Section 9.3.2	Modified ALI secondary endpoint definition to be hospitalisation or referral to a specialist for ALI in patients with and without predisposing conditions	To have a secondary outcome that can be assessed in the three data sources included in the study
5.0	17 May 2018	Section 9.3.2	Modified the AKI primary outcome definitions to be hospitalisation or referral to a specialist for AKI	To have a primary outcome that can be assessed in the three data sources included in the study. Outpatient AKI cannot be assessed in Denmark due to lack of primary care data but will be assessed in a sensitivity analysis in the CPRD and HIRD
5.0	17 May 2018	Section 9.3.2	Deleted UTI secondary outcome and modified primary UTI outcome to include outpatient diagnosis	Based on PRAC feedback to versions 1 and 2 of the protocol

Version number	Date	Section of study protocol	Amendment or update	Reason
5.0	17 May 2018	Section 9.3.2	Deleted the genital infection secondary outcome	The genital infection primary outcome already includes outpatient and inpatient cases and is considered appropriate and adequate to address this safety concern
5.0	17 May 2018	Section 9.3.3	Updated covariates section	Updated based on type and availability of data in each data source
5.0	17 May 2018	Section 9.4	Added data sources and extended the study period	Per EMA feedback to the second interim report
5.0	17 May 2018	Section 9.7.1	Clarified the selected approach for using propensity scores to adjust for confounding	Based on literature, we have selected stratification by propensity score categories to adjust for confounding
5.0	17 May 2018	Section 9.7.4	Specified stratification variables during follow-up and deleted mention of the method	Specific method to evaluate potential confounders during follow-up will be defined in the statistical analysis plan
5.0	17 May 2018	Section 9.7.7	Added analysis using positive predictive value	To correct for outcome misclassification
5.0	17 May 2018	Section 9.7.8	Added meta-analysis	To combine the incidence rate ratios obtained from the main analysis of the cohort study performed in the different data sources
5.0	17 May 2018	Section 9.8	Updated quality-control section	To describe quality-control process in the newly added data sources
5.0	17 May 2018	Section 9.9	Updated limitations	To discuss limitations driven by changes in the protocol given the addition of new data sources
5.0	17 May 2018	Annex 4 and 5	Read codes have been deleted	To retain codes that are common to the three data sources, streamline the protocol, and because CPRD Aurum, which uses other coding systems, may be used

Version number	Date	Section of study protocol	Amendment or update	Reason
5.0	17 May 2018	Annex 7 and 8	These annexes were deleted. Details on the CPRD in previous Annex 7 have been integrated with details on the Danish Health Registries and HIRD. Information on Spanish data sources has been deleted	To facilitate comparison of the three data sources in one table instead of two or three tables and to update the annex to include the new study data sources
4.0	10 Jun 2016	Cover page and abstract	Addition of Synjardy (empagliflozin/ metformin)	Synjardy came on the market in August 2013, after the approval of protocol version 3.0
4.0	10 Jun 2016	Section 7.1	Added data on DKA in clinical trials and safety studies	New safety outcome added to the study
4.0	10 Jun 2016	Section 7.7	Added data on the epidemiology of DKA	New safety outcome added to the study
4.0	10 Jun 2016	Section 8	Added the evaluation of the risk of DKA as a primary objective	New safety outcome added to the study
4.0	10 Jun 2016	Section 9.3.2 and Annex 5	Added definition of the DKA outcome	New safety outcome added to the study
4.0	10 Jun 2016	Section 9.6	Changed chi-square sample size calculation method for Poisson method	The Poisson method is considered more appropriate for rare events, and confidence intervals are also going to be estimated using Poisson method
4.0	10 Jun 2016	Annex 6	Added covariates to be considered for inclusion in the propensity score model for DKA	New safety outcome added to the study

AKI = acute kidney injury; ALI = acute liver injury; CPRD = Clinical Practice Research Datalink; DKA = diabetic ketoacidosis; EMA = European Medicines Agency; HIRD = HealthCore Integrated Research DatabaseSM; MAH = marketing authorisation holder; PASS = post-authorisation safety study; PRAC = Pharmacovigilance Risk Assessment Committee; SGLT2 = sodium-glucose cotransporter 2; UTI = urinary tract infection.

6. MILESTONES

Milestone	Planned Date	Actual / Revised Date
Protocol (version 3.0) endorsed by the EMA	January 2016	January 2016 Protocol amendment no 1 (version 4.0 to add DKA) endorsed by EMA in September 2017
Start of data collection[1]	15 March 2016	15 March 2016 (monitoring in the UK) September 2019 (first data source) through December 2019 (last data source)
End of data collection[2]	31 December 2017	July 2020 (first data source) through March 2022 (dependent on the start of data collection and duration of validation activities in each data source)
First interim report	Based on data available 19 months after use of empagliflozin was first captured in the CPRD Expected in June 2016	Submitted in June 2016
Second interim report	Based on data available 26 months after use of empagliflozin was first captured in the CPRD Expected in June 2017	Submitted in June 2017
Third interim report	Based on data available 37 months after use of empagliflozin was first captured in the CPRD Expected in June 2018	Submitted in June 2018
Registration in the EU PAS Register[3]	10 May 2016	10 May 2016 (last updated 03 Jun 2020)
Annual progress reports[4]	Expected in June 2019 and June 2020	Submitted in June 2019 and June 2020

Milestone	Planned Date	Actual / Revised Date
Final report of study results [5]	As per protocol version 3.0 final report expected in July 2018. As per Protocol version 7.0, the planned date for completion of the study report is September 2021. As per Protocol version 8.0, the planned date for completion of the study report is December 2022.	December 2022

CPRD = Clinical Practice Research Datalink; EMA = European Medicines Agency; UK = United Kingdom.

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.

1. Start of data collection: the date from which information on the first study subject is first recorded in the study data set or, in the case of secondary use of data, the date from which data extraction starts [IR Art 37(1)]. Simple counts in a database to support the development of the study protocol, for example, to inform the sample size and statistical precision of the study, are not part of this definition [R13-5420]. For additional details on the dates of start of data collection, see the feasibility assessment (listed in Annex 1).
2. End of data collection: the date from which the analytical data set is completely available [IR Art 37(2)] [R13-5420].
3. ENCEPP/SDPP/13413; http://www.encepp.eu/encepp/viewResource htm?id=13414.
4. Will include monitoring the number of users in the databases, including CPRD Aurum, and study status.
5. Study milestone for the final report is dependent on the initiation of validation activities in CPRD. December 2022 is an estimate based on communication received from CPRD on 09 Jul 2021, indicating that validation would be initiated end of July 2021.

7. RATIONALE AND BACKGROUND

Jardiance (empagliflozin), a highly potent and selective inhibitor of the sodium-glucose cotransporter 2 (SGLT2), was approved in Europe in May 2014 for the treatment of type 2 diabetes mellitus (T2D) to improve glycaemic control in adults. SGLT2 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 for monotherapy and add-on combination therapy with other glucose-lowering medicinal products including insulin. 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].

The overall frequency of treatment-emergent adverse events observed in the empagliflozin clinical trials and safety studies was comparable between treatment groups, between 70% and 74% [P14-17456].

As part of the risk management plan, Boehringer Ingelheim International GmbH (BI) has committed to conduct a post-authorisation safety study (PASS) to evaluate the liver safety of empagliflozin due to a higher frequency of serious hepatic events in clinical trials and renal

safety of empagliflozin due to its mechanism of action. For the available information on renal and liver safety of empagliflozin, please refer to Section 7.1.

The study will also evaluate the risks of (1) severe complications of urinary tract infection (UTI) and (2) genital infection (GI). The rationale for looking at these risks is the fact that inhibition of SGLT2 in patients with T2D leads to excess glucose excretion in the urine [R14-4617], which, together with hyperglycaemia, may be the main cause of increased susceptibility of patients with diabetes to UTI and GI. Although this mechanism is not completely understood, it is known that increased glucose levels in genitourinary tissues enhance yeast adhesion and growth; thus, by providing a favourable growth environment for otherwise commensal microorganisms, glycosuria could potentially increase the risk for UTIs, vulvovaginitis, and balanitis. Moreover, hyperglycaemia not only impairs various aspects of host defence, including neutrophils and complement proteins, but also promotes the virulence of infecting organisms in patients with diabetes [P14-02878, R14-5237].

In addition, version 4.0 of the protocol added DKA as a safety topic of interest. The rationale for assessing this risk was that cases of DKA occurred in patients taking SGLT2 inhibitors for T2D, and a number of these cases had been atypical, with patients not having blood sugar levels as high as expected or even with levels in the normal range [P15-08785]. Clinically, DKA is defined by the biochemical triad of ketosis, hyperglycaemia, and acidosis. The condition usually occurs as a consequence of absolute or relative insulin deficiency that is accompanied by an increase in counter-regulatory hormones (i.e., glucagon, cortisol, growth hormone, catecholamines). This type of hormonal imbalance enhances hepatic gluconeogenesis and glycogenolysis, resulting in severe hyperglycaemia. Enhanced lipolysis increases serum free fatty acids that are then metabolised as an alternative energy source in the process of ketogenesis. This results in accumulation of large quantities of ketone bodies and subsequent metabolic acidosis. Although DKA has been considered to be indicative, or even diagnostic, of type 1 diabetes mellitus (T1D), cases of ketone-prone T2D are increasingly being recognised [R16-1372, R16-1373]. Atypical DKA reported in patients with T2D treated with SGLT2 has a different origin than that observed in patients with type 1 diabetes mellitus (T1D). Full-dose SGLT2 inhibition induces a rapid increase in urinary glucose excretion, ranging from 50 to 100 g/day, lasting slightly more than 24 hours. Concomitant insulin intensification therapy was a common factor in most if not all cases. Depending on body size, glomerular filtration rate, and degree of hyperglycaemia, SGLT2-induced glucose loss can make up a substantial fraction of daily carbohydrate availability [P15-08785]. The different pathophysiology of DKA vs. atypical DKA induced by SGLT2 inhibitors is that (1) in the latter, insulin deficiency and insulin resistance are milder, with less glucose overproduction and underutilisation, which in most cases led to significant reductions in total daily insulin requirements, consequently leading to hypoinsulinaemia, and (2) renal glucose clearance (i.e., the ratio of glycosuria vs. glycaemia) is twice as large in atypical DKA as in DKA. Ketoacidosis follows the same sequence of events in both presentations of DKA. However, in SGLT2-treated patients with T2D, the lower insulin-to-glucagon ratio stimulates lipolysis (increase in free fatty acids) and enhances lipid oxidation at the expense of carbohydrate oxidation. At low glucose concentrations, non-oxidative glucose disposal falls. The augmented free fatty acids delivered to the liver result in mild stimulation of ketogenesis, while fasting and mean post-meal beta-hydroxybutyrate levels increase;

conversely, plasma lactate levels decrease as an expression of reduced carbohydrate utilisation [P15-08785].

This protocol describes a cohort study, to be conducted among patients with T2D, comparing the incidence of the outcomes of interest in patients initiating empagliflozin compared with patients initiating a DPP-4 (dipeptidyl peptidase-4) inhibitor. For each outcome, different inclusion/exclusion criteria will be applied, resulting in slightly different study cohorts to be assessed for the risk of each outcome.

7.1 DATA ON THE OUTCOMES OF INTEREST IN EMPAGLIFLOZIN CLINICAL TRIALS AND SAFETY STUDIES

When the original protocol was developed, according to the Jardiance assessment report, published 20 March 2014 and including data across empagliflozin clinical trials and safety studies, as of 31 August 2012, there were 3,522 patients in the placebo group, with a total duration of exposure of 2,758.1 patient-years; 3,630 patients in the empagliflozin 10-mg group, with a total duration of exposure of 3,258.2 patient-years; and 4,602 patients in the empagliflozin 25-mg group, with a total duration of exposure of 4,448.1 patient-years [P14-17456].

The frequency of hepatic injury (Standard MedDRA Query) was low and similar for all treatment groups (see Table 1) [P14-17456]. However, the number of serious hepatic events was higher in patients treated with empagliflozin than in the placebo group, with 19 of the 22 cases of serious liver enzyme elevations occurring in one of the empagliflozin groups. There was also some imbalance seen for elevations of alanine aminotransferase (ALT) and/or aspartate transaminase (AST) (≥ 5 x the upper limit of the normal range [ULN], ≥ 10 x ULN, and ≥ 20 x ULN), when comparing empagliflozin and placebo users. A total of 7 patients (5 during treatment with empagliflozin, 1 after treatment with empagliflozin, and 1 during the screening period) had laboratory values consistent with Hy's Law of drug-induced liver injury [R14-5256], although none of them were finally qualified as drug-induced liver injury due to plausible alternative hypotheses.

The frequency of decreased renal function was low and similar for all treatment groups (see Table 1) [P14-17456]. The most common adverse event was renal impairment: 0.5% among the placebo group and 0.7% each among the empagliflozin 10-mg and 25-mg groups.

The frequency of UTIs was similar for all treatment groups (see Table 1) [P14-17456]. The incidence rate of UTI was 10.9 per 100 patient-years among the placebo group, 10.5 per 100 patient-years among the empagliflozin 10-mg group, and 9.6 per 100 patient-years among the empagliflozin 25-mg group.

The frequency of GI was consistently higher among empagliflozin groups, compared with placebo (see Table 1) [P14-17456]. The proportion of women with GI was 2-fold higher among women than among men. Empagliflozin groups had a 4-fold higher rate of GI than comparator groups irrespective of sex [P14-17456].

In a retrospective analysis of randomised phase 2 and 3 empagliflozin trials (13,000 participants with T2D), there were eight events consistent with DKA, with no imbalance observed between patients treated with empagliflozin 10 mg (two events), empagliflozin 25 mg (one event), and placebo (five events). In the cardiovascular outcome trial EMPA-REG, with approximately 7,000 patients, the frequency of reported blinded events of DKA was less than 0.1% [P15-08785].

A meta-analysis of clinical trials of users of SGLT2 inhibitors vs. users of other antidiabetic medications reported an overall odds ratio of AKI of 0.80 (95% CI, 0.67-0.96), of UTI of 1.15 (95% CI, 1.00-1.33), of male genital infections of 3.61 (95% CI, 3.10-4.19), of female genital infections of 3.17 (95% CI, 2.15-4.68), and of DKA of 1.96 (95% CI, 0.77-4.98) [P18-00901].

Table 1 Frequency of adverse events of interest by treatment group in the empagliflozin clinical trials and safety studies – all patients

Adverse event	**Frequency (%) of adverse events by treatment group**		
	Placebo (N = 3,522)	**Empagliflozin 10 mg (N = 3,630)**	**Empagliflozin 25 mg (N = 4,602)**
Hepatic injury	1.5	1.2	1.4
Decreased renal function	1.0	1.1	1.3
UTI	8.1	8.9	8.8
GI	1.0	4.4	4.7
DKA[1]	5	2	1

DKA = diabetic ketoacidosis; GI = genital infection; UTI = urinary tract infection.

1 Source for diabetic ketoacidosis: Rosenstock and Ferrannini [P15-08785]. The total number of patients in each treatment group does not include the diabetic ketoacidosis outcome, for which these numbers were not reported.

Source for adverse events except for diabetic ketoacidosis: Jardiance, European Public Assessment Report [P14-17456].

7.2 EPIDEMIOLOGY OF ACUTE LIVER INJURY

7.2.1 Epidemiology of acute liver injury in general population

The incidence of acute liver injury (ALI) and drug-induced liver injury in the general population has been poorly investigated. Data from several observational studies suggest that the annual incidence of drug-induced ALI in the general population ranges from 0.7 cases per 100,000 persons (95% confidence interval [CI], 0.6-0.9) [P03-00488] to 13.9 cases per 100,000 persons [P02-05969].

The study with the lowest incidence was performed in a hospital network surveillance system in Spain between 1993 and 1998 and identified 107 cases of ALI with a total follow-up of 14.6 x 106 person-years [P03-00488]. The study with the highest incidence was conducted in France using intensive surveillance of cases in a well-defined geographic region between

1997 and 2000 and identified 34 cases of ALI [P02-05969]. Other cohort studies reported an incidence rate (per 100,000 persons) of drug-induced ALI of 2.3 (n = 77 cases) in a hospital outpatient hepatology clinic in Sweden between 1995 and 2005 [P06-11008] and 3.4 (N = 461 cases) in a Spanish regional register between 1994 and 2004 [P05-08822]. Similarly, in a study conducted in the United Kingdom (UK) Clinical Practice Research Datalink (CPRD) between 1994 and 1999, 128 cases of ALI were identified, and the estimated incidence rate was 2.4 per 100,000 person-years (95% CI, 2.0-2.8) [P04-07683].

7.2.2 Epidemiology of acute liver injury in the diabetes population

Even fewer data are available on the incidence of ALI among patients with diabetes. The incidence rate of ALI among patients with diabetes without risk factors for liver disease in the UK CPRD between 1994 and 1998 was estimated to be 14.2 cases per 100,000 person-years (n = 14 cases) [P03-03701]. Among users of oral glucose-lowering drugs (GLDs) the incidence rate was 22.0 cases per 100,000 person-years (n = 9 cases), and among users of insulin the incidence rate was 13.2 per 100,000 person-years (n = 3 cases). Among patients with diabetes, the adjusted relative risk of ALI among users of GLDs or insulin, compared with non-users of these medications, was 2.8 (95% CI, 0.6-12.5). In the same study, the incidence rate of ALI was estimated to be 8.8 per 100,000 person-years in the general population without diabetes [P03-03701].

In a cohort study performed in the United States (US) Veterans Affairs database between 1985 and 2000, 173,643 hospitalised veterans with type 1 and type 2 diabetes, identified through ICD-9 (International Classification of Diseases, 9th Revision) code 250, were matched to 650,620 hospitalised veterans without diabetes. The incidence rate of ALI was 2.31 per 10,000 person-years (n = 346 cases after 1,494,995 person-years of follow-up) among patients with diabetes, and 1.44 per 10,000 person-years (n = 942 cases after 6,556,350 person-years of follow-up) among patients without diabetes, with an adjusted hazard ratio of 1.44 (95% CI, 1.26-1.63). Chronic liver disease and increasing age increased the risk of ALI. Congestive heart failure was more frequent among patients with diabetes and ALI (31%) than patients without diabetes and ALI (22%). The 6-week mortality after hospitalisation with acute liver failure was 60% among patients with diabetes and 63% among patients without diabetes [R12-3632].

For the study size and power calculations (see Section 9.5), the estimates of the ALI incidence in the population with diabetes were used as reported by the two studies above [R12-3632, P03-03701]. Only one of these studies was performed in CPRD; however, as this study used data from 1994-1998 and was in a highly selected population (diabetes patients without risk factors for liver disease), both available studies were considered for the study size calculations.

7.3 EPIDEMIOLOGY OF ACUTE KIDNEY INJURY

Acute kidney injury (AKI) is characterised by an abrupt decline in renal function. The definition of acute renal failure in epidemiologic studies has been based on absolute increases of serum creatinine from normal values (1.7 or 2 x ULN) or change from baseline (20% to 50%) or both [P02-05226, P05-04032, P14-17372, P14-17373, P14-17376, P14-17377].

7.3.1 Epidemiology of acute kidney injury in general population

Two studies have used large administrative and/or claims databases to examine secular trends in the epidemiology of AKI among the general population in the US. The first study was based on a total of 5,403,015 hospital discharges for AKI based on 5% of Medicare claims and reported that the incidence of AKI rose from 14.6 to 36.4 per 1,000 discharges between 1992 and 2001 [R11-5329]. The second study, using the Nationwide Inpatient Sample, analysed a total of 1,083,745 AKI discharges (320,370 of which required dialysis) and reported that AKI incidence rose from 4 to 21 per 1,000 discharges between 1988 and 2002. For AKI requiring dialysis, the percentage of annual discharges increased from 0.3 in 1988 to 2 per 1,000 discharges in 2002. For AKI requiring dialysis, the incidence rose from 4 per 100,000 population in 1988 to 27 per 100,000 population in 2002 [R14-5285].

7.3.2 Epidemiology of acute kidney injury in the diabetes population

Among patients with diabetes, a study performed in the UK CPRD between 2003 and 2007 reported an incidence rate of 198 per 100,000 person-years among patients with T2D and 27 per 100,000 patient-years among those without T2D. The adjusted incidence rate ratio (IRR) of AKI was 2.5 (95% CI, 2.2-2.7) among patients with T2D compared with patients with no T2D [R11-5319]. Although the manuscript does not provide the incidence of hospitalisation for AKI, it could be estimated to be 65% of the above incidence rates given that this percentage of patients with AKI was identified through a referral or hospitalisation for AKI.

Scarce data have been published on the risk of AKI associated with exposure to antidiabetic medications. A nested case-control study performed in the UK CPRD between 1997 and approximately 2004 (end of study period not reported) reported a relative risk of AKI of 2.5 (95% CI, 0.8-7.8) among patients with current exposure to insulin (n = 5) compared with unexposed patients (n = 98). Similarly, the relative risk of AKI was 1.3 (95% CI, 0.5-3.1) among patients exposed to other antidiabetic drugs (n = 11), compared with unexposed patients (n = 88) [P05-04032]. A cohort study compared the risk of AKI among SGLT2 inhibitor users and non-users using data from the Mount Sinai chronic kidney disease registry and the Geisinger Health System cohort. Hazard ratios of AKI among users vs. non-users were 0.4 (95% CI, 0.2-0.7) and 0.6 (95% CI, 0.4-1.1), respectively [P17-09718].

For the study size and power calculations, the estimates of the AKI incidence in the population with diabetes were used as reported in the UK CPRD study [R11-5319]. In addition, due to sparse data available, estimates reported in general population [R14-5258] have been used as well; see Section 9.5.

7.4 EPIDEMIOLOGY OF CHRONIC KIDNEY DISEASE

In a cross-sectional analysis of random samples from the nationally representative Health Survey for England that took place in England in 2003 and in 2009-2010, there were 13,896 adults aged 16 years or older participating. Among 305 participants in 2003 and 322 participants in 2009-2010 with doctor-diagnosed diabetes, the prevalence of chronic kidney disease (CKD) was 17.3% [R15-3134].

In a retrospective, longitudinal study assessing adults with prevalent or incident CKD (identified using estimated glomerular filtration rate readings and/or primary care codes) in the CPRD in 2010, the prevalence of stage 3-5 CKD in 2010 was estimated to be 5.9% (n = 165,942), and the prevalence of mildly impaired eGFR was 21.2% (n = 602,437), being the denominator for prevalence calculations the total CPRD population (n = 2,836,476). The prevalence of diabetes was 19.2% among patients with CKD stage 3-5, 13.2 among patients with mildly impaired eGFR and 6.1% in the general population [R15-3139]

The UK Prospective Diabetes Study (UKPDS) [R03-0585] was a clinical trial designed to evaluate the effects of improved blood glucose control and/or blood pressure control on the incidence of complications in patients with hypertension and newly diagnosed type 2 diabetes. The study included 5,102 participants 25-65 years of age (mean, 53 years), 60% male; 4,031 without albuminuria and 5,032 with normal plasma creatinine at diagnosis. All patients were followed until the trial ended in 1997. The renal outcomes assessed were two measures of albuminuria (micro- and macroalbuminuria) and two measures of renal impairment (reduced creatinine clearance and doubling of baseline plasma creatinine). A reduced glomerular filtration rate was defined as an estimated creatinine clearance ≤ 60 mL/min per 1.73 m2. Altogether, 1,544 of 4,031 patients (38%) developed albuminuria and 1,449 of 5,032 (29%) developed renal impairment over a median of 15 years after diagnosis of type 2 diabetes [R12-1479].

In the same UKPDS study, the absolute risk of renal failure was 1.4 events per 1,000 patient-years among patients with tight control of the diabetes, and 2.3 events per 1,000 patient-years among patients with less tight control of diabetes. The absolute risk of death from renal failure was 0.3 event per 1,000 patient-years among patients with tight control of the diabetes, and 1 event per 1,000 patient-years among patients with less tight control of diabetes [R03-0585].

7.5 EPIDEMIOLOGY OF SEVERE COMPLICATIONS OF URINARY TRACT INFECTION

Untreated UTIs can lead to acute or chronic kidney infections (pyelonephritis), which could permanently damage the kidneys, and lead to urosepsis [R14-5259, R14-5286]. People with T2D are at higher risk for infections than people without diabetes [R12-1100], and women with T2D have a 2-fold higher risk of culture-confirmed UTIs than women without diabetes [R12-1083, R14-5236, R14-5283].

7.5.1 Epidemiology of severe UTI in general population

In a US hospital-based study of the general population, using the 1997 Healthcare Cost and Utilization Project Nationwide Inpatient Sample Release 6, the incidence of hospital discharges for acute pyelonephritis was estimated to be 11.7 per 10,000 persons-years in females (n = 160,848), and 2.4 per 10,000 persons-years in males (n = 30,718) [R14-5258]. Similarly, in another US study, 4,887 enrolees of Group Health Cooperative, based in Seattle, Washington (US), received a diagnosis of acute pyelonephritis from 1997 through 2001. The annual rates of pyelonephritis among women were 12 to 13 outpatient cases per 10,000 population and 3 to 4 inpatient cases per 10,000 population; among males, the annual rates of pyelonephritis were 2 to 3 outpatient cases per 10,000 population and 1 to 2 inpatient cases per 10,000 population [P14-17370].

7.5.2 Epidemiology of severe UTI in the diabetes population

In a review of hospital charts of pyelonephritis (n = 838 cases, 35 among patients with diabetes) and UTI (n = 976 cases, 11 among patients with diabetes) conducted in Canada between 1991 and 1992, the rate of hospitalisations due to pyelonephritis ranged from 66 to 144 per 10,000 person-years among women with diabetes and from 24.7 to 34.3 per 10,000 person-years among men with diabetes, compared with 6.0 to 11.2 per 10,000 person-years among females without diabetes and 1.7 to 10 per 10,000 person-years among males without diabetes [R14-5251].

Another cohort study performed in Denmark evaluated 10,063 individuals from the Danish general population who were participants in The Copenhagen City Heart Study and estimated the risk of hospitalisation caused by any infectious disease between 1991 and 2000. The study reported a total of 314 hospitalisations due to UTI and an incidence rate for pyelonephritis of 158 per 10,000 person-years among patients with diabetes and 41 per 10,000 person-years among patients from the general population [R12-1080].

A cohort study of the computerised medical database of the University Medical Center Utrecht General Practitioners Research Network selected all patients aged 45 years or older with diabetes between 1995 and 2003; the estimated annual incidence rate of pyelonephritis among patients with diabetes was 3 per 1,000 person-years (n = 16 cases), while the rate of outpatient diagnosis of UTI was 101 per 1,000 person-years (n = 2,000 cases) [R12-1105].

In a more recent study performed in the UK CPRD for the period 1990 to 2007, the incidence of outpatient diagnosis of UTI per 1,000 person-years was 46.9 (95% CI, 45.8-48.1) among patients with diabetes and 29.9 (95% CI, 28.9-30.8) among patients without diabetes. The study reported that, over the 1-year follow-up period, 5,967 UTI events were observed among patients with diabetes, and 3,708 UTI events were observed among patients without diabetes. The relative risk of UTI was 1.46 (95% CI, 1.40-1.53) among newly diagnosed patients with diabetes compared with patients without diabetes. Similarly, the relative risk of UTI was 2.08 (95% CI, 1.93-2.24) among patients with previously diagnosed diabetes compared with patients without diabetes. The risk was higher among females, older patients, patients with

previous diagnosis of diabetes, worst level of diabetes control, and a recent history of UTI [R12-5226].

7.5.3 Epidemiology of sepsis in general population and in diabetes patients

The reported incidence rates of sepsis vary by study design and methods. A review of the variations in the incidence and mortality of severe sepsis in the US used four methods or definitions and found that the annual incidence of sepsis ranged from 3 to 10.3 per 1,000 patient-years, depending on the method used [R14-5287].

Among patients with diabetes, the incidence rate of sepsis was evaluated in a cohort study comparing all people with diabetes in Ontario, Canada, between 1999 and 2000, to a matched cohort without diabetes (513,749 patients in each group). The rate of sepsis among the diabetes cohort was 5.4 per 1,000 patient-years (number of cases not reported), and the risk ratio of sepsis vs. the non-diabetes cohort was 2.45 (95% CI, 2.23-2.68) [R10-6632]. The rate of septicaemia among patients with T2D in Denmark in 2004-2012 was 5.5 per 1,000 person-years, with an adjusted rate ratio of 1.60 (95% CI, 1.53-1.67) compared with individuals without diabetes [P16-10382].

For the study size and power calculations, the incidence estimates of the acute pyelonephritis in the population with diabetes were drawn from the study in Netherlands [R12-1105]; estimates from the study performed in UK CPRD [R12-5226] could not be used as it had evaluated all forms of UTI including mild and not resulting in hospitalisation. Incidence rate of UTI hospitalisations reported in Denmark [R12-1080] has been used for the power calculations as well; see Section 9.5.

7.6 EPIDEMIOLOGY OF GENITAL INFECTIONS

Individuals with T2D are at higher risk for infections, including GI, than patients without diabetes [R10-6632, R12-1080, R12-1088, R12-1100, R12-3639, R14-5248], and T2D is a risk factor for vaginal infections and balanitis [R12-2432, R14-5237, R14-5260]. Moreover, patients with diabetes have up to a 4-fold greater risk of dying due to infectious disease than patients without diabetes [R10-6632, R12-1080]. Women with diabetes have increased rates of asymptomatic vaginal carriage of *Candida* species and increased frequency of symptomatic infections [R12-2432, R14-5254].

While assessing the incidence or prevalence of GI using existing databases it is important to remember that the rates may be underestimated, as some of these cases may be self-limiting or may be treated with over-the-counter medications only.

An observational cohort study in the UK evaluated the risk of vaginitis and balanitis among patients with T2D and patients without diabetes in the UK CPRD database using electronic medical records between 1990 and 2007 [R12-3639]. Among females with T2D, there were 1,243 cases of vaginitis: 87.5% recorded as vaginitis/candidiasis, 8.6% vulvitis, 2.2% bacterial vaginosis, and 1.7% related sexually transmitted diseases (STDs). The incidence rate of vaginitis was 21 per 1,000 person-years among patients with T2D and 10.3 per

1,000 person-years among females without diabetes (adjusted relative risk [RR], 1.81; 95% CI, 1.64-2.00). Similarly, there were 592 cases of balanitis among patients with T2D classified as 85.4% balanitis/candidiasis, 8.1% penile/Candida infection, and 6.5% other (no STDs). The incidence rate of balanitis was 8.4 per 1,000 person-years among males with T2D and 2.5 per 1,000 person-years among males without diabetes (RR, 2.85; 95% CI, 2.39-3.39). The incidence rate of vaginitis and balanitis decreased with increasing age for males and females with T2D and females without diabetes, but not for males without diabetes (see Table 2). Among patients with T2D, previous history of GI predisposed to GI, with an adjusted RR of 6.99 (95% CI, 5.97-8.18) for vaginitis and an adjusted RR of 11.22 (95% CI, 9.02-13.97) for balanitis. Also, worse control of diabetes, assessed by measuring glycated haemoglobin (HbA1c), was associated with an increased risk of GI among both treated and untreated patients with T2D (see Table 2).

Table 2 Incidence of vaginitis and balanitis among patients with T2D and patients without diabetes in the UK CPRD, 1990-2007

	Vaginitis (females)		**Balanitis (males)**	
	T2D (N = 62,537, 59,279 patient-years)	**No diabetes** (N = 62,700, 57,844 patient-years)	**T2D** (N = 73,383, 70,088 patient-years)	**No diabetes** (N = 73,220, 67,676 patient-years)
All ages	21 (19.8-22.1)	10.3 (9.5-11.1)	8.4 (7.8-9.1)	2.5 (2.1-2.9)
Age 18-39	~53[1]	~26	~18	~2
Age 40-49	~42	~20	~16	~2
Age 50-59	~24	~11	~9	~2
Age 60-69	~19	~10	~7	~3
Age 70+	~12	~8	~5	~3
Not treated fair control	11.7 (10.0-13.5)		4.6 (3.6-5.6)	
Not treated poor control	38.2 (31.6-44.8)		20.7 (16.6-24.8)	
Treated fair control	15.0 (12.4-17.7)		4.4 (3.1-5.7)	
Not treated poor control	23.9 (20.8-27.0)		11.5 (9.6-13.5)	

T2D = type 2 diabetes mellitus.

Note: Incidence rates (95% CIs) are per 1,000 person-years.

1. Incidence rates with an ~ symbol are approximate estimates derived from the published graph [R12-3639].

Another cohort study conducted in Canada (1999-2000) compared administrative data from all people with diabetes (n = 513,749) in Ontario with a matched cohort of 513,749 of people without diabetes. The rate of GI among male patients with diabetes was 1,340 per

100,000 patients. The rate of GI among female patients with diabetes was 234 per 100,000 patients [R10-6632].

For the study size and power calculations, the incidence estimates of the GI in men and women with diabetes were drawn from the study performed in UK CPRD [R12-3639], as it is most relevant for the study setting; see Section 9.5.

7.7 EPIDEMIOLOGY OF DIABETIC KETOACIDOSIS

DKA is the most common hyperglycemic emergency in patients with diabetes mellitus, leading to more than 100,000 hospital admissions each year in the US and comprising 4% to 9% of all hospital discharge summaries among patients with diabetes mellitus. In Denmark, the annual incidence of DKA in the general population was estimated at 12.9 per 100,000 [R16-4879]. In 20% to 30% of the cases of DKA, DKA may be the initial manifestation of diabetes, particularly for T1D [R16-1981]. Although DKA occurs most frequently in patients with T1D, it can also occur in patients with T2D [R15-2053]. The incidence of DKA is difficult to establish, ranging from 4.6 to 8 episodes per 1,000 patients with diabetes in population-based studies from the US [R16-1373]. In England, according to the National Diabetic Audit, the 5-year prevalence of recorded DKA was over 12% in people with T1D and less than 1% in people with T2D. DKA occurred in 3.9% of the people with T1D and 0.48% of all people with diabetes (T1D and T2D) during 2009-2010; the occurrence of DKA in people with T2D was not reported [R16-1374]. In a recent UK study based on the CPRD 1998-2013, the incidence of hospital admission for DKA in T1D was 36 per 1,000 person-years, and in T2D was 0.9 per 1,000 person-years [R18-1217].

The incidence of DKA among patients with T2D is not well established. A review of the medical records of all adult patients admitted to the medical intensive care unit of a medical centre in an Arizona (US) identified 226 patients with DKA: 47% had T1D, 26% had T2D, and 27% had DKA as the initial manifestation of diabetes (type not specified) [R16-1375]. A retrospective population-based study performed in Sweden on data from 1997 through 2000 was performed to determine the occurrence of DKA in adult patients with T1D and T2D [R14-3272]. All adult patients with severe hyperglycaemia or suspected DKA admitted to The Umeå University Hospital with a diagnosis code of DKA (ICD-10 codes: E10.0, E10.1, E11.0 and E11.1) were included. The average annual incidence rate for DKA was 5.9 per 100,000 adult inhabitants. The annual incidence rate for DKA in patients with diabetes was 50 per 100,000 adult patients with T2D and 1,585 per 100,000 adult patients with T1D. A total of 25 patients developed DKA: 8 (32%) had T2D, while 17 (68%) had T1D [R14-3272]. In a similar study performed in an Arizona (US) hospital, 226 patients with DKA were identified: 47% had T1D, 26% had T2D, and 27% had DKA as the initial manifestation of diabetes (type not yet specified) [R16-1375].

The incidence of DKA among patients with T2D using SGLT2 inhibitors in the context of routine clinical practice is not well understood. The low incidence of DKA in patients with T2D did not allow for conclusions on a potential association with SGLT2 inhibitor use. A cohort study in the Truven MarketScan database, published in 2017, evaluated the risk of DKA among patients with diabetes mellitus who newly initiated either an SGLT2 inhibitor or

a DPP-4 inhibitor and found that the risk of DKA within 180 days of initiating an SGLT2 inhibitor was 2.2 times greater than the risk for those initiating a DPP-4 inhibitor (hazard ratio after propensity score matching, 2.2; 95% CI, 1.4-3.6). However, it is important to note that this study, based on a limited number of cases, only reported data on SGLT2 inhibitors as a class and no drug-specific data were presented [R18-1215].

In 2016, a 5-year enhanced pharmacovigilance surveillance study of DKA (1245.146) was initiated by BI upon request by the US FDA. The first interim report for the DKA surveillance study was submitted and assessed by the EMA in 2017 [c19231123-01].

7.8 EPIDEMIOLOGY OF DIABETES AND ANTIDIABETIC TREATMENT PATTERNS

In the UK, the prevalence of diabetes has increased 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.

In Denmark, 252,750 people were diagnosed with diabetes as of 01 January 2016, which corresponds to 4.4% of the population. The number of people with diabetes (80% T2D) has more than doubled from 2000 to 2016. In 2016, about 16,300 new cases of diabetes were diagnosed in Denmark. About 60,000 Danes do not yet know that they have T2D, and another 300,000 are estimated to have precursors for T2D (prediabetes) [R18-1554].

In the US in 2015, 30.3 million Americans had diabetes (9.4% of the population); most had T2D, and 1.25 million (~4.3%) had T1D. Approximately 1.5 million Americans are diagnosed with diabetes every year. Of the 30.3 million adults with diabetes, 23.1 million were diagnosed and 7.2 million were undiagnosed; 84.1 million Americans aged 18 years or older had prediabetes [R18-1560]. From 1988-1994 to 2011-2012, the prevalence of diabetes increased significantly in the US population. The estimated prevalence of diabetes among US adults was 12% to 14% from 2011-2012 [R18-1558].

During the period 1996-2005, a change in oral GLD 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].

A change in the use of GLDs in the last decade was also observed in a study performed in Denmark using data from medstat.dk, a publicly accessible webpage from the Danish Health Data Authority that provides aggregate statistics on sale of pharmaceuticals in Denmark, based on individual-level data. During the period 1999 through 2014, the annual prevalence of GLD users increased more than twofold, from 19 per 1,000 inhabitants (n = 98,362) in 1999 to 41 per 1,000 (n = 233,230) in 2014. The most frequently prescribed GLDs were metformin (72% of all persons using GLDs), followed by insulin (33%) and sulfonylureas (15%). Use of sulfonylureas decreased since 2007. In contrast, prescribing of DPP-4 inhibitors increased steadily since their introduction in 2007; in 2014, 3 per 1,000 used a combination pill of metformin and a DPP-4 inhibitor and 3 per 1,000 redeemed a prescription of a DPP-4 inhibitor noncombination pill. SGLT-2 inhibitor use reached 0.8 per 1,000 within the third year after introduction to the Danish market (123 using a combination pill with metformin and SGLT-2 inhibitor and 4,398 using a noncombination pill). The prevalence of thiazolidinedione users decreased from 0.7 per 1,000 in 2007 to 0.03 per 1,000 in 2014. Prescriptions of alpha-glucosidase inhibitors and meglitinides remained low over time. Despite a steadily increasing absolute number of insulin users, the proportion of all GLD users who used insulin declined from 41% in 1999 to 33% in 2014[R18-1559].

In the US, among participants with self-reported physician diagnosis of diabetes in the 2000 through 2009 National Health Interview Surveys, 38% reported use of antidiabetic monotherapy, 52% use of combined antidiabetic therapy, and 10% reported no drug use for diabetes. Among users of monotherapy, the most frequent medication was insulin (33.0% to 37.2%), followed by metformin (27.3% to 33.3%) and sulfonylureas (21.9% to 25.9%). The most common combination therapies were metformin with sulfonylureas (20.1% to 21.9%), followed by insulin with any oral GLD (15.3% to 16.9%); thiazolidinediones with any other oral GLD (13.1% to 13.6%); metformin with sulfonylurea and thiazolidinediones (6.8% to 7.8%); and metformin, sulfonylureas, and insulin (5.0% to 6.9%). Any other combination of GLDs was used by 36.3% to 36.5% of the patients using combination therapies [R18-1557].

8. RESEARCH QUESTION AND OBJECTIVES

The primary research question is to evaluate whether, among patients with T2D, initiation of empagliflozin changes the adjusted incidence of outcomes compared with initiation of a DPP-4 inhibitor.

The primary objectives of the study are as follows:

- To estimate the adjusted IRR of hospitalisation, emergency department (ED) visit, or specialist visit for ALI in patients with no predisposing conditions, by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of hospitalisation, ED visit, or specialist visit for AKI by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of severe complications of UTIs (pyelonephritis and urosepsis) (inpatient and outpatient) by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of GI (inpatient and outpatient) by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of hospitalisation or ED visit for DKA by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.

The secondary objectives of the study are as follows:

- To estimate the adjusted IRR of hospitalisation, ED visit, or specialist visit for ALI in patients with or without predisposing conditions, by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of CKD (inpatient and outpatient), by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate the adjusted IRR of severe GI by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.
- To estimate adjusted IRRs for each of the primary outcomes (ALI in patients with no predisposing conditions, AKI, UTI, GI, and DKA) and secondary outcomes (ALI in patients with or without predisposing conditions, CKD, severe GI)—stratified by categories of insulin use at the index date, age, sex, and other variables of interest such as diabetes control—by comparing patients initiating empagliflozin with patients initiating a DPP-4 inhibitor, among patients with T2D.

- To estimate adjusted incidence rates for each of the primary outcomes (ALI in patients with no predisposing conditions, AKI, UTI, GI, and DKA) and secondary outcomes (ALI in patients with or without predisposing conditions, CKD, severe GI)—overall and stratified by categories of insulin use at the index date, age, sex, and other variables of interest such as diabetes control—among patients with T2D initiating empagliflozin or a DPP-4 inhibitor.

9. RESEARCH METHODS

9.1 STUDY DESIGN

An observational cohort study will be conducted in the CPRD in the UK, and for the evaluation of the rarest outcomes (ALI, AKI, and DKA) also in the Danish Population Registries in Denmark, and in the HealthCore Integrated Research DatabaseSM (HIRD) in the US. The study will use a new-user (also known as incident-user) design and will compare new users of empagliflozin with new users of DPP-4 inhibitors.

The new-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 starting empagliflozin will be required to be new users, defined as having no exposure to empagliflozin, another SGLT2 inhibitor drug, or a DPP-4 inhibitor during the 12 months before or at the index date. Patients starting a DPP-4 inhibitor will be required to have no exposure to a DPP-4 inhibitor, empagliflozin, or another SGLT2 inhibitor during the 12 months before or at the index date (see further details in Section 9.2.3,New user definition) [R13-1120, R14-4378].

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 naive. 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 GLD 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. According to newer guidelines, double and even triple therapy may be started at diabetes debut if there is poor glucose control [P18-01920]. In this scenario, if analysis evaluating confounding and interaction suggest that the effect of drug exposure varies between patients on second- vs. third-line therapy, and if the number of events is sufficient, additional analysis will be done to achieve a fair comparison [R13-1120]. In the same way, patients starting a combination of empagliflozin and metformin (whether a fixed-dose or free combination) will be compared with patients starting a combination of a DPP-4 inhibitor and metformin.

A cohort design will allow direct estimation of the absolute rates, rate differences, and relative risk or hazard ratios of multiple outcomes of interest among new users of empagliflozin compared with new users of a DPP-4 inhibitor. A cohort study design will also allow accurate chronologic confounder assessment and assessment of the outcomes at

multiple time points. The covariate information will be assessed during the time preceding treatment initiation and will include all historical information available for each patient. Follow-up will start the day after treatment initiation. In the context of data sources such as the CPRD, Danish Population Registries, and HIRD, 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].

DPP-4 inhibitors have been selected as a comparator group for several reasons. First, the National Institute for Health and Care Excellence (NICE) appraisal of dapagliflozin (an SGLT2 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, SGLT2 inhibitors, and thiazolidinediones (pioglitazone) have similar indications and target population, while dual therapy with GLP-1 (glucagon-like peptide-1) analogues has a restricted target population [P14-17374]. Finally, the use of thiazolidinediones has decreased in recent years, given increasing concerns about their safety, 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 (see Section 9.4) [R14-5244, R14-5249].

Propensity scores will be estimated for each cohort member based on information before or at the index date. Propensity scores will incorporate measured potential predictors of the outcome as independent variables and exposure group status as the dependent variable. Propensity scores will be used to minimise confounding.

9.2 SETTING

9.2.1 Study population

Empagliflozin is expected to be prescribed mainly by general practitioners (GPs) and specialists, and most of the follow-up prescriptions (for chronic treatment) will also be issued by GPs or primary care physicians. Thus, the selected study populations will be adult patients identified using data on prescriptions written by GPs in the CPRD, dispensings in community pharmacies in Denmark, and health insurance claims for outpatient medication dispensings in the HIRD during the study period.

For ascertainment of hospitalisation-related study outcomes, hospital data from each data source will be obtained for more complete information on the study outcomes. For the CPRD, linkage with Hospital Episode Statistics (HES) will be available for a subset of available patients. Among patients not linkable to HES in the CPRD, GP records of hospitalisations will be used to identify hospitalisation-related study outcomes. For Denmark, ascertainment of hospitalisation-related study outcomes will be based on data from inpatient hospital discharge diagnoses that are available for all patients across Denmark. For the HIRD, ascertainment of hospitalisation-related study outcomes will be based on data from inpatient hospital discharge diagnoses that are available for all patients in the HIRD database.

The study population will include eligible adult male and female patients with T2D initiating treatment with empagliflozin or initiating a DPP-4 inhibitor. Type 2 diabetes mellitus will be identified based on the type and availability of data in each data source. In the CPRD, patients with T2D will be identified based on a combination of outpatient and inpatient codes for T2D, T1D, diabetes unspecified, and GP prescriptions for insulin and non-insulin GLD using an adapted definition from Holden et al. [R13-3433]. In Denmark, patients with T2D will be identified based on a combination of community pharmacy prescription data, hospital codes, and primary care procedure codes [R18-1213]. In the HIRD, patients with T2D will be identified based on outpatient and inpatient health care claims with ICD-9-CM or ICD-10-CM diagnosis codes for T2D, as well as pharmacy dispensing claims for antidiabetic medications. These criteria are based on published studies in which similar algorithms had PPVs greater than 85% for identifying T2D in health care claims data [R16-3197, R18-3576, R18-3667].

9.2.2 Study period

The study period will start in August 2014, the month of empagliflozin launch in the UK, Denmark, and the US. Based on interim counts, the study end date (originally August 2017) will be extended through August 2019.

9.2.3 New user definition

New users will be patients with a prescription/dispensing for empagliflozin or a DPP-4 inhibitor during the study period and no prescription/dispensing for empagliflozin, another SGLT2 inhibitor drug, or a DPP-4 inhibitor during the previous 12 months. Patients will be allowed to be new users of a study medication only once during the study period but will be allowed to be new users of the other study medication if they fulfil the inclusion/exclusion criteria.

9.2.4 Index prescription definition

The index prescription will be the first prescription for the study medication of interest that fulfils the definition of new user during the study period. Index prescriptions/dispensings of the study drugs include the single study drugs or fixed-dose combinations of the study drugs with metformin, when available.

9.2.5 Index date

The index date will be defined as the date on which each identified new user receives the index prescription for empagliflozin or DPP-4 inhibitor comparator.

9.2.6 Baseline and lookback period

To characterise the empagliflozin and DPP-4 inhibitor cohorts at the time of study drug initiation, all information available during the lookback (pre-index) time period will be collected. The lookback time period is defined as the time period ending on the index date, i.e., will include the index date, unless otherwise specified. As all cohort members are

required by inclusion criteria to have at least 12 months of data before the index date (baseline period), the lookback period will include at least 365 days during which covariates can be evaluated. For some of the cohort members, more data on covariates might be available beyond 365 days, and all available information will be considered for covariate classification related to diabetes, diabetes medications, and concomitant chronic conditions. Nevertheless, for comedications (i.e., for diseases other than diabetes) the lookback time period will be limited to 180 days before or at the index date. Lookback time periods for a small number of specific covariables may be adapted in each data source, e.g., to define body mass index (BMI) in the CPRD, the closest data in the 3 years before or at the index date will be used [R15-4888, R16-1231].

If the distribution of the duration of lookback time period is different among empagliflozin and DPP-4 inhibitor groups, categories of lookback 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.

9.2.7 Inclusion criteria

All patients will be required to meet all of the following criteria:

- Be aged 18 or more years at the index date.
- Have at least 12 months of continuous registration before or at the index date. In the CPRD this means registration in a primary care practice with up-to-standard data. In Denmark, this means residency in the country. In the HIRD, this means enrolment in the health care plan.
- Have T2D ever before or at the index date: the algorithm to identify patients with T2D will be adapted to the type of data available in each data source. This algorithm may include medication codes and will be described in the statistical epidemiological analysis plan.

The empagliflozin-exposed population must also meet the following criteria:

- Have at least one prescription for empagliflozin or fixed-dose combination of empagliflozin with metformin, with or without treatment with another GLD.
- Have no prior prescriptions of any SGLT2 inhibitor (including empagliflozin) or a DPP-4 inhibitor alone or in fixed-dose combination during the previous 12 months.
- The population exposed to a DPP-4 inhibitor must meet the following criteria:
- Have at least one prescription for a DPP-4 inhibitor or a fixed-dose combination of a DPP-4 inhibitor with metformin with or without treatment with other GLDs.
- Have no prior prescriptions of a DPP-4 inhibitor or an SGLT2 inhibitor (including empagliflozin) alone or in fixed-dose combination during the previous 12 months.

9.2.8 Exclusion criteria

Patients with a confirmed diagnosis of T1D before or at the index date will be excluded from the study. The final algorithm to identify patients with T1D will be adapted to the type of data available in each data source. This algorithm may include a combination of diagnosis and drug prescription codes and will be described in the statistical epidemiological analysis plan.

Patients prescribed/dispensed combinations of SGLT2 inhibitors with DPP-4 inhibitors at the index date (as fixed-dose combinations such as Glyxambi® [empagliflozin and linagliptin], or as non–fixed-dose combinations of the two individual medications prescribed on the same date) will be excluded.

9.2.8.1 Exclusion criteria by outcome of interest

Different exclusion criteria will be applied to generate sets of cohorts for the analysis of the outcomes of interest.

For analysis of the primary outcome "ALI in patients with no predisposing conditions," the following set of exclusion criteria will be applied:

1. A diagnosis of ALI is recorded any time before or at the index date (i.e., during the available lookback time). Annex 3 contains ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) codes for these exclusion conditions (Annex 3 Table 1).
2. Pregnancy at the index date, because pregnancy can be associated with an increased risk of hepatic injury. Specific liver disorders associated with pregnancy include preeclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hyperemesis gravidarum [P15-00346]. Pregnancy will be identified through diagnosis codes compatible with initiation and/or termination of pregnancy, and duration of pregnancy will be estimated through specific time windows set up around the date of diagnosis.
3. A diagnosis of the following chronic conditions recorded any time before or at the index date—Annex 3 contains ICD-10 codes (Annex 3 Table 1):
    - Chronic liver disease
    - Chronic alcoholism
    - Chronic infectious hepatitis
    - Chronic disease involving the liver or causing hyperbilirubinaemia
    - Chronic cholelithiasis and cholecystitis
    - Intra- or extrahepatic biliary obstruction
    - Chronic pancreatic disease
    - Primary or secondary hepatic, biliary, or pancreatic cancer
    - Congestive heart failure

For analysis of the ALI secondary outcome, patients will be excluded if they meet any of the following criteria:

1. A diagnosis of any of the following acute conditions within 6 months before or at the index date—Annex 3 contains ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) codes (Annex 3 Table 1) for these exclusion conditions:
   - ALI
   - Acute infectious hepatitis
   - Acute cholelithiasis and cholecystitis
   - Acute intra- or extrahepatic biliary obstruction
   - Acute pancreatic disease
   - Decompensated congestive heart failure (i.e., hospitalisation)
2. Pregnancy at the index date

For analysis of the AKI primary outcome, patients will be excluded if they meet any of the following criteria:

- A diagnosis of AKI is recorded within 6 months before or at the index date. Annex 3 contains ICD-10 codes (Annex 3 Table 2) for AKI.
- A diagnosis of CKD is recorded any time before or at the index date (i.e., during the available lookback time). Annex 3 contains ICD-10 codes (Annex 3 Table 3) for CKD.

For the analysis of CKD (kidney secondary outcome), patients will be excluded if they meet the following criterion:

- A diagnosis of CKD is recorded any time before or at the index date (i.e., during the available lookback time). Annex 3 contains ICD-10 codes (Annex 3 Table 3) for CKD.

For analysis of the UTI outcome, patients will be excluded if they meet any of the following criteria:

- The patient experienced chronic or acute pyelonephritis within the 6 months before or at the index date (see Annex 3 Table 4 for ICD-10 codes for pyelonephritis).

9.2.9 Follow-up of subjects

Follow-up will start the day after the index date, which will be the date of the first prescription for empagliflozin or a DPP-4 inhibitor.

For the analysis of each outcome, follow-up time in a given cohort in a given exposure category for each patient will end at whichever of the following dates occurs first:

- The date of the outcome event; a diagnosis of ALI, AKI or CKD, severe complications of UTI, GI, or DKA.
- The date of death.
- The date of study end.

- The date of transfer out of the practice or last collection date of a practice in the CPRD, emigration date in Denmark, or end of health plan eligibility in the HIRD.
- The date that outcome-specific exclusion criteria are met (see exclusion criteria in Section 9.2.8). Exclusion criteria will be specific for primary and secondary outcomes, e.g., for the ALI primary outcome, criteria such as chronic liver disease occurring any time before or at the index date.
- The end date of the first continuous treatment of the index drug (empagliflozin or DPP-4 inhibitor) plus a defined grace period (see also Section 9.3.1.1):
  - Current use analysis (main analysis): 30 days after the end of the last prescription's supply.
  - Recent use analysis: 90 days after the end of current use, i.e., 120 days after the end of the last prescription's supply.
  - Sensitivity analysis: the earliest of 90 days after the end of the last prescription's supply or the date on which a new treatment episode starts with the same index drug.
  - Intention-to-treat sensitivity analysis: follow-up will not be censored at the end of the first continuous treatment of the index drug, i.e., this criterion will not be applied for this analysis.
- Within each exposure cohort, the date on which a new treatment episode with any of the other index drugs or other SGLT2 inhibitors starts. This criterion will not be applied for the intention-to-treat sensitivity analysis.

Patients will be able to re-enter another study cohort (not the original cohort) if they fulfil inclusion and exclusion criteria, including no prior use of any of the study medications during the previous 12 months. Follow-up will not be censored if oral or injectable GLDs other than the index drugs are prescribed in addition to empagliflozin or a DPP-4 inhibitor after the index date. All censoring criteria, except otherwise specified, will be applied in the intention-to-treat analysis.

9.3 VARIABLES

9.3.1 Exposures

For this study, eligible patients will be identified from prescriptions/dispensings of the study medications of interest listed in the data sources included in the study.

Empagliflozin alone (Jardiance, ATC code A10BX12 before 2016, A10BK03 beginning in 2017) or in fixed-dose combination with metformin hydrochloride (Synjardy, ATC code A10BD20) will be the study drug of interest. The oral GLDs currently suggested as comparators are DPP-4 inhibitors (see Sections 9.1 and 9.4). Empagliflozin in fixed-dose combination with linagliptin, a DPP-4 inhibitor (ATC code A10BD19) will not be included as a study medication group. Currently, the following DPP-4 inhibitors, as available in each country, will be comparator medications for study purposes:

- Sitagliptin: ATC code A10BH01
- Vildagliptin: ATC code A10BH02

- Saxagliptin: ATC code A10BH03
- Alogliptin: ATC code A10BH04
- Linagliptin: ATC code A10BH05
- Sitagliptin and metformin hydrochloride: ATC code A10BD07
- Vildagliptin and metformin hydrochloride: ATC code A10BD08
- Saxagliptin and metformin hydrochloride: ATC code A10BD10
- Alogliptin and metformin hydrochloride: ATC code A10BD13
- Linagliptin and metformin hydrochloride: ATC code A10BD11

Users of SGLT2 inhibitors in fixed-dose combination with DPP-4 inhibitors, such as Glyxambi (empagliflozin/linagliptin), will not be included in the study. If additional DPP-4 inhibitor drugs are marketed in the future in the UK, Denmark, or the US during the study period, they will also be considered to be members of the comparator group.

In clinical practice, new users of index drugs (empagliflozin and DPP-4 inhibitors) may be prescribed the drug in the context of (1) adding it to an existing GLD regimen as double or triple therapy or (2) switching from one GLD to the study drug as monotherapy or combination therapy. Relatively few patients with T2D will be starting DPP-4 inhibitors naive to other GLDs since monotherapy indication for these drugs is restricted to patients intolerant to metformin and sulfonylureas. Thus, new users of the study drugs would usually be (1) switching from monotherapy with another GLD to monotherapy with a study drug, (2) switching from dual or triple therapy with another GLD to dual or triple therapy with a study drug and other GLDs, or (3) adding a study drug to therapy with one or two other GLDs to become patients on dual or triple therapy. However, new users of the study drugs may also have started SGLT2 inhibitors or DPP-4 inhibitors from the start of diabetes therapy if hyperglycaemia was substantial at diabetes diagnosis, i.e., new users may be combination therapy users from the beginning. This approach is consistent with the new-user design described previously in Section 9.1 [R13-1120]. Information on whether patients received prior GLD therapy or if they were “added on” or “switched to” empagliflozin or a DPP-4 inhibitor at the time of inclusion in the study will be collected. Patients will be classified according to their treatment complexity as receiving mono vs. dual vs. triple therapy.

9.3.1.1 Exposure and time at risk

For this study, it will be assumed that the risk of ALI, AKI, severe complications of UTI, DKA, and GI related to use of empagliflozin or DPP-4 inhibitors (the index drugs) increases at the beginning of therapy, is maintained at an increased level for the duration of treatment and decreases gradually to the background risk once treatment is stopped.

Only use at the first continuous treatment will be considered, defined as having consecutive prescriptions/dispensings separated by 30 days or less. Therefore, the risk/exposure time window for each new user of an index drug—empagliflozin or DPP-4 inhibitor—will be categorised into two mutually exclusive categories of risk, as follows (Figure 1):

- Current use (current time at risk): The risk/exposure window for current use starts on the date of the prescription and ends 30 days after end of supply. The main analysis is

based on current use. This time at risk will be used for comparisons and estimation of IRRs in the study main analysis.

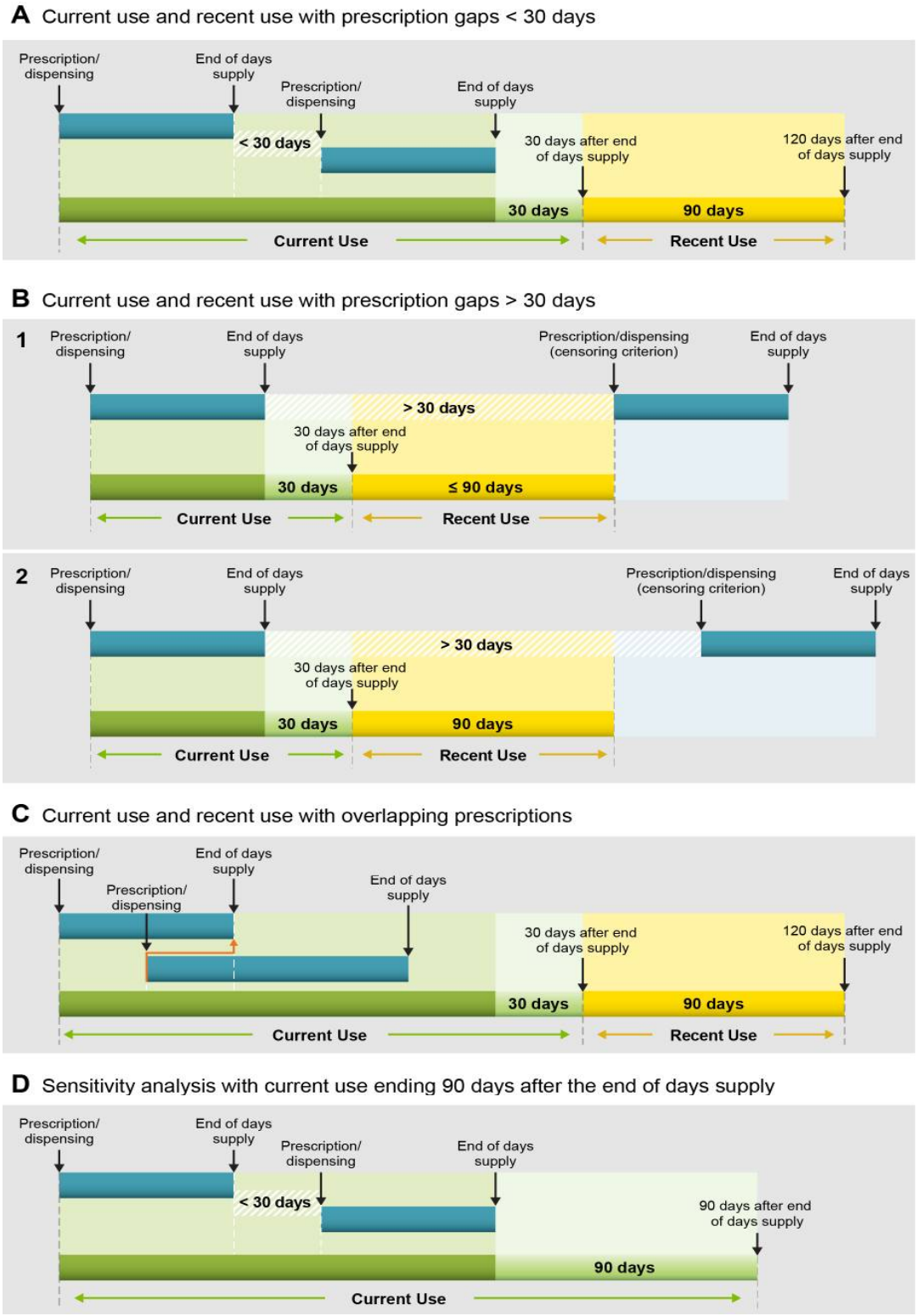
- Recent use (recent time at risk): The risk/exposure window for recent use starts at the end of current use (30 days after end of supply) and ends the earliest of 90 days later (which is 120 days after end of supply) or at any censoring event (including a new prescription for a study medication). The time at risk for recent use will be used in an additional analysis (see Section 9.7.3).

Overlapping time at risk from current use for consecutive prescriptions/dispensings of the index medication will be concatenated, with the overlapping time added at the end of the concatenated prescription. For consecutive prescriptions/dispensings of the index medication separated by gaps of 30 days or less, time at risk from current use will include the gaps between prescriptions/dispensings.

In the CPRD and in the US, since most oral GLD prescriptions/dispensings are supplied for 30 days, for the majority of cohort members, the risk window defining current use will end 60 days after the last prescription date. By adding 30 days to the end of the days' supply, a delayed increase in risk for ALI, AKI, severe complications of UTI, DKA, or GI after termination of the index drugs can be detected. To define duration of use similarly in the three data sources, the carry-over period in the main analysis will also be 30 days in Denmark. However, in Denmark, some medications are supplied for 90 days; for this reason, in a sensitivity analysis, the risk/exposure window for current use starts on the date of the prescription and ends 90 days after the end of supply. This sensitivity assessment will allow exploration of any further potential delay in effect. A 90-day period was selected because this time is long enough to account for non-adherence and extended use of the discontinued index drug and a delay in effect.

Duration of exposure will be based on the duration of current use. Categories of duration will be defined based on available data.
Dose will be the dosage at index date. When the dose is missing, the dose will be estimated from the available recorded information (e.g., strength, number of units, amount of drug prescribed). Evaluation of different doses of empagliflozin will be performed if variation in the dose used by the empagliflozin cohort is observed and an adequate number of events occur within the different daily dose categories.

Figure 1 Exposure Definition



Note: End of supply will be estimated according to prescription instructions in the CPRD or based on available information on the duration of dispensings (e.g., number of packages bought, strength, number of pills) in Denmark and the HIRD.

9.3.2 Study outcomes

The primary outcomes of interest for this study are ALI in patients with no predisposing conditions, AKI, severe complications of UTI, DKA, and occurrence of GI. Secondary outcomes are ALI in patients with or without predisposing conditions, CKD, and severe GI.

Scarce literature exists on the validity of algorithms used to identify GI and hospitalisations due to severe complications of UTI or to DKA. Validation of the algorithms used to identify cases to confirm diagnosis and date of the event will be implemented for all primary and secondary outcomes and for ALI and AKI non-specialist outpatient cases (CPRD, US). Each outcome-specific section that follows includes a case definition to be used for validation purposes and the algorithm of codes that will be used for case identification. The final section describes the common validation process that will be followed for all outcomes (see Section 9.3.2.9).

9.3.2.1 ALI: Hospitalisation, emergency department (ED) visit, or specialist visit for ALI in patients with no predisposing conditions, primary outcome

- Outcome type: primary
- Secondary outcomes within this section: yes (hospitalisation, ED visit, or specialist visit for ALI in patients with or without predisposing conditions)
- Further outcomes within this section: none
- Outcome name: ALI in patients with no predisposing conditions
- Time frame: up to 5 years
- Safety issue: no

Acute liver injury has been defined in terms of an elevation in the serum concentration of ALT or AST, conjugated bilirubin, or alkaline phosphatase (ALP). It has been considered that elevations of ALT/AST are indicators of liver injury, whereas increases of conjugated bilirubin are measures of overall liver function. Liver injury alone may not lead to clinically significant liver damage, whereas impaired liver function is a marker of severe drug-induced hepatotoxicity. Thus, a combined elevation of ALT or AST and conjugated bilirubin without evidence of intra- or extra-biliary obstruction (i.e., no significant elevation of ALP) could be used to define potentially clinically significant elevations of serum liver enzyme levels [P06-02059]. The concept of combining markers of liver injury and function evolved from the observation of Hyman Zimmerman [R05-1093] that “drug-induced hepatocellular jaundice is a serious lesion.” Zimmerman noted that the combination of pure hepatocellular injury (ALT elevation without much ALP elevation) and jaundice among patients with drug-induced liver injury had a poor prognosis, with a mortality of 10% to 15% [R14-5256, R05-1093, P14-17375, P09-12413]. This observation is referred to as “Hy’s Law” by the FDA and has been used by the FDA over the years to assess the potential for a drug to cause severe liver injury—that is, irreversible liver failure that is fatal or requires liver transplantation [P09-12413, P14-17375, R14-5256].

According to the international Expert Working Group on drug-induced liver injury [R14-1933], ALI is defined as abnormal liver function test as summarised in Table 3. According to the Working Group, persistent drug-induced liver injury is defined as evidence of continued liver injury more than 3 months after hepatocellular or mixed liver injury and more than 6 months after cholestatic liver injury; increases of these parameters for more than 1 year are compatible with chronic liver injury.

Table 3 Clinical criteria for "severe" or "clinically significant" acute liver injury

Endpoint	Definition
Clinically significant ALI[1]	All the following criteria[2]: • ALT or AST ≥ 5 × ULN, *or* • ALT or AST ≥ 3 × ULN *and* total bilirubin ≥ 2 × ULN, *or* • ALP ≥ 2 × ULN

ALI = acute liver injury; ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; ULN = upper limit of normal range.

1. Sources: Aithal et al. (2011) [R14-1933].
2. For the secondary endpoint, where patients with predisposing conditions are also included, if a patient has previous liver disease, ULN is replaced by the mean baseline values obtained before exposure to the suspect drug, and the changes should be proportionate to this modified baseline (i.e., 5 x baseline for ALT, 2 x baseline for ALP, and 2 x baseline for total bilirubin with associated 3 x baseline elevation in ALT).

9.3.2.1.1 Validation case definition

In this study, a case of ALI is defined as any person with a recorded diagnosis compatible with "severe" or "clinically significant" ALI who meets the criteria recommended by the international Expert Working Group on drug-induced liver injury [R14-1933] (Table 3).

In addition, among those ALI cases identified through diagnosis codes and that will undergo validation, the number of cases that had elevated liver enzymes ALT and/or AST ≥ 3 x ULN but < 5 ULN and therefore did not fulfil the Aithal et al. criteria [R14-1933] will also be described.

9.3.2.1.2 Case identification

Potential cases of liver injury will be identified by the following process:

- In the CPRD, potential cases will be identified through primary care codes (currently Read codes but may also be SNOMED codes and local EMIS codes in the future) and liver function test results associated with codes for hospitalisation, ED visit, or specialist visit. In HES data, potential cases will be identified through hospital discharge ICD-10 codes suggestive of ALI. No ED visit information is available in HES; ED visits will be identified from the CPRD primary care records.

- In Denmark, potential cases will be identified through hospital discharge ICD-10 codes, ED visit ICD-10 codes, or hospital outpatient clinic ICD-10 codes suggestive of ALI.
- In the HIRD, potential cases will be identified through hospital claims, ED visit claims, or specialist claims with an ICD-10-CM code and/or laboratory results suggestive of ALI.

Annex 4 contains a preliminary list of ICD-10 codes suggestive of ALI (Annex 4 Table 1). For the assessment of this primary outcome, predisposing exclusion criteria will be applied to the cohorts as described in Section 9.2.8.1. Follow-up will be censored at the occurrence of any exclusion criteria for this cohort, e.g., any predisposing condition such as chronic liver disease, as described in Section 9.2.9.

- ALI sensitivity, all cases: a sensitivity analysis will be performed, also including outpatient cases of ALI in the data sources where primary care data are available (the CPRD and HIRD). The exclusion and censoring conditions will be the same as for the primary outcome.
- ALI sensitivity, based on results from validation: another sensitivity analysis will repeat the primary outcome analysis for ALI, including only confirmed cases; however, when less than 70% of the cases are validated, other methods will be explored, e.g., correct the IRRs for the positive predictive value (PPV) [R18-1561].

9.3.2.2 ALI: Hospitalisation, ED visit, or specialist visit for ALI in patients with or without predisposing conditions, secondary outcome

- Outcome type: secondary
- Secondary outcome: none
- Further outcomes within this section: none
- Outcome name: ALI in patients with or without predisposing conditions
- Time frame: up to 5 years
- Safety issue: no

For the assessment of this secondary outcome, ALI in the 6 months before or at the index date and pregnancy at the index date will be used, and exclusion criteria will be those described in Section 9.2.8.1. Follow-up will be censored by the occurrence of the outcome (ALI) or any exclusion criteria applied to the selection of this cohort, such as pregnancy exclusion criteria, but not by the occurrence of an exclusion criterion used for the primary outcome, e.g., any predisposing condition such as chronic liver disease, as described in Section 9.2.9.

9.3.2.3 Hospitalisation, ED visit, or specialist visit for AKI, primary outcome

- Outcome type: primary
- Secondary outcomes within this section: Yes (CKD)
- Further outcomes within this section: none
- Outcome name: AKI
- Time frame: up to 5 years

- Safety issue: no

Multiple definitions of AKI have been used in clinical practice and research.

- In 2002, an international consensus conference of the Acute Dialysis Quality Initiative (ADQI) proposed the RIFLE criteria for defining and classifying AKI [R14-5242]. The RIFLE acronym indicates Risk of renal dysfunction, Injury to the kidney, Failure of kidney function, Loss of kidney function, and End-stage kidney disease. The RIFLE criteria are based on three levels of renal dysfunction (risk, injury, and failure) and two clinical outcomes (loss of function and end-stage renal disease). The RIFLE classification includes separate criteria for serum creatinine, glomerular filtration rate (GFR), and urine output. The serum creatinine (SCr) and GFR criteria are based on changes from baseline values. An increase of 1.5 times the baseline SCr indicates risk of kidney dysfunction, 2 times the baseline SCr indicates injury to the kidney, and 3 times the baseline SCr indicates failure of kidney function. A decrease of > 25% to ≤ 50% of the baseline GFR indicates risk of kidney dysfunction, a decrease of > 50% and ≤ 75% of the baseline indicates injury to the kidney, and a decrease of > 75% of the baseline indicates failure of kidney function.
- According to the most recent Kidney Disease: Improving Global Outcomes (KDIGO) guideline on AKI [R13-4387], AKI is defined as any of the following (Not Graded): an increase in SCr of by ≥ 0.3 mg/dL (≥ 26.5 µmol/l) within 48 hours; or an increase in SCr to ≥ 1.5 times baseline, which is known or presumed to have occurred within the prior 7 days; or a urine volume < 0.5 mL/kg/h for 6 hours.

The definition of AKI in epidemiologic studies has been based on absolute increases of serum creatinine from normal values (1.7 or 2 times the ULN) or changes from baseline (20% to 50%) or both [P02-05226, P05-04032, P14-17372, P14-17376, P14-17377].

9.3.2.3.1 Validation case definition

In this study, a case of AKI is defined as any person with AKI according to the following criteria:

- At least a 2-fold increase in serum creatinine from the lowest baseline value recorded at any time before or at the index date, and the value is above the ULN; or
- An increase in serum creatinine to at least 2 times the ULN in the absence of a recorded baseline value; and
- Absence of a recorded diagnosis of chronic renal failure (defined as CKD stage 3 or higher) at any time before or at the index date.

9.3.2.3.2 Case identification

Potential cases of hospitalisation, ED visit, or specialist visit for AKI will be identified by the following procedure:

- In the CPRD, potential cases will be identified using primary care codes suggestive of AKI and associated with codes for hospitalisation, ED visit, or specialist visit. In the

HES, potential cases will be identified through hospital discharge ICD-10 codes suggestive of AKI. No ED visit information is available in the HES; ED visits will be identified from the CPRD primary care records.
- In Denmark, potential cases will be identified through hospital discharge ICD-10 codes, ED visit ICD-10 codes, or hospital outpatient clinic ICD-10 codes suggestive of AKI.
- In the HIRD, potential cases will be identified through hospital claims, ED visit claims, or specialist claims with an ICD-10 code or laboratory results suggestive of AKI.

Preliminary lists of ICD-10 codes are presented in Annex 4 Table 3. For the assessment of this AKI primary outcome, CKD exclusion criteria will be applied to the cohort as described in Section 9.7.2.1. Follow-up will be censored at the occurrence of a code for CKD as described in Section 9.2.9.

- AKI sensitivity, all cases: a sensitivity analysis will be performed, also including outpatient cases of AKI, in the data sources where primary care data are available (CPRD and HIRD). The exclusion and censoring conditions will be the same as for the primary outcome.
- AKI sensitivity, based on results from validation: another sensitivity analysis will repeat the primary outcome analysis for AKI, including only confirmed cases; however, when less than 70% of the cases are validated, other methods will be explored, e.g., correct the IRRs for the PPV [R18-1561].

9.3.2.4 Chronic kidney disease (inpatient and outpatient), secondary outcome

- Outcome type: secondary
- Secondary outcome: none
- Further outcomes within this section: none
- Outcome name: CKD
- Time frame: up to 5 years
- Safety issue: no

CKD will be defined based on the estimated glomerular filtration rate (eGFR), as calculated using the 2009 CKD-EPI equation as described below [R13-4387; R12-1392; R15-5270].
141 × min(SCr/k, 1)α × max(SCr/k, 1)-1.209 × 0.993Age [× 1.018 if female] [× 1.159 if black]

SCr is serum creatinine (in mg/dL), k is 0.7 for females and 0.9 for males, α is −0.329 for females and −0.411 for males, min is the minimum of SCr/k or 1, and max is the maximum of SCr/k or 1. As race is not available for all patients in CPRD, the race coefficient will not be applied if race is missing; this may result in underestimation of eGFR in black patients.

CKD will be differentiated from AKI by requesting confirmation of the first abnormal test result.

9.3.2.4.1 Validation case definition

In this study, a case of CKD is defined as any person meeting the following criteria after the index date:

- Estimated GFR < 60 mL/min/1.73 m2

AND

- Estimated GFR < 60 mL/min/1.73 m2 is confirmed in a separate test result performed at least 3 months after the initial post-index date at which the abnormal result was identified.

AND

- Absence of a recorded diagnosis of CKD or an estimated GFR < 60 mL/min/1.73 m^2 at any time before or at the index date.

9.3.2.4.2 Case identification

Potential cases of CKD will be identified by diagnosis codes suggestive of CKD (CKD stage 3 or higher). In the CPRD, potential cases will be identified using any primary care codes suggestive of CKD, irrespective of whether these are associated with a specialist visit, referral, hospitalisation, or ED visit [R15-3136]. In the HES, potential cases will be identified through hospital discharge ICD-10 codes suggestive of CKD [R15-3138]. Preliminary lists of codes are presented in Annex 4 Table 5 (ICD-10 codes). All codes selected and displayed in the tables had PPVs larger than 70% [R15-3136, R15-3138].

9.3.2.5 Severe complications of UTI (inpatient and outpatient pyelonephritis and urosepsis), primary outcome

- Outcome type: primary
- Secondary outcome within this section: No
- Further outcomes within this section: none
- Outcome name: severe complications of UTI
- Time frame: up to 5 years
- Safety issue: no

9.3.2.5.1 Validation case definition

To define **pyelonephritis**, the criteria established by Patkar et al. (2009) [R14-5253] will be used.

1. At least two of the following will have to be present:
    - History of fever or documented fever > 38.0°C or 104.0°F
    - Dysuric complaints
    - Flank pain/costovertebral angle tenderness
    - Leukocytosis (white blood cell count > 12,000/cubic mm)
    - Abnormal urine (cloudy, frank pus or blood in urine, foul smell)

AND (2 OR 3)

2. Any one of the following:

- Computed tomography, magnetic resonance imaging, or ultrasonography findings consistent with renal inflammation
- Computed tomography, magnetic resonance imaging, or ultrasonography findings consistent with renal abscess
- Computed tomography, magnetic resonance imaging, or ultrasonography findings consistent with hydronephrosis

OR

3. Any one of the following:
   - Blood cultures and urine cultures positive for the same organism
   - Blood cultures positive for Gram-negative organisms, Enterococcus species, or Staphylococcus saprophyticus
   - Urine culture positive for more than 105 Gram-negative organisms (e.g., Escherichia coli), Enterococcus species, or S. saprophyticus
   - Urine culture positive for fewer than 105 of any organism AND patient treated for at least 7 days with antibiotics

Urosepsis is clinically defined as sepsis caused by infection of the urinary tract and/or male genital organs (e.g., prostate). Patients are affected by microorganisms capable of inducing inflammation within the urinary and male genital tract. The following criteria from Wagenlehner et al. (2008) [R14-5286] need to be met for the diagnosis of urosepsis:

1. Diagnosis of infection of the urinary tract

AND

2. One of the following criteria:
   - Proof of bacteraemia
   - Clinical suspicion of sepsis

AND

3. Two or more of following, which indicate the presence of systemic inflammatory response syndrome (SIRS):
   - Body temperature ≥ 38°C or ≤ 36°C
   - Tachycardia (≥ 90 beats per minute)
   - Tachypnoea (≥ 20 breaths per minute)
   - Respiratory alkalosis (PaCO2 ≤ 32 mm Hg)*
   - Leucocytes ≥ 12,000 per µL or ≤ 4,000 per µL or band forms > 10%

The case definition allows a patient to meet the criteria for both pyelonephritis and urosepsis, but the case will be counted only once for analysis purposes.

*PaCO2 = partial pressure of carbon dioxide.

9.3.2.5.2 Case identification

The outcome severe complications of UTI comprises any of the following conditions:

- Hospitalisation, ED visit (as recorded in the CPRD primary care database), or a GP record for pyelonephritis (ICD-10 codes in Annex 4 Table 7)
- Hospitalisation, or ED (as recorded in the CPRD primary care database), or outpatient visit for urosepsis. Combination of diagnosis code for UTI and a diagnosis code for sepsis within 1 week (ICD-10 codes for sepsis in Annex 4 Table 7; ICD-10 codes for UTI in Annex 4 Table 9).

9.3.2.6 Genital infections (inpatient and outpatient), primary outcome

- Outcome type: primary
- Secondary outcome: severe GI
- Further outcomes within this section: none
- Outcome name: GI
- Time frame: up to 5 years
- Safety Issue: yes

Diabetes is a risk factor for vaginal infections and balanitis, especially those produced by *Candida* species, with studies suggesting that the rate of vulvovaginal candidiasis and balanitis is higher among patients with diabetes [R14-5237, R12-3639, R12-2432, R14-5260, R14-5243].

This study will include only non-sexually transmitted GI. For women, bacterial vaginosis and vulvovaginal candidiasis will be included. For men, only non-sexually transmitted cases of balanitis will be included, which comprise Candida balanitis and aerobic balanitis.

Clinical diagnosis of bacterial vaginosis is presumptive and based on the presence of typical symptoms of vulvovaginitis, elevated pH (> 4.7 or > 4.5), and the presence of clue cells in saline wet mount or Gram stain of vaginal discharge. Diagnosis is enhanced by fishy odour of vaginal discharge after addition of 1-2 drops of 10% potassium hydroxide. Cultures are not useful. *Gardnerella vaginalis* is commonly found in women with bacterial vaginosis. Other organisms associated with bacterial vaginosis include Prevotella species, *Mycoplasma hominis*, and *Mobiluncus* species [R14-5247, R14-5252].

Clinical diagnosis of vulvovaginal candidiasis is presumptive if there are typical symptoms of vulvovaginitis and microscopic identification of yeast forms or hyphae in Gram stain or potassium hydroxide wet mount preparations of vaginal discharge. Diagnosis is definitive by positive culture for *C. albicans* (or other *Candida* species) in symptomatic women [R14-5247, R14-5252].

Balanitis is defined as an inflammation of the penis that often involves the prepuce (balanoposthitis). Clinical diagnosis of balanitis is established based on clinical symptoms followed by culture confirmation; in the case of Candida balanitis, the isolation of yeast is

definitive proof of fungal infection. There is a wide variety of causes and predisposing factors such as not being circumcised, neutropenia, or diabetes [R14-5248, R14-5243]. Balanitis can be due to several microorganisms, but *Candida* is the most common organism in patients with diabetes [R12-3639].

Non-sexually transmitted balanitis may also be caused by anaerobic and aerobic organisms, (such as Gardnerella vaginalis and Group B Streptococcus). Typical symptoms of Candida balanitis include burning and itching of the penis with generalised erythaema of the glans and/or prepuce (which may have a dry, glazed appearance) and with erosions, papules, and white discharge. In patients with diabetes, the presentation may be more severe, with oedema and fissuring of the foreskin, which may become non-retractile. Symptoms of Gardnerella balanitis are milder and include irritation of the prepuce and glans penis, macular erythaema, and fishy subpreputial discharge. Symptoms of balanitis due to Group B Streptococcus include non-specific erythaema with or without exudate [R14-5243].

9.3.2.6.1 Validation case definition

Vulvovaginitis: all potential cases (outpatient and inpatient) of non-sexually transmitted vulvovaginitis will be included in the study and classified as vulvovaginal candidiasis, bacterial vaginosis, or non-specific vulvovaginitis or vulvitis.

Based on the clinical diagnosis mentioned above [R14-5247], either of the following criteria will need to be met for the diagnosis of vulvovaginal candidiasis:

1. A specific diagnosis of vulvovaginitis due to *Candida* AND any of the following
   a. Treatment with antifungals
   b. Yeast in Gram stain
   c. Culture positive for *Candida*
2. Non-specific diagnosis of vulvovaginitis OR symptoms of vulvovaginitis AND two or more of the following:
   a. Treatment with antifungals
   b. Yeast in Gram stain
   c. Culture positive for *Candida*

Based on the clinical diagnosis mentioned above [R14-5247], either of the following criteria will need to be met for the diagnosis of bacterial vaginosis:

1. A specific diagnosis of vulvovaginitis due to bacterial vaginosis AND any of the following:
   a. Treatment with metronidazole
   b. Clue cells in saline wet mount or Gram stain
   c. Culture positive for *Gardnerella*
   d. pH > 4.7
2. Non-specific diagnosis of vulvovaginitis OR symptoms of vulvovaginitis AND two or more of the following:
   a. Treatment with metronidazole

b. Clue cells in saline wet mount or Gram stain
c. Culture positive for *Gardnerella*
d. pH > 4.7

Cases with diagnosis or symptoms of vulvovaginitis or vulvitis that do not fulfil the above criteria will be classified as "non-specific vulvovaginitis or vulvitis" or excluded from the analysis if the event is suggestive of an alternative diagnosis such as a sexually transmitted GI.

Balanitis: all potential cases of non-sexually transmitted balanitis (outpatient and hospitalised) will be included in the study and classified as Candida balanitis, aerobic balanitis, or non-specific balanitis.

Based on the clinical diagnosis mentioned above [R14-5247], either of the following criteria will need to be met for the diagnosis of *Candida* balanitis:

1. A specific diagnosis of balanitis due to *Candida* AND any of the following:
   a. Treatment with antifungals
   b. Yeast in Gram stain
   c. Culture positive for *Candida*
2. (Non-specific diagnosis of balanitis OR symptoms of balanitis) AND two or more of the following:
   a. Treatment with antifungals
   b. Yeast in Gram stain
   c. Culture positive for *Candida*

Based on the clinical diagnosis mentioned above [R14-5247], either of the following criteria will need to be met for the diagnosis of aerobic balanitis:

1. Non-specific diagnosis of balanitis AND culture positive for a non-sexually transmitted microorganism such as *Gardnerella* or Group B *Streptococcus*.
2. Symptoms of balanitis AND culture positive for a non-sexually transmitted microorganism such as *Gardnerella* or Group B *Streptococcus*.

Cases with diagnosis or symptoms of balanitis that do not fulfil the above criteria will be classified as "non-specific balanitis" or excluded from the analysis if the event is suggestive of an alternative diagnosis such as a sexually transmitted GI.

9.3.2.6.2 Case identification

Vulvovaginitis: potential cases of vulvovaginitis will be identified by the presence of an outpatient diagnosis or a hospitalisation or ED visit (as recorded in the CPRD primary care database) for vulvovaginitis, specifically, codes for specific diagnosis of bacterial vaginosis or vulvovaginal candidiasis, codes for non-specific diagnosis of vulvovaginitis, codes for specific microbiology results for *Candida* or *Gardnerella*, or diagnosis codes for non-specific positive microbiology results (see ICD-10 codes in Annex 4 Table 11, and Read codes in

Annex 4 Table 12). Cases will be excluded if there is a diagnosis code with a specific diagnosis for a sexually transmitted infection within 30 days before or after the index infection date (see Annex 4 Table 13).

Balanitis: potential cases of balanitis will be identified by the presence of an outpatient diagnosis of or a hospitalisation or ED visit (as recorded in the CPRD primary care database) for balanitis, including Candida balanitis, aerobic balanitis, and non-specific balanitis (see ICD-10 codes in Annex 4 Table 11, and Read codes in Annex 4 Table 14). Cases will be excluded if there is a diagnosis code with a specific diagnosis for a sexually transmitted infection within 30 days before or after the index infection date (Annex 4 Table 15).

9.3.2.7 Severe genital infections: secondary outcome

- Outcome type: secondary
- Secondary outcome: none
- Further outcomes within this section: none
- Outcome name: incidence of severe genital infections
- Time frame: up to 5 years
- Safety issue: yes

Among the cases of GI identified in the primary outcome, those that resulted in hospitalisation will be classified as severe GI. To ensure that GI was the reason for hospitalisation and not a nosocomial infection, only primary discharge diagnoses will be used to identify GI.

Those GIs identified through an outpatient diagnosis and that required systemic treatment with antifungals or antibiotics (as opposed to topical or vaginal treatment) will also be classified as severe GI. Systemic treatment will be considered if prescribed on the date of the GI diagnosis or within 30 days before or after the GI diagnosis date. The UK recommended systemic treatment for vulvovaginitis includes metronidazole and tinidazole for bacterial vaginosis and fluconazole and itraconazole for vulvovaginal candidiasis [R14-5247]. The systemic treatment for balanitis recommended by the European Union is fluconazole for *Candida* balanitis, erythromycin or amoxicillin/clavulanic acid for aerobic balanitis, and metronidazole or amoxicillin/clavulanic acid for anaerobic balanitis [R15-0067].

Complications and severe consequences of GI will be described among severe GI cases. This will include patients with codes such as ulcerations, abscesses, boils, or Fournier´s gangrene (see ICD-10 codes in Annex 4 Table 11).

9.3.2.8 Hospitalisation or ED visit for DKA: primary outcome

- Outcome type: primary
- Secondary outcomes within this section: no
- Further outcomes within this section: none
- Outcome name: incidence of DKA
- Time frame: up to 5 years

- Safety issue: no

DKA is defined by the biochemical triad of ketosis, hyperglycaemia, and acidosis. Serious complications of DKA and its treatment are hypokalaemia and hyperkalaemia, hypoglycaemia, and cerebral oedema [R16-1372, R16-1373].

9.3.2.8.1 Validation case definition

In this study, a case of DKA is defined as any person with a recorded diagnosis compatible with DKA who meets the following criteria recommended by the Joint British Diabetes Societies [R16-1372] and the American Diabetes Association [R16-1371, R15-2058]:

- Ketonaemia > 1.5 mmol/L or ketonuria (+)
- pH less than 7.3. In the absence of pH values: anion gap > 10 mEq/L or venous bicarbonate (HCO_3) below 15 mmol/L

Although blood glucose levels are usually a parameter used to define DKA, this criterion will not be used because atypical DKA can occur among users of SGLT2 inhibitors, including empagliflozin.

In this study, cases of DKA will include only patients requiring hospitalisation for DKA because some authorities contended that true DKA will always require hospitalisation or ED admission for treatment; otherwise, this study would include patients with milder forms of ketosis and unconfirmed DKA events [R16-1371, R16-1372].

9.3.2.8.2 Case identification

The outcome “hospitalisation due to DKA” comprises hospitalisation, ED visit, or a GP record of hospitalisation or referral to hospitalisation for DKA. Potential cases of DKA will be identified by diagnosis codes suggestive of DKA.

- In the CPRD, potential cases will be identified using primary care codes and biochemistry test results, together with codes for hospitalisation or ED visit. In HES data, potential cases will be identified through hospital discharge ICD-10 codes in the primary or secondary discharge position. No ED visit information is available in the HES; ED visits will be identified from the CPRD primary care records.
- In Denmark, potential cases will be identified through hospital discharge ICD-10 codes or ED visit ICD-10 codes suggestive of DKA.
- In the HIRD, potential cases will be identified through hospital claims, or ED visit claims with an ICD-10 code suggestive of DKA.

Preliminary list of ICD-10 codes for DKA are available in Annex 4 Table 16.

9.3.2.9 Case validation process

The goal of case validation is to validate the algorithm used to identify the outcomes of interest. The target is a random sample of 100 cases each of the primary and secondary outcomes that will be validated in each data source where the outcome is being evaluated. In addition, a random sample of 100 outpatient or primary care cases of ALI and AKI will also be validated in the CPRD and HIRD.

Validation will be performed through general practitioner questionnaires in the CPRD and through medical record data abstraction in Denmark and the HIRD. The relevant clinical information from these sources will be abstracted using a standardised abstraction form. Final confirmation of cases will be conducted independently by medical epidemiologists who will be blinded to medication exposure. Difficult cases will be evaluated by consensus between the validation physicians.

For all outcomes, all cases (whether or not they have been selected for the validation process) will be included in the analysis. However, the PPV estimates available from the validation will help during interpretation of the results. A sensitivity analysis is planned for the validated outcomes that will repeat the primary analysis, including only confirmed cases; however, when less than 70% of the cases are validated, other methods will be explored, e.g., correct the IRRs for the PPV [R18-1561].

9.3.3 Covariates

Exclusion diagnoses will be identified based on recorded GP diagnoses or hospital outpatient or inpatient diagnoses during the lookback period. Definitions of specific variables will be adapted to the type and availability of data in each data source.

Variables potentially associated with the outcomes of interest—such as sociodemographic variables including age, sex, socioeconomic status or urban/rural area of residence in the HIRD, BMI, smoking or alcohol consumption, concomitant medications, comorbidities, Deyo-Charlson Comorbidity Index, health care utilisation, and duration of lookback period (see Annex 5)—will be identified for all cohort members before or at the index date, when available. Severity of T2D will be assessed by HbA1c values, diagnosis codes, and duration since the first diagnosis, when available. These and other variables that can differ by exposure group will be considered for inclusion as predictors in the logistic regression models used to generate the propensity scores. Propensity scores will be used in the analysis to quantify the probability of receiving empagliflozin at the index date for new users of empagliflozin or DPP-4 inhibitors.

At the index date, cohort members will also be classified by indicator variables on the calendar year at the index date, whether the index treatment (empagliflozin or DPP-4 inhibitor) was added to existing medication (adding on) or initiated as a replacement for another GLD (switching from the existing GLD to empagliflozin or DPP-4 inhibitor), and whether this treatment was received as monotherapy or as dual or triple therapy. A variable indicating whether or not patients were receiving insulin at the index date will also be

created. At the index date, patients will be classified as "adding on" or "switched to" empagliflozin or DPP-4 inhibitor based on whether at least one prescription/dispensing of the previous GLD treatment recorded in the 3 months before or at the index date is also recorded within 3 months after the index date.

The approaches to handling concomitant GLDs (including oral and injectable treatments such as insulin) in the analyses are summarised in Table 4.

Table 4 Approaches to handling concomitant glucose-lowering drugs

Timing and Type of GLD Prescription	**Analysis Approach**
At the index date, any GLD taken during baseline that is not the new prescription and is not eligible to be a study exposure	Include in propensity score
At the index date, any drug combination that includes the study drugs and metformin (i.e., empagliflozin plus metformin or a DPP-4 inhibitor plus metformin)	Include in propensity score and conduct additional stratified analyses by metformin use at index date (Yes or No)
Designation whether the GLD initiated at the index date is an add-on to current medication or a switch to a different medication	Include in propensity score
Designation whether the GLD initiated at the index date is a first-, second-, or third-line therapy	Include in propensity score
Insulin use at the index date	Include in propensity score and conduct additional stratified analyses by insulin use at index date (Yes or No)

GLD = glucose-lowering drug.

9.3.3.1 Description of cases of elevated liver enzymes

In addition to the evaluation of the ALI endpoint, this study will describe patients with elevated liver enzymes irrespective of whether the patients had a diagnosis or symptom code suggestive of ALI. Cases of elevated liver enzymes will be classified in the following categories:

- ALT and/or AST $\geq 3 \times$ ULN
- ALT and/or AST $\geq 5 \times$ ULN
- ALT and/or AST $\geq 10 \times$ ULN
- ALT and/or AST $\geq 20 \times$ ULN

Availability of laboratory data and identification of patients with cases of increase of liver enzymes varies by data source, as follows:

- In the CPRD, patients with elevated liver enzymes will be identified using outpatient laboratory results available in the CPRD primary databases.
- In Denmark, patients with elevated liver enzymes can be identified through linkage with the clinical laboratory information system databases, where data from all hospital-based laboratory tests performed during inpatient stays and outpatient hospital visits can be accessed electronically. In addition, if a GP sends samples to hospital-based laboratories, the result of the test will be recorded and available even if a patient is seen outside the hospital. For the period 2000-2015, linkage with the laboratory database is currently possible in the Central Region of Denmark, which corresponds to 23% of the total Danish population. On a nationwide level, laboratory data for all 5.77 million Danish inhabitants have recently become available for research in the new nationwide Register of Laboratory Results for Research (LAB_F), tracking all laboratory test results from both primary and secondary care, with complete data coverage beginning in 2015 at the Danish Board of Health [R19-1785]. Therefore, identification of cases of ALT and/or AST ≥ 3 × ULN will be possible in Denmark for the Central Denmark region (1.30 million people) since approximately the year 2000 and for all of Denmark (5.77 million people) since approximately the year 2015 [R18-3665]. It is still unclear if it will be possible for researchers to retrieve individual person-identifiable laboratory data for the entire country of Denmark for local validation purposes, e.g., through medical records, because nationwide research data at the Danish Board of Health are normally de-identified.
- In the HIRD, cases of elevated liver enzymes will be identified using information from outpatient laboratory test results available electronically for 30% of the patients in the HIRD.

The number and proportion of patients with elevated liver enzymes identified in each data source will be described overall and stratified by age and sex, by treatment group, and by prior history of predisposing conditions.

9.4 DATA SOURCES

Protocol version 4.0 included the CPRD data source in the UK. A feasibility assessment of several data sources based on actual and projected user counts, type and availability of data, and possibility of validation, revealed that the most efficient approach to increase study size was to include the Danish Population Registries and the HIRD in the US in the study (Annex 1).

9.4.1 Clinical Practice Research Datalink (CPRD)–UK

Ideally, the proposed study design requires data sources that longitudinally capture inpatient and ED (and outpatient for some of the outcomes) diagnoses and procedures; capture prescription information; and allow validation of the outcomes of interest. A data source meeting all those study requirements is the UK CPRD (website: cprd.com), which is proposed as the study data source.

In the UK, nearly all residents are registered in a general medical practice that uses electronic medical records. Some of those records are available for research purposes in the CPRD. The CPRD contains diagnostic and prescribing information recorded by GPs as part of their routine clinical practice in the UK. The CPRD primary care data are divided into two databases (CPRD GOLD and CPRD Aurum) that differ in the electronic patient record system software. CPRD GOLD uses the VISION software and was the only software available until recently. Since 2016-2017, primary care practices across the UK are progressively switching to the Egton Medical Information Systems electronic patient record system (EMIS-Web) software, and data from these practices are now collected into CPRD Aurum. Both databases, i.e., CPRD GOLD and CPRD Aurum, will be used for this study.

As of April 2019, CPRD GOLD contains data for over 16.7 million patients, with research-quality data from 790 UK practices; 2.6 million of these patients are active (still registered with one of the 296 contributing GP practices) [R19-1781]. As of April 2019, CPRD Aurum contains data for approximately 22.7 million patients, with research-quality data from 873 GP practices in England; 7.3 million of these patients are active (registered with one of the 754 contributing GP practices). No data on practices from Scotland, Wales, or Northern Ireland are available in CPRD Aurum [R19-1782]. It is possible that patients in practices that switched from CPRD GOLD to CPRD Aurum may have contributed to both databases. Based on the monitoring of users, between 1% and 2% of the users of empagliflozin who are in CPRD GOLD are also in CPRD Aurum. Duplicate practices can be identified [R18-0349], and patients will be removed from CPRD GOLD because when patients switch from CPRD GOLD to CPRD Aurum, validation is possible only in CPRD Aurum. Patients registered in CPRD GOLD and CPRD Aurum are representative of the whole UK population in terms of age and sex. A large proportion of patients are linkable to central mortality records.

A large and growing proportion of patients can also be linked to hospitalisation records (hospital discharge diagnoses and procedures are coded using ICD-10 codes in the HES) and to death registration data records from the Office for National Statistics (ONS). Linkage is done via the patient's National Health Service number, sex, date of birth, and postal code. Updated, valid, linked CPRD data are available through the CPRD Division of the UK Medicines and Health Care Products Regulatory Agency. The HES data set to be used is HES Admitted Patient Care (APC), which includes information on admissions (including day cases) to English NHS health care providers, with admission and discharge dates, diagnoses, specialists seen, and procedures undertaken (using OPCS codes). Although a HES Accident and Emergency (A&E) linkage database exists and includes individual records of patient care administered in accident and emergency settings at English NHS health care providers and treatment centres, it is not used because data coverage is incomplete in comparison with national A&E data attendances. For this study, information on ED visits will come from CPRD primary care records, which can be in the form of referrals to A&E or other administrative codes indicating that a diagnosis is related to an ED visit.

CPRD GOLD linkage of data to HES and ONS covers approximately 75% of practices contributing to CPRD GOLD in England and approximately 54% of practices contributing to CPRD GOLD in the UK. Within the participating practices, 88% of research-quality

(acceptable) patients have the necessary data to allow linkage to HES and ONS and have not dissented from disclosure of personal confidential data to NHS Digital [R19-1734].

CPRD Aurum linkage of data to HES and ONS covers 93% of all CPRD Aurum practices available in the June 2018 build, all of which are in England. The large volume of data that the CPRD are incrementally onboarding from CPRD Aurum practices means that additional practices will be added over subsequent linkage sets [R19-1734].
Detailed information on prescriptions written by GPs, including prescribed dose and duration, is routinely recorded in the data source. Read codes are used for diagnoses in CPRD GOLD and CPRD Aurum, and SNOMED codes and EMIS local codes are also used in CPRD Aurum for diagnoses. Gemscript codes are used for medications in CPRD GOLD, and medications are coded using the Dictionary of Medicines and Devices (DM+D) in CPRD Aurum. Additional diagnostic and treatment information can be found in letters from specialists and hospitals, and other sources. Because GPs serve as the gatekeepers for all medical services, any visit to a specialist or hospital requires communication back to the GP, who might enter that information into the medical record.

CPRD GOLD and CPRD Aurum contain information on lifestyle factors with a variable proportion of missing values. Although information on race is not available, other user characteristics of interest are likely to be captured. For example, data on body weight and height, smoking, and alcohol use were available for approximately 70% of patients in the CPRD [R14-5279]. In contrast, the pharmaceutical exposures and comorbidities are expected to be based on outpatient prescriptions and to be complete. The diagnosis of T2D, after excluding individuals with diagnosis codes for T1D, has been validated in the CPRD and found to have a high PPV: 98.6% [R14-5280].

The number of practices contributing to the CPRD GOLD database is decreasing, and the number of practices contributing to the CPRD Aurum database is increasing. The flow of patients between CPRD GOLD and CPRD Aurum is not clearly understood at this stage and will be monitored closely. The number of empagliflozin users in CPRD Aurum was low as of December 2017, but monitoring of users shows a rapid increase in the number of empagliflozin users in CPRD Aurum, with a total of 17,650 users between August 2014 and 31 March 2019. Monitoring of empagliflozin users in CPRD Aurum and CPRD GOLD will be reported in the annual progress reports.

The UK is an ideal setting for population-based studies of diabetes because diabetes care is largely coordinated by the GP, and metabolic parameters, cardiovascular risk factors, diabetes comorbidities, and disease outcomes are collected electronically. Furthermore, clinical guidelines in the UK facilitate consistency in patterns of care [R14-5255]. In general, the validity of the former General Practice Research Database, upon which the CPRD was founded, as a reliable data source for drug safety studies in numerous therapeutic areas is well established [R11-2162, R99-1044]. However, all of the outcomes of interest in this study have not been well validated in the CPRD. For example, the validity of the diagnosis codes for severe complications of UTI or pyelonephritis has not yet been determined and does not appear in a recent systematic review of the validated outcomes in the CPRD. For ALI and AKI, the PPV of the codes analysed in one study seems to be below 50% [R11-5210]. On the

other hand, the combined infectious and parasitic endpoints studied in the CPRD have a median proportion of cases confirmed of 93%; for the combined genitourinary system endpoints, the median proportion of cases confirmed is 91% [R11-5211].

Access to de-identified, patient-level data from the CPRD is available following approval of the study protocol by the Independent Scientific Advisory Committee. A more detailed description of data available from the CPRD is shown in Annex 6.

9.4.2 The Danish Population Registries

The Danish health care system provides universal coverage to all Danish residents (5.7 million inhabitants; https://www.sundhed.dk/service/english/an-ehealth-nation/healthcare-in-dk/) [R18-0353]. Health care coverage includes visits to GPs and specialists, hospital admissions, and outpatient visits. The costs of medicines are partially covered by the Danish health system. The centralised Civil Registration System in Denmark allows for personal identification of each person in the entire Danish population [R16-2604] and for the possibility of linkage to all Danish registries containing civil registration numbers, such as the Danish National Patient Register [R18-0262], Danish National Prescription Registry [R18-0263], and the clinical laboratory information system research databases [R18-0264]. Data collected in these registries are available for research purposes after following a standard application procedure to the relevant data board. The process requires collaboration with a local university or investigator affiliated with a research institute to access the data; Danish Data Protection Agency approval to handle data; data release by the Danish National Data Board; and, for accessing medical charts, approval of a Patient Safety Board [R15-3135, R15-3137]. All applications have to be submitted in Danish.

Denmark's primary health care sector, which includes GPs, specialists, and dentists, generates about 96% of the prescription sales, most of which are reimbursable and are dispensed by community pharmacies. Each dispensing record contains information on the patient, drug, and prescriber. Dispensing records retain the patient's universal personal identifier, allowing for individual-level linkage to all Danish registries and medical databases.

Three national registries—Danish National Patient Registry, Danish National Prescription Registry, and Danish National Database of Reimbursed Prescriptions—will be of particular interest. In addition, the Danish National Civil Registration System will be used to obtain information on death and migration status.

- The Danish National Patient Register includes data on all hospital admissions since 01 January 1977 and on outpatient clinic and ED visits since 1995 [R15-3137]. Hospital discharge diagnoses and information on surgical procedures, in-hospital deaths, and some selected drugs are recorded. After 1993, hospital discharge diagnoses are coded using ICD-10 codes.
- The National Health Services Prescription Database (formerly known as the Danish National Database of Reimbursed Prescriptions) encompasses the reimbursement records of all reimbursed drugs sold in community pharmacies and hospital-based outpatient pharmacies in Denmark since 2004 [R15-3140]. On average,

approximately 3.5 million users are recorded in the database each year. Individuals are identified by the unique central personal registration (CPR) number assigned to all persons born in or immigrating to Denmark. This new data source avoids restrictions imposed on data use at the Danish National Prescription Registry. Most importantly, CPR numbers are reversibly encrypted, which allows re-identification of medication users. These features are very important for validation purposes.

- Laboratory results back to 2015 have become available nationwide [R19-1785]. For a subpopulation of Danish drug users, i.e., inhabitants of the Central Denmark Region (population, 1.3 million; approximately 23% of the Danish population), acute liver and kidney injury outcomes can also be assessed through population-based laboratory databases with complete coverage back to 2000 [R18-0264, R18-067]. Thus, liver injury can be assessed by elevated liver enzyme values (e.g., alanine aminotransferase [ALT] values increasing to > 437 U/L for men older than 16 years or ALT > 282 U/L for others), and kidney injury through elevated serum creatinine and decreased estimated glomerular filtration rate (eGFR) (e.g., 1.5-fold increase in serum creatinine from baseline or eGFR decreasing to a different kidney dysfunction severity level < 60 mL/min/1.73).

Identification and validation of the outcomes of interest in Denmark are summarised in Table 5.

Table 5 Evaluation of acute liver injury, acute kidney injury, and hospitalisation for diabetic ketoacidosis in the Danish Population Registries

Case ascertainment through electronic algorithm: information available/not available	**Medical record abstraction: information available and implications for outcome validation**
• Hospital discharge ICD-10 codes are available for all inpatient episodes during the study period. • Hospital outpatient ICD-10 codes are available for all outpatient specialist visits during the study period. • Hospital procedure codes for acute dialysis are available to identify the most severe AKI cases. • Outpatient primary care diagnoses are not available; outpatient primary care cases of AKI and ALI cannot be identified. • Laboratory results data are available for patients treated in Hospitals in the Central Denmark Region from 2000 [R18-0264] and nationwide from 2015.	• Hospital records for both outpatient clinic and inpatient episodes can be abstracted for almost 100% of all potential cases identified by the electronic algorithms. • Laboratory results from hospital inpatient and outpatient care and outpatient primary care can be abstracted from hospital medical records, when available. • No validation of outpatient primary care cases of ALI and AKI is possible, but most ALI and AKI outpatient cases are managed in the hospital setting.

AKI = acute kidney injury; ALI = acute liver injury; ICD-10 = *International Statistical Classification of Diseases and Related Health Problems, 10th Revision.*

A more detailed description of data available from the Danish Population Registries is shown in Annex 6.

9.4.3 HIRD, US

The HIRD is a large administrative health care database maintained by HealthCore for use in health outcomes and pharmacoepidemiologic research. The HIRD contains longitudinal medical and pharmacy claims data from health plan members across the US. The database represents claims information from one of the largest commercially insured populations in the US.

Member enrolment, medical care (professional and facility claims), outpatient prescription drug use, outpatient laboratory test results data available for 30% of the patients, and health care utilisation may be tracked for patients in the database dating back to January 2006. Diagnoses and procedures are identified by *International Classification of Diseases, 9th Revision, Clinical Modification codes* (ICD-9-CM), *International Classification of Diseases, 10th Revision, Clinical Modification codes* (ICD-10-CM); Current Procedural Terminology (CPT) codes, and Healthcare Common Procedure Coding System (HCPCS) codes for both

outpatient visits and inpatient stays. Drug claims are captured by National Drug Codes, which can be translated to broader categories such as Generic Product Identifier codes. Information on physician speciality is also available in the HIRD.

In addition, the HealthCore Integrated Research Environment has the ability to link claims data in the HIRD to complementary data sources, including inpatient and outpatient medical records for the health plan members represented in the HIRD; identify and contact providers and members for survey research through vendor relationships; and link data to national vital records, such as the National Death Index (NDI), for date and cause of death [R18-0256]. Death records are added to the NDI file annually, approximately 12 months after the end of the calendar year, although early release files are available with a lag of around 5 months and are usually more than 95% complete. Cause-of-death codes can be obtained using the NDI Plus service.

Using these resources, HealthCore conducts a range of real-world research studies, including retrospective database studies, medical record review studies, cross-sectional and longitudinal patient and provider surveys, and prospective site-based studies. This research team has a long experience on pharmacoepidemiology studies, including validation studies and studies on ALI and AKI [R18-0255, P14-10229, R18-0350].

Identification and validation of the outcomes of interest in the HIRD are summarised in Table 6.

Table 6 Evaluation of acute liver injury, acute kidney injury, and hospitalisation for diabetic ketoacidosis in the HIRD

Case ascertainment through electronic algorithm: information available/not available	**Medical record abstraction: information available and implications for endpoint validation**
• Hospital discharge ICD-9 or ICD-10 codes (depending on the time period) are available for all inpatient, outpatient, and primary care claims during the study period. • Laboratory results from large national reference laboratories from the outpatient setting and primary care are available electronically for approximately 30% of the patients.	• HealthCore has permission to obtain medical records for a subset of patients (approximately 50%). Redacted copies of medical records are typically obtained for approximately 60% of the patients for whom HealthCore has access to their protected health information.[1] • These medical records can be from any providers, including primary care providers, hospitals, emergency departments, and providers of any speciality care. • Diagnoses from primary care claims or any other outpatient or hospital setting can be validated. • Validation is performed by a trusted third party. • Outpatient laboratory results are available electronically for a subset of approximately 30% of the patients (no need to abstract outpatient laboratory results from medical records, although it is possible to access medical records for around 15% of these patients). • When outpatient laboratory results are not available electronically, these can be obtained through abstraction of information in outpatient medical records, if available. • Inpatient laboratory results can be obtained through abstraction of information in hospital medical records, if available.

ICD-9 = *International Classification of Diseases, 9th Revision*; ICD-10 = *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*.

1 HealthCore is successful in obtaining redacted copies for 60% of the patients of the ~50% of patients for whom HealthCore has permission to obtain medical records.

A more detailed description of data available from the HIRD is shown in Annex 6.

9.5 STUDY SIZE

The study size will be driven by the uptake of empagliflozin following approval and launch of empagliflozin for the treatment of T2D to improve glycaemic control in adults in each country.

The required study size to detect an IRR of 3 among empagliflozin new users compared with DPP-4 inhibitor new users, with a comparator: empagliflozin ratio of 10:1 and a power of 80%, would be between 18,300 and 30,000 person-years of empagliflozin use for liver injury, approximately 8,400 person-years for DKA, and 1,400 to 3,200 person-years for kidney injury. For all other outcomes, the number of empagliflozin new-user person-years required to detect an IRR of 3 would be less than 1,400 (Table 7).

As of June 2018, the actual and projected number of users of empagliflozin in CPRD GOLD, Denmark, and the HIRD are presented in the feasibility assessment (Annex 1). The actual number of users of each of the study medications in each data source suggests that the DPP-4 inhibitor: empagliflozin ratio is likely to be 10:1 or lower. The actual and projected number of users of empagliflozin suggest that if the study includes CPRD GOLD, Denmark, and the HIRD, at the time of data extraction, there would be around 109,885 patients and approximately 21,978 person-years of empagliflozin exposure, which would be sufficient to provide a probability of 0.80 that the upper bound of the IRR for empagliflozin compared with DPP-4 inhibitors is less than 3 if the true IRR of ALI and DKA is 1.0, and less than 1.5 if the true IRR of AKI is 1.0.

As of 31 March 2019, the actual number of empagliflozin users in CPRD Aurum (N = 17,650) is higher than the number projected in the feasibility assessment (N = 2,541 projected up to December 2019). Because CPRD Aurum now contains a larger number of users of empagliflozin, this database will also be included in the study to increase the study size and person-years of exposure in the UK. Similarly, as of 31 March 2019, the actual number of empagliflozin users in CPRD GOLD (N = 8,316) is also higher than the number projected in the feasibility assessment (N = 5,874). Of the 8,316 users in CPRD GOLD, 303 users are also in CPRD Aurum and 8,013 of users are unique to CPRD GOLD. As of 31 March 2019, the actual and projected numbers of empagliflozin users suggest that if the study includes CPRD GOLD, CPRD Aurum, Denmark, and the HIRD, at the time of data extraction, there would be approximately 151,184 patients and approximately 30,237 person-years of empagliflozin exposure, which would be sufficient to provide a probability of 0.80 that the upper bound of the IRR for empagliflozin compared with DPP-4 inhibitors is less than 3 if the true IRR of ALI is 1.0, less than 2 if the true IRR of DKA is 1.0, and less than 1.5 if the true IRR of AKI is 1.0.

Table 7 Number of empagliflozin-exposed person-years needed to detect an IRR of 1.5, 2, 3, or 4 with a power of 80% and alpha = 0.05 (using a two-sided test for the ratio of two Poisson rates)

Outcome	**Background incidence rates[1]**	**Comparator:empagliflozin ratio, 10:1**				**Comparator:empagliflozin ratio, 20:1**			
		IRR, 1.5	**IRR, 2**	**IRR, 3**	**IRR, 4**	**IRR, 1.5**	**IRR, 2**	**IRR, 3**	**IRR, 4**
ALI	0.14[a]	309,102	92,008	30,027	16,360	293,301	86,858	28,101	15,200
	0.23[b]	188,149	56,005	18,277	9,958	178,531	52,870	17,105	9,252
AKI	1.29[c,d]	33,546	9,985	3,259	1,776	31,831	9,427	3,050	1,650
	1.98[c]	21,856	6,506	2,123	1,157	20,739	6,142	1,987	1,075
	2.88[e]	15,026	4,473	1,460	795	14,258	4,222	1,366	739
Acute pyelonephritis	3[f]	14,425	4,294	1,401	764	13,687	4,053	1,311	709
UTI leading to hospitalisation	15.8[g]	2,739	815	266	145	2,599	770	249	135

Table 7 (cont'd) Number of empagliflozin-exposed person-years needed to detect an IRR of 1.5, 2, 3, or 4 with a power of 80% and alpha = 0.05 (using a two-sided test for the ratio of two Poisson rates)

Outcome	Background incidence rates[1]	Comparator:empagliflozin ratio, 10:1				Comparator:empagliflozin ratio, 20:1			
		IRR, 1.5	IRR, 2	IRR, 3	IRR, 4	IRR, 1.5	IRR, 2	IRR, 3	IRR, 4
Vaginitis[2]	21[h]	2,061	613	200	109	1,955	579	187	101
Balanitis[3]	8.4[h]	5,152	1,534	501	273	4,888	1,448	468	253
DKA in T2D	0.5[i]	86,549	25,762	8,408	4,581	82,124	24,320	7,868	4,256

AKI = acute kidney injury; ALI = acute liver injury; DKA = diabetic ketoacidosis; IRR = incidence rate ratio; T2D = type 2 diabetes mellitus; UTI = urinary tract infection.

1 Incidence rates per 1,000 patient-years.

2 Patient-years restricted to females.

3 Patient-years restricted to males.

Sources: a Huerta et al. (2002) [P03-03701]; b El-Serag and Everhart [R12-3632]; Girman et al. (2012) [R11-5319]; d Estimate calculated as the 65% of the incidence reported by the study, 2 per 1,000 person-years. Only 65% of those were based on hospital records; e Waikar et al. (2006) [R14-5285] (the results reported are on patients without T2D); f Venmans et al. (2009) [R12-1105]; g Benfield et al. (2007) [R12-1080]; h Hirji et al. (2012) [R12-3639]; i Wang et al. (2008) [R14-3272].

9.6 DATA MANAGEMENT

All conversion of the original data to analysis variables will be performed using SAS software version 9.2 or higher (SAS Institute, Inc., Cary, North Carolina). Data management for CPRD data will be carried out in accordance with RTI Health Solutions (RTI-HS) standard operating procedures. Routine procedures include checking electronic files, maintaining security and data confidentiality, following the statistical epidemiological analysis plan, and performing quality-control checks of all programmes.

Security processes will be in place to ensure the safety of all systems and data. Every effort will be made to ensure that data are kept secure so that they cannot be accessed by anyone except select study staff. Appropriate data storage and archiving procedures will be followed (i.e., storage on CD-ROM and DVD), with periodic backup of files to tape. A more complete description of the data management procedures will be included in the statistical epidemiological analysis plan.

9.7 DATA ANALYSIS

The final approach to data analysis will be presented in a separate statistical epidemiological analysis plan, to be developed before data collection.

9.7.1 Propensity score approach

Decisions to begin a specific GLD are influenced by demographic, medical, and clinical factors, and those same factors might be associated with the outcomes of interest. In the context of this study where the expected number of patients meeting the case definition is small for some of the outcomes (e.g., ALI), the number of covariates that can be used in a regression model predicting those outcomes is limited [R08-1486, R08-1494]. To overcome this problem, the set of confounding variables will be summarised into a single summary confounder score, a propensity score. The propensity score is the predicted probability of being assigned to a particular treatment conditional on a set of observed covariates. Because the models predict not the probability of experiencing the outcome but the probability of being treated with empagliflozin in this study, many more variables can be used in the predicting regression model [P12-04844, R14-5241, R14-5284, R14-5389]. Propensity scores for the comparison of empagliflozin vs. DPP-4 inhibitors will be generated. Furthermore, given the different inclusion/exclusion criteria used for each of the outcomes, the propensity scores will be cohort specific to the outcome being analysed (i.e., propensity scores will be calculated for only those patients included in the specific outcome analysis).

As a first step, a propensity score is estimated for each cohort member at the index date, based on the values of the observed covariates. In this study, propensity scores will be estimated by conducting multivariable logistic regression modelling and incorporating measured potential predictors of therapy as independent variables and exposure group status (empagliflozin group vs. DPP-4 inhibitor group) as the outcome. “Combination with metformin” status will be taken into account by indicator variables that will be used in the analysis stratified by use of combinations with metformin at the index date.

The selection of variables to be included in the propensity score modelling (see Section 9.3.3) will be based on examination of exposure group differences in the distribution of each covariate and within categories of insulin use at the index date. Ideally, the included covariates should 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. Thus, selection of variables to be included in the propensity score models will be based on their independent associations with the outcome rather than with the exposure. Potential associations will be evaluated from the literature and, when needed, from bivariate associations with the outcome within the data or from methods that account for small numbers of outcomes (e.g., Poisson regression or logistic regression using Firth penalised likelihood approach), if appropriate.

The variables listed in Annex 4 Table 17 are potential candidates for inclusion in the propensity score model: Annex 5 Table 1(ALI), Annex 5 Table 2 (AKI), Annex 5 Table 3 (UTI), Annex 5 Table 4(GI), and Annex 5 Table 5 (DKA).

Those variables will be assessed at the index date and during the previous lookback time period. Duration of the lookback period will be categorised, with indicator variables to be used for propensity score development and potential adjustment in multivariable regression models (see Section 9.7.1).

Prescription patterns change over time, and the confounding influence of the determinants of the prescription may also change. To allow for changing prescription patterns for empagliflozin from the time it is first available through the date of receipt of the data, a full-interaction propensity score model will be generated where calendar year will be included as a covariate along with its interaction with each of the other covariates. This full-interaction propensity score model will allow the influence of each covariate in predicting treatment to vary across calendar years (thus accounting for potential channelling bias) and has an advantage of efficiency in providing one overall comprehensive model compared with generating separate propensity score models by calendar year [R18-1214].

Descriptive analyses of covariates at baseline will include means, standard deviations, and medians/interquartile ranges, when appropriate, for continuous variables and percentages for categorical variables. Next, selective removal of observations, known as “trimming” [P12-04844], which occurs at both ends of the propensity score range, will be implemented. At the low end of the range, all patients, exposed or unexposed, with a propensity score below the 2.5 percentile value of the distribution of scores in the exposed group (i.e., empagliflozin) will be excluded. At the upper end of the range, we will exclude all patients, exposed and unexposed to empagliflozin, with scores greater than the 97.5 percentile of scores among the comparator patients (i.e., DPP-4 inhibitor). Trimming will be performed separately for each outcome-specific cohort and, within each cohort, for each set of propensity scores.

After trimming is completed, data will be stratified into deciles of propensity scores based on the distribution among empagliflozin new users. However, if the resulting decile strata are too small (i.e., have zero events in both treatment groups), deciles may be combined (e.g., to form quintiles), or alternative methods of producing summary estimates in the presence of sparse data will be explored, such as matching weights, which has been demonstrated to yield the smallest amount of bias in health care database studies with rare binary outcomes [P18-11001]. The propensity scores obtained after trimming will be used in the analysis to control for measured confounding variables simultaneously while reducing the loss in degrees of freedom that would have been experienced with a multivariable model, which is particularly important in studies with rare events. The propensity score methodology to be applied in this study will be further detailed in the statistical epidemiological analysis plan.

9.7.2 Primary and secondary objectives: estimate adjusted incidence rate ratios and compare adjusted incidence rates for each of the study outcomes

Adjusted incidence rates of ALI, AKI/CKD, severe complications of UTI, and incidence rates of GI among new users of empagliflozin and of DPP-4 inhibitors will be estimated and compared. Incidence rates will be reported as point estimates (in cases per 1,000 person-years) and 95% CIs. Ascertainment during follow-up will allow estimation of the number of new cases for each of the primary outcomes. Current use person-time for each patient will be allocated as the time between the date of the first prescription for either empagliflozin or a DPP-4 inhibitor and the end of current time at risk (see Section 9.3.1.1 for time at risk definitions). The total person-time of observation among individuals at risk will then be calculated.

9.7.2.1 Main analysis

For each of the primary outcomes (primary objective) and the secondary outcomes (secondary objective) of interest, estimation of adjusted current use IRRs with 95% CIs will be considered the main analysis of interest. Adjustment will be implemented by stratifying for propensity score deciles among empagliflozin new users vs. DPP-4 inhibitor new users. More details on the analysis methods will be included in the statistical epidemiological analysis plan.

Although IRRs will be estimated using two different comparison groups and for different outcomes, no Bonferroni type I error adjustment for multiple comparisons is planned for this study [R14-1393, R14-2938].

Crude IRRs will facilitate comparison with the adjusted IRRs to provide an indication of the degree of confounding. However, crude IRRs cannot be used for any comparisons given that this is a non-randomised, observational study and crude IRRs are expected to be biased due to channelling and confounding. Secondary objectives are to estimate adjusted incidence rates of each of the outcomes among empagliflozin and DPP-4 inhibitor new users and estimate adjusted IRRs stratified by insulin use at baseline.

9.7.2.2 Secondary outcome analyses

The following adjusted incidence rates for each of the primary and secondary outcomes of interest among empagliflozin new users and among DPP-4 inhibitor new users will be estimated:

- Cumulative incidence function plots will be reported graphically to depict the cumulative incidence rates of events during the follow-up.
- Adjusted incidence rates of the primary and secondary outcomes, overall and stratified by categories of insulin use at the index date, age, sex, and other variables of interest such as diabetes control.

Because the patients with T2D treated with insulin may have diabetes in a more advanced stage, may receive a different GLD treatment, and may be at different risk of the outcomes of interest, the incidence rate for each of the primary outcomes of interest will be estimated stratified by insulin use at the index date in each cohort. Sensitivity analysis will include the combined analysis of patients with T2D treated with and without insulin, and tests of interaction will be performed to evaluate whether these estimates are biased. Incidence rates will be reported as point estimates (in cases per 1,000 person-years) and 95% CIs. The following estimates and comparisons will be generated:

- Summary IRRs after adjusting, by stratifying for propensity score deciles, overall and by categories of insulin use at index date, age, sex, and other variables of interest such as diabetes control, comparing empagliflozin new users with DPP-4 inhibitor new users.

9.7.3 Duration, dose, and recent use effects analysis

- Adjusted IRRs by categories of duration of exposure will be estimated among empagliflozin new users vs. DPP-4 inhibitor new users. For example, patients with at least 1 year of continuous empagliflozin exposure will be compared with patients with at least 1 year of continuous DPP-4 inhibitor exposure; patients with less than 1 year of exposure to empagliflozin will be compared with patients with less than 1 year of exposure to DPP-4 inhibitors. Categories of duration will be defined based on available data. This analysis will be done if there is enough variation in the duration of use within the empagliflozin group.
- Adjusted IRRs by exposure dose categories will be estimated among empagliflozin new users vs. DPP-4 inhibitor new users. This analysis will be done if dose variation occurs within the empagliflozin group.
- Main analyses repeated to estimate adjusted IRRs using recent use (recent time at risk) instead of current use (current time at risk).

9.7.4 Interim reports to monitor accrual of empagliflozin users and the event rates of acute liver injury and acute kidney injury

Accrual of empagliflozin users was monitored and reported annually in three interim reports [BI document numbers c10503662-01, c17493314-01, and c24037075-01]. The interim reports included data up to 19, 24, and 36 months after use of empagliflozin was first captured in the CPRD were released in June of each year. Crude incidence rates of acute liver and kidney injury outcomes (overall, not stratified by treatment) were generated for the second and third interim reports and were in line with what it is expected considering that no validation was performed. Given the current event rates, the number of new users of empagliflozin accrued at 36 months after launch was too low to yield acceptable precision and statistical power, the study population has been expanded by including countries with better market uptake of empagliflozin.

9.7.5 Imputation of missing values

In the CPRD, no high frequency of missing values is expected for most variables, except for lifestyle variables. If missing data are common for lifestyle variables, multiple imputation methods will be used to replace missing values during propensity score generation and multivariable analysis. Additional details on when and how multiple imputation methods will be used will be provided in the statistical epidemiological analysis plan. We propose to use multiple imputation methods because in most cases they allow for better bias correction than alternative methods and are more efficient than the complete-case approach. The complete-case approach can be very costly of information in a body of high-dimensional data, since the proportion of complete cases will decline with the increase in the number of variables [R07-2456, R14-5281]. Multiple imputation methods assume that the data are missing at random. If it is determined based on previously established knowledge of the data source that the data for particular variables in question are not missing at random (e.g., the data collection mechanism is such that sicker patients do not have those variables collected), then alternative methods will be explored. Analyses evaluating potential effects of variables that are missing not at random on the estimation of the treatment effect are inherent in the sensitivity analysis outlined in Section 9.7.6, item 5. Details will be specified in the statistical epidemiologic analysis plan.

In Denmark and the HIRD, no data will be available for smoking, alcohol, BMI, HbA1c (available in around 30% of the patients in the HIRD), but no high frequency of missing values is expected for most of the other variables. Effect of unmeasured confounders will be evaluated in a sensitivity analysis.

Of note, for the medical history conditions and 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.

9.7.6 Sensitivity analyses

The following sensitivity analyses will be conducted:

1. In the main analysis (see Figure 1) current use ends 30 days after end of supply but in this sensitivity analysis current use will end 90 days after end of supply. This change will be applied to all exposure groups: empagliflozin new users and DPP-4 inhibitor new users. New adjusted IRRs will then be estimated for empagliflozin users vs. DPP-4 inhibitor users.
2. For cases of ALI and AKI, a sensitivity analysis including outpatient cases of ALI or AKI, in the data sources where primary care data are available (CPRD and HIRD). The exclusion and censoring conditions will be the same as those for the primary outcome.
3. For all outcomes, conduct analyses including only validated cases or correct estimates with PPVs: i.e., estimate summary adjusted IRRs among validated cases or, when less than 70% of the cases are validated, explore other methods to correct for disease misclassification (e.g., correct the IRRs for the PPV obtained from the validation) [R18-1561]..
4. Compare characteristics of validated cases with those of cases that were not validated.
5. For cases of CKD, conduct analyses to explore the time between exposure and CKD using varying time windows to explore the lag time to the outcome. The time window will be defined in the statistical epidemiological analysis plan. Exploratory analyses using different definitions of CKD will also be conducted to compare definitions and to facilitate comparison to other studies. The definitions will be defined in the statistical epidemiological analysis plan.
6. Assess the potential effect of unmeasured confounders on the association between empagliflozin use and, for example, ALI, by using the method described by Lash et al. [R14-5373]. These methods may be of special interest to evaluate HbA1c as a measure of disease control, in all data sources, or smoking, BMI, and other variables in the HIRD and Denmark, where no data are available. Additionally, these methods apply to variables that are missing not at random, as the missingness of these variables is ultimately unmeasured confounding. More details and examples of how this bias analysis method will be used will be provided in the statistical epidemiological analysis plan.
7. Estimate summary adjusted IRRs using intention-to-treat analysis, carrying forward the initial exposure status and disregarding changes in treatment status during all available follow-up time.
8. Tabulate the crude count of each study outcome by DPP-4 inhibitor type to provide an indication regarding whether the distribution of study outcomes in each individual type of DPP-4 inhibitor is different from the expected count based on patient use.
9. Conduct a sensitivity analysis of the meta-analysis (described in Section 9.7.7) to combine the IRRs obtained from the main analyses performed in the cohort studies in the different data sources. The main meta-analysis will be a Poisson regression model with random intervention effects, while the sensitivity analysis will be a Poisson regression model with fixed intervention effects.

9.7.7 Further analysis: combined analysis of incidence rate ratios from the different data sources

If enough events occur in each data source and depending on the results (e.g., the risk estimates do not go in different directions), meta-analytic techniques will be used to combine the IRRs obtained from the main analysis performed in the cohort study in the different data sources. If meta-analysis is performed, it will be to estimate summary IRRs and 95% CIs of the rare outcomes (with possibility of zero events) from three data sources. To account for this situation, Poisson regression models with random intervention effects will be implemented to estimate summary treatment effects across data sources. This approach has been demonstrated to perform particularly well in meta-analyses with zero events, while traditional inverse-variance methods and fixed-effects Poisson regression models have shown evidence of bias and poor coverage in similar situations [R18-3569]. Furthermore, because of potential differences in patient characteristics across data sources, the true effect size may differ between the data sources, which further supports the a priori decision to implement a random-effects approach. Although the fixed-effects Poisson regression approach has been demonstrated to have more potential for bias [R18-3569], it will be implemented as a sensitivity analysis.

9.8 QUALITY CONTROL

Standard operating procedures or guidance documents at the participating institutions will be used to guide the conduct of the study.

At RTI-HS, these procedures include internal quality audits, rules for secure and confidential data storage, methods to maintain and archive project documents, quality-control procedures for programming, standards for writing analysis plans, and requirements for senior scientific review. All programming written by one study analyst will be reviewed independently by a different analyst, with oversight by a senior statistician. All key study documents, such as the analysis plan, abstraction forms, and study reports, will undergo quality-control review, senior scientific review, and editorial review.

At RTI-HS, an independent Office of Quality Assurance (OQA) will perform audits and assessments that involve various aspects of the project, including but not limited to education and training documentation, data entry and data transfer procedures and documentation, and IRB documentation. Such audits will be conducted by the OQA according to established criteria in standard operating procedures and other applicable procedures. Standard procedures will be in place to restore files in the event of a hardware or software failure.

Aarhus University will follow its standard procedures for data management, data analysis, and interpretation and other quality-control and contingency-planning procedures. All reports undergo senior review.

At HealthCore, all programming required for study database extraction and creation of the analytic data sets from the HIRD and analysis are performed in accordance with HealthCore Programming Standards. The HealthCore Programming Standards are a set of documents

describing data extraction methods that are referenced in HealthCore Standard Operating Procedures and provide a guideline for basic, frequently used terms and definitions and respective coding information to maintain operational consistency. These documents contain confidential and proprietary information but can be available for audit if necessary.

At HealthCore, data quality checks include, but are not limited to, programming checks by an individual who is not the main programmer for the study, internal data set consistency, and checks to ensure that protocol criteria were met. The distribution and range of all covariates are examined to verify that they are within the expected range. If any unexpected distributions are identified, or for specific covariates, the raw claims for a sample of not less than 15 patients are reviewed to confirm that the questioned variables have been correctly specified. Additionally, all the diagnosis, procedure, and drug codes are reviewed in the coding scripts provided by a programmer. All quality checks and resolution of issues identified are documented in a log.

A quality-assurance audit of this study may be conducted by the sponsor or the sponsor's designees.

9.9 LIMITATIONS OF THE RESEARCH METHODS

There are several methodological challenges when conducting epidemiologic studies to evaluate the association between glucose-lowering medications and outcomes of interest among patients with diabetes. These challenges include, but are not limited to, changes in treatment in response to advancing diabetes or due to adverse effects of specific drugs and time-varying risk of an outcome depending on duration of exposure [P12-13528, P14-17457, R14-4378].

9.9.1 Confounding

Although use of propensity scores will facilitate the control of measured confounders, unmeasured and unidentified confounders could still introduce bias if they are differentially distributed among the exposed and comparator groups and are related to the outcome. As an example, use of over-the-counter medications will remain unmeasured in this study.

Confounding by indication or severity, also known as channelling bias, is a common bias in pharmacoepidemiology. Patients starting treatment with a newly marketed drug might have more severe disease than patients not taking the medication either because of self-selection or because of physician preference. They may also have a less severe form of the disease if physicians prefer to test new drugs with a less familiar safety profile in less severely affected patients. New medications may also be prescribed differentially by physicians who are "early adopters" of new technologies and who systematically treat more severely affected patients with the new medications. The use of propensity scores and of DPP-4 inhibitors as the comparator group—relative to other GLDs such as sulfonylureas (DPP-4 inhibitors were recently introduced in the market)—reduces the risk of this type of bias, but residual confounding could still operate. For example, empagliflozin could be preferentially prescribed to patients with more severe diabetes or for whom other GLDs have failed.

Empagliflozin could also be more likely prescribed to patients with fewer risk factors for severe complications of UTI. Similarly, the decision for hospitalisation could be affected by the perceived side effects profile. These channelling patterns could bias the risk estimate towards or away from the null.

The occurrence of the outcomes of interest may also vary over time according to the duration or cumulative dose of drug exposure. To account for these time-varying hazards, a stratified analysis estimating the IRRs of the outcomes of interest by categories of cumulative duration of exposure will be conducted.

9.9.2 Other biases

Misclassification bias can occur when study patients are not correctly assigned to the outcome and/or exposure. Because prescribing records will be used, misclassification of exposure is unlikely. However, analyses will not control for non-adherence to the study drugs. Moreover, misclassification of new users could happen if free samples of empagliflozin or DPP-4 inhibitor are provided to patients for different periods of time.

Misclassification of the outcome will be reduced by the validation process that has been planned for each of the outcomes of interest. Among females, misclassification of Fournier's gangrene, a severe complication of GI, is possible because there are no ICD-10 or Read codes for this disease among females.

Empagliflozin will be compared with DPP-4 inhibitors as a group. Information about the risk of the outcomes of interest among patients using specific DPP-4 inhibitors is scarce; therefore, the comparator groups will include specific drugs within each drug class that may or may not have a differential risk for some of the outcomes of interest. This could potentially bias empagliflozin risk estimates for this study.

The main analysis in this study will be performed on data from patients as treated and during current use that will terminate after the days' supply of the last prescription has elapsed plus 30 days. However, if drug discontinuation predicts future outcomes of interest, then an informative censoring bias may occur because we are removing outcomes from their appropriate exposure category. In this study, a sensitivity analysis with varying latency periods after drug discontinuation will evaluate the potential for informative censoring bias.

To address potential informative censoring, an additional sensitivity analysis will be performed by conducting an intention-to-treat analysis, which will carry forward the initial exposure status and disregard changes in treatment status over time.

9.9.3 Strengths and limitations concerning each study database

9.9.3.1 CPRD, UK

Strengths and limitations of the CPRD:

- Data from the primary care database (CPRD GOLD) cover 3.91% of the UK population, and data from CPRD Aurum cover 11.12% of the UK population. However, CPRD GOLD includes practices from England, Scotland, Wales, and Northern Ireland, while CPRD Aurum includes only practices from England (e-mail correspondence, 01 April 2019, on file).
- The CPRD contains information on primary care for all patients and from hospital data for a subset of approximately 54% of patients, although hospital coverage may vary in the future. As of April 2019, CPRD Aurum has linkage of data from 93% of patients with hospital and death data [R19-1734].
- Outpatient laboratory data are available, but no inpatient laboratory data. Inpatient laboratory data can be obtained only if results are recorded in the discharge letter and if the discharge letter is available to the GP at the time of the GP questionnaire.
- Although most of the laboratory results are incorporated in the electronic system automatically, microbiology test results are often missing, but the degree of missingness has not been described. This may have consequences for UTI and GI case validation where microbiology results are needed. The validation of these outcomes will include degrees of certainty that can be included in the analysis.
- Identification of cases of hospitalisation for ALI, AKI, DKA, or GI is possible through linkage to HES. For those patients who are in practices not linkable to HES, identification of hospitalisations can be done through algorithms that identify codes for diagnosis and codes for hospitalisation within a specific time window. The time windows commonly used are 7 days or 30 days [R18-3575]. This study will use a time window of 30 days to increase the sensitivity for detecting hospitalisations, although it may reduce the specificity vs. a time window of 7 days. Recording of hospitalisations in the CPRD primary care is not 100% (i.e., the hospitalisation is not recorded, or the discharge diagnosis is not recorded). In addition, when the hospitalisation is recorded, its association with a diagnosis code is not straight-forward, because there may be other diagnosis codes within the same time window. It is expected that after validation, some of the hospitalisations for ALI, AKI, DKA, or GI will be discarded [R15-4888, R18-3577]. Emergency department visits can only be identified in primary care as recorded by the GP. Although a HES A&E linkage database exists, it is not used because data coverage is incomplete in comparison with national A&E data attendances.
- Validation is possible through questionnaires sent to the GPs, who can access primary care medical records. Source hospital records cannot be accessed, although the GP can access discharge letters recorded in the primary care database.

9.9.3.2 Danish Population Registries

Strengths and limitations of the Danish data sources:

- Data from national registries covers all the Danish population, i.e., includes all age ranges in the population.
- At the national level, information on all reimbursed care is available.
- Primary care outpatient data are not available. Underascertainment of diabetes and some of the comorbidities of interest is possible and likely for conditions that are managed in the primary care setting. However, the study will include data from the outpatient hospital setting, secondary discharge codes, and medication use to define some of the covariates of interest. Definitions will be adapted to each data source depending on data availability and on the existence of previously validated algorithms.
- Source medical records can be accessed for selected projects and with special approvals for studies conducted in the Danish data sources.
- Most potential confounders can be obtained from the national databases, although only hospital-based diagnoses are available. Medication data can be used to identify proxies for diseases such as diabetes [R18-0259].
- Since approximately 2000, the clinical laboratory information system research databases contains results of laboratory tests performed at hospital-based laboratories for the subpopulation of patients treated in the Central Denmark Region, including tests ordered by primary care clinicians from those laboratories [R18-0264]. On a nationwide level, laboratory data for all 5.77 million Danish inhabitants have recently become available for research with complete data coverage beginning in 2015 and onwards in the new nationwide Register of Laboratory Results for Research (LAB_F), tracking all laboratory test results from both primary and secondary care [R18-3665].
- Identification of hospitalised cases of ALI and AKI is possible. No primary care cases can be identified through GP diagnoses, but isolated cases of liver enzyme elevations (e.g., ALT and/or AST $\geq 3 \times$ ULN) or eGFR-defined kidney disease in primary care can be identified through blood samples taken in primary care. For validation, hospital laboratory results can be used in the Central Denmark Region subpopulation since 2000 (e.g., through liver enzyme test values). For the remainder of Denmark's population, laboratory data have become available at the Danish Board of Health starting in 2015, and it is still unclear whether retrieving individual person-identifiable laboratory data for validation purposes on a nationwide scale will therefore be possible.
- Identification of hospitalisations for DKA is possible. For validation, hospital laboratory results in the subpopulation of the Central Denmark Region can be used; for the remainder of the population, hospital laboratory results would have to be extracted from hospital medical records, if available.
- Validation is possible through access to source medical records. This can be done for selected projects and with special approvals for studies conducted in the Danish data sources.

9.9.3.3 HIRD, US

Strengths and limitations of the HIRD:

- A large percentage of the US population is covered in these data sources.
- Patients aged 65 years or older are underrepresented in these data sources. Data are available for those who continue to have insurance through employment after the age of 65 years and for a subset of Medicare patients. The proportion of patients aged 65 years or older will be described for the HIRD and for the other study data sources.
- Access to medical records is possible for many but not all of the cases requiring validation. Medical record retrieval rates have been low for some studies. Redacted copies of hospital medical records are expected to be obtained for approximately 60% of the patients.
- These are claims-based data sources with limited clinical information. Clinical information can be available via linkage to patients' medical records.
- Ascertainment of ALI and AKI: identification of inpatient and outpatient cases of ALI and AKI is possible, although primary care laboratory results are available electronically for approximately 30% of the patients; inpatient laboratory results can be obtained only through abstraction of data from hospital medical records, with an expected retrieval rate of approximately 60%.
- Ascertainment of hospitalisations for DKA: identification of cases is possible, but validation will be limited because inpatient laboratory results can be obtained only through abstraction of data from hospital medical records, with an expected retrieval rate of approximately 60%.
- Claims data may include rule-out diagnoses. To increase sensitivity, for the outcomes of interest, a record for one of the diagnoses of interest will be enough to be considered a potential case. Type 2 diabetes will be defined using ICD-9-CM or ICD-10-CM diagnosis codes from both outpatient and inpatient health care claims, as well as pharmacy dispensing claims for antidiabetic medications. This algorithm was derived from similar algorithms with PPVs greater than 85% for identifying T2D in health care claims data [R18-3576, R16-3197, R18-3667]. For other covariates of interest, algorithms including more than one claim with a diagnosis code or a combination of diagnosis and medication use codes may be required. As discussed in the definitions of the covariates (Section 9.3.3), the definitions will be adapted to each data source depending on data availability and on the existence of previously validated algorithms.

9.9.4 Limitations due to the new-user design definition

A new-user design is used for this study to overcome bias associated with the inclusion of prevalent users and to better distinguish medication effect from T2D-progression effect. However, a new-user design can restrict inclusion (1) to new users of each study medication, or (2) as in this study, to new users of a specific study medication without prior use of any of the other study medication groups.

The "medication class new-user design" used for the second interim report [c17493314-01] used all information available before or at the index date to consider a patient as a new user. As a result, more than 50% of patients in the empagliflozin group and in the SGLT2 inhibitors group were excluded due to prior use of DPP-4 inhibitors. To avoid this loss of patients, the medication class new-user design was maintained but evaluated use of medications only in the previous 12 months. That is, new users were defined as patients who received a first prescription of empagliflozin, other SGLT2 inhibitor, or DPP-4 inhibitor during the study period, with no prescriptions of any of the study drugs recorded in the 12 months (instead of ever) before the start of the study. Based on an exploratory analysis conducted for the second interim report, the number of patients increased by approximately 40%. Analysis in the third interim report with 1 year of additional data identified 1,744 and 1,660 new users of empagliflozin in the ALI and AKI cohorts, respectively.

Bias due to including switchers from DPP-4 inhibitors that may have experienced disease progression will be addressed in the final analysis by considering the inclusion of disease progression variables and prior use of DPP4-inhibitors (e.g., number of prior prescriptions) in the propensity score.

9.9.5 Limitation due to study size

As described in more detail in Section 9.4.1, the CPRD is probably the richest data source to conduct this study since it captures lifestyle factors not available in other data sources. However, the number of patients available for analysis have been monitored during the 3 years after empagliflozin launch in the CPRD. Actual and projected number of users of empagliflozin in the CPRD were low and not enough to address the study objectives. To overcome this limitation, the protocol has been amended to extend the study period and add the Danish Population Registries and the HIRD data sources.

The new projected study size is estimated to be sufficient to address the study objectives, but some of the subgroup analyses may be underpowered. The purpose of using multiple databases is to increase study size and to be able to accrue more users in a shorter period of time; however, analysis at each individual data source may be underpowered for some of the outcomes of interest, especially for ALI and DKA in the CPRD UK and in Denmark. Due to the rarity of some of the outcomes and the study size in some data sources, it is possible that at the time of the analysis, some propensity score strata may contain zero events. In this case, alternative methods, such as matching weights, may be used to reduce the amount of bias.

9.9.6 Generalisability

Generalisations from study findings depend on the category of the finding [R14-2938, R14-1409].

The CPRD GOLD is a population-based database that provides data from practices in England, Scotland, Wales, and Northern Ireland, entered by primary care practitioners in a routine clinical care setting. Therefore, the study results can be generalised to similar patients with T2D in other geographic settings, including most industrialised countries. For the same

reason, because CPRD Aurum is also included, although it provides data only from practices in England, it is also expected that the study results can be generalised to similar patients with T2D in other geographic settings.

The Danish Population Registries cover the full population of Denmark, and thus results from this study are likely to be generalisable to the Danish population of new users of the study medications. It is also likely that results from Denmark can be generalised to most industrialised countries, especially to the other Scandinavian countries with similar health care systems.

The HIRD is a broad, clinically rich and geographically diverse spectrum of longitudinal claims data from 40 million health plan members in the Northeastern, Mid-Atlantic, Southeastern, Midwest, Central, and Western regions of the US. It includes both commercially insured patients aged less than 65 years and Medicare-insured patients aged 65 years or more. However, patients aged 65 years or older are underrepresented in this data source. Data are available for those who continue to have insurance through employment after 65 years of age and for a subset of Medicare patients.

10. PROTECTION OF HUMAN SUBJECTS

This is a non-interventional study using an existing database (secondary data) and does not pose any risks for patients. All data collected in the study will be de-identified with no breach of confidentiality with regard to personal identifiers or health information. RTI-HS will apply for an independent ethics committee review according to local regulations in the UK; in addition, RTI-HS will obtain approval from the RTI International† IRB.

Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants.

10.1 RTI INTERNATIONAL

RTI International holds a Federal-Wide Assurance from the Department of Health and Human Services Office for Human Research Protections that allows the organisation to review and approve human subjects protocols through its IRB committees. RTI International currently has three IRB committees available to review research protocols. One IRB committee is constituted to review medical research and has two members who are physicians. These IRBs have been audited by the US Food and Drug Administration and are fully compliant with applicable regulatory requirements.

10.2 CPRD

RTI-HS will submit the final study protocol for approval to the Independent Scientific Advisory Committee (ISAC) (http://www.cprd.com/ISAC). The CPRD has obtained ethical approval from a Multicentre Research Ethics Committee for all observational research using

† RTI Health Solutions is a business unit of RTI International, a private, not-for-profit research organisation.

CPRD data without patient involvement; however, ISAC may recommend that the Multicentre Research Ethics Committee review the study documentation if any ethical issues arise. The possibility of unintentional (deductive) disclosure arises when cells with small numbers of patients are quoted. When reporting the data, CPRD policy is that no cell should contain less than 5 events.

10.3 THE DANISH POPULATION REGISTRIES

The study must be approved by the Danish Data Protection Agency. Access to medical charts will be approved by the Patient Safety Board, following all standard and required procedures for such approval. The data will be handled according to the Danish Act on Processing of Personal Data. The possibility of unintentional (deductive) disclosure arises when cells with small numbers of patients are quoted. When reporting the data, Danish policy is that no cell should contain less than 5 events.

10.4 HIRD

This study is based on medical and pharmacy claims data from a large insured population in the US. There is no active enrolment or active follow-up of study subjects, and no data will be collected directly from individuals.

HealthCore maintains Data Sharing Agreements and Business Associate Agreements with all covered entities who provide data to the HIRD. HealthCore's access, use, and disclosure of protected health information (PHI) are in compliance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (45 CFR Part 160 and Subparts A and E of Part 164). HealthCore does not access, use, or disclose identifiable PHI unless under a specific waiver of authorisation (e.g., a HIPAA Waiver of Authorization from an institutional review board [IRB]). HealthCore accesses the data in a manner that complies with federal and state laws and regulations, including those related to the privacy and security of individually identifiable health information.

As PHI must be accessed to acquire medical records to validate electronic case-finding algorithms, a HIPAA Waiver of Authorization will be applied for from an IRB. HealthCore will submit the protocol to a central IRB for review and approval. Approval is typically provided within 2 to 3 weeks of submission. Once IRB approval is obtained, HealthCore's vendor will proceed with the conduct of medical record acquisition. If changes to the protocol are required, HealthCore will submit an amendment to the IRB. As the IRB is independent, HealthCore cannot control the approval or whether there are conditions for the approval.

Notwithstanding receipt of approval from a central IRB, in some instances, individual institutions may require approval from their local IRB, which would require a separate protocol submission and, in some cases, additional fees. In these cases, HealthCore, RTI Health Solutions (RTI-HS), and Boehringer Ingelheim will need to agree whether or not to proceed with chart acquisition from these institutions.

HealthCore will provide to the vendor only the minimum amount of patient information that is necessary to acquire the needed medical records. HealthCore uses only vendors that follow federal and state laws and regulations, including but not limited to privacy and security rules such as HIPAA.

At no time during the conduct of this study will HealthCore provide patient- or provider-identifying information to RTI-HS or Boehringer Ingelheim. Only aggregated data will be reported to RTI-HS or Boehringer Ingelheim. The possibility of unintentional (deductive) disclosure arises when cells with small numbers of patients are quoted. When reporting the data, HIRD policy is that no cell should contain less than 10 events.

10.5 OTHER GOOD RESEARCH PRACTICE

This study adheres to the *Guidelines for Good Pharmacoepidemiology* Practices (GPP) of the International Society for Pharmacoepidemiology [R11-4318] and has been designed in line with the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) *Guide on Methodological Standards in Pharmacoepidemiology* [R14-5282]. The ENCePP *Checklist for Study Protocols* [R13-1395] is included in Annex 2.

The study is a PASS and will comply with the definition of the non-interventional (observational) study referred to in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use tripartite guideline *Pharmacovigilance Planning* E2E [R11-2259] and provided in the EMA *Guideline on Good Pharmacovigilance Practices* (GVP) *Module VIII: Post-Authorisation Safety Studies* [R13-5420] and with the 2012 European Union pharmacovigilance legislation, adopted 19 June 2012 [R14-5246]. Following EMA regulations, the study protocol has been registered at the EU PAS Register, and an abstract of results will be registered at the EU PAS Register (registration number: ENCEPP/SDPP/13413; http://www.encepp.eu/encepp/viewResource.htm?id=13414).

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 milestones are agreed with the EMA. The study progress was reported by BI in regulatory communications in line with the risk management plan, Periodic Safety Update Reports, and other regulatory milestones and requirements. Study reports are prepared using a template following the *Guideline on Good Pharmacovigilance Practices (GVP), Module VIII*, Section B.6.3 [R13-5420]. The planned periodic interim reports at 12, 24, and 36 months were reported within the earliest corresponding Periodic Safety Update Report.

Section V of *Guidelines for Good Pharmacoepidemiology Practices (GPP)* [R11-4318] contends that "there is an ethical obligation to disseminate findings of potential scientific or public health importance"; for example, results pertaining to the safety of a marketed medication. Moreover, a well-developed publication strategy is encouraged in the *Guideline on Good Pharmacovigilance Practices, Module VIII*, Section B.7 [R13-5420].

RTI Health Solutions, Aarhus, and HealthCore reserve the right to submit the results from any of the study analyses for publication and commits that at least the final results will be published. Any publications will follow guidelines, including those for authorship, established by the International Committee of Medical Journal Editors [R13-5418]. When reporting results of this study, the appropriate STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist will be followed [R13-2485].

13. REFERENCES

13.1 PUBLISHED REFERENCES

P02-05226 Griffin MR, Yared A, Ray WA. Nonsteroidal antiinflammatory drugs and acute renal failure in elderly persons. Am J Epidemiol 2000; 151 (5): 488-96.

P02-05969 Sgro C, Clinard F, Ouazir K, Chanay H, Allard C, Guilleminet C, et al. Incidence of drug-induced hepatic injuries: a French population-based study. Hepatology 2002; 36 (2): 451-5.

P03-00488 Ibáñez L, Pérez E, Vidal X, Laporte JR, Grup d'Estudi Multicènteric d'Hepatotoxicitat Aguda de Barcelona. Prospective surveillance of acute serious liver disease unrelated to infectious, obstructive, or metabolic diseases: epidemiological and clinical features, and exposure to drugs. J Hepatol 2002; 37 (5): 592-600.

P03-03701 Huerta C, Zhao SZ, García Rodríguez LA. Risk of acute liver injury in patients with diabetes. Pharmacotherapy. 2002 Sep; 22(9):1091-6.

P04-07683 de Abajo FJ, Montero D, Madurga M, García Rodríguez LA. Acute and clinically relevant drug-induced liver injury: a population based case-control study. Br J Clin Pharmacol. 2004 Jul; 58 (1): 71-80.

P05-04032 Huerta C, Castellsague J, Varas-Lorenzo C, García Rodríguez LA. Nonsteroidal anti-inflammatory drugs and risk of ARF in the general population. Am J Kidney Dis 2005; 45 (3): 531-9.

P05-08822 Andrade RJ, Lucena MI, Fernández MC, Pelaez G, Pachkoria K, García-Ruiz E, et al. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 2005 Aug; 129 (2): 512-21.

P06-02059 Navarro VJ, Senior JR. Drug-related hepatotoxicity. N Engl J Med. 2006; 354 (7): 731-9.

P06-11008 De Valle MB, Av Klinteberg V, Alem N, Olsson R, Björnsson E. Drug-induced liver injury in a Swedish University hospital out-patient hepatology clinic. Aliment Pharmacol Ther 2006; 24 (8): 1187-95.

P09-12413 FDA. Guidance for industry. Drug-induced liver injury: premarketing clinical evaluation. Silver Spring (MD): US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research(CBER); July 2009.

website: https://www.fda.gov/downloads/Guidances/UCM174090.pdf (access date: 18 April 2018).

P12-04844 Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 2006; 98 (3): 253-9.

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.

P14-02878 Geerlings S, Fonseca V, Castro-Diaz D, List J, Parikh S. Genital and urinary tract infections in diabetes: impact of pharmacologically-induced glucosuria. Diabetes Res Clin Pract 2014; 103 (3): 373-81.

P14-10229 Kaye JA, Castellsague J, Bui CL, Calingaert B, McQuay LJ, Riera-Guardia N, et al. Risk of acute liver injury associated with the use of moxifloxacin and other oral antimicrobials: a retrospective, population-based cohort study. Pharmacotherapy 2014; 34 (4): 336-49.

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

P14-17370 Czaja CA, Scholes D, Hooton TM, Stamm WE. Population-based epidemiologic analysis of acute pyelonephritis. Clin Infect Dis 2007; 45 (3): 273-80.

P14-17372 Pérez Gutthann S, García Rodríguez LA, Raiford DS, Duque Oliart A, Ris Romeu J. Nonsteroidal anti-inflammatory drugs and the risk of hospitalization for acute renal failure. Arch Intern Med 1996; 156 (21): 2433-9.

P14-17373 Goettsch WG, Heintjes EM, Kastelein JJ, Rabelink TJ, Johansson S, Herings RM. Results from a rosuvastatin historical cohort study in more than 45,000 Dutch statin users, a PHARMO study. Pharmacoepidemiol Drug Saf 2006; 15 (7): 435-43.

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

P14-17375 Temple RJ. Hepatotoxicity through the years: impact on the FDA (presentation). 2001. website: http://www.fda.gov/downloads/Drugs/ScienceResearch/ResearchAreas/ucm122149.pdf (access date: 09 October 2014).

P14-17376 García-Rodríguez LA, Massó-González EL, Wallander MA, Johansson S. The safety of rosuvastatin in comparison with other statins in over 100,000 statin users in UK primary care. Pharmacoepidemiol Drug Saf 2008; 17 (10): 943-52.

P14-17377 García Rodríguez LA, González-Pérez A, Stang MR, Wallander MA, Johansson S. The safety of rosuvastatin in comparison with other statins in over 25,000 statin users in the Saskatchewan Health Databases. Pharmacoepidemiol Drug Saf 2008; 17 (10): 953-61.

P14-17413 Gallagher AM, Smeeth L, Seabroke S, Leufkens HG, van Staa TP. Risk of death and cardiovascular outcomes with thiazolidinediones: a study with the general practice research database and secondary care data. PLoS One 2011; 6 (12): e28157.

P14-17456 EMA. Assessment report: Jardiance. EMA/CHMP/137741/2014. London: European Medicines Agency, Committee for Medicinal Products for Human Use; 20 March 2014. Available at: website: ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002677/WC500168594.pdf. Accessed 09 October 2014.

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

P15-00346 Hay JE. Liver disease in pregnancy. Hepatology. 2008 Mar; 47 (3): 1067-76.

P15-08785 Rosenstock J, Ferrannini E. Euglycemic diabetic ketoacidosis: a predictable, detectable, and preventable safety concern with SGLT2 inhibitors. Diabetes Care. 2015 Sep;38(9):1638-42.

P15-11541 Roach P, Skierczynski P. Euglycemic diabetic ketoacidosis in a patient with type 2 diabetes after treatment with empagliflozin. Diabetes Care. 2016 Jan;39(1):e3.

P16-03420 European Commission. SGLT2 inhibitors and diabetic ketoacidosis. Notification to the PRAC/EMA Secretariat of a referral under Article 20 of Regulation (EC) 726/2004. 10 June 2015. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Referrals_document/SGLT2_inhibitors__20/Procedure_started/WC500187925.pdf. Accessed 18 March 2016.

P16-10382 Mor A, Berencsi K, Nielsen JS, Rungby J, Friborg S, Brandslund I, et al. Rates of community-based antibiotic prescriptions and hospital-treated infections in individuals with and without type 2 diabetes: a Danish nationwide cohort study, 2004-2012. Clin Infect Dis 2016; 63 (4): 501-11.

P17-09718 Nadkarni GN, Ferrandino R, Chang A, Surapaneni A, Chauhan K, Poojary P, et al. Acute kidney injury in patients on SGLT2 inhibitors: a propensity-matched analysis. Diabetes Care. 2017; 40(11): 1479-85.

P18-00901 Zhang XL, Zhu QQ, Chen YH, Li XL, Chen F, Huang JA, et al. Cardiovascular safety, long-term noncardiovascular safety, and efficacy of sodium-glucose cotransporter 2 inhibitors in patients with type 2 diabetes mellitus: a systemic review and meta-analysis with trial sequential analysis. J Am Heart Assoc. 2018; 7(2).

P18-01920 American Diabetes Association. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes 2018. Diabetes Care 2018; 41 (Suppl. 1): S73–S85.

P18-11001 Franklin JM, Eddings W, Austin PC, Stuart EA, Schneeweiss S. Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. Stat Med. 2017 May 30;36(12):1946-1963.

P19-04515 Patorno E, Pawar A, Franklin JM, Najafzadeh M, Déruaz-Luyet A, Brodovicz KG, et al. Empagliflozin and the risk of heart failure hospitalization in routine clinical care: a first analysis from the empagliflozin comparative effectiveness and safety (EMPRISE) study [supplemental material]. Circulation. 2019 Apr 8. doi: 10.1161/CIRCULATIONAHA.118.039177. [Epub ahead of print].

R03-0585 UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. BMJ 1998; 317 (7160): 703-13.

R05-1093 Zimmerman HJ. Drug-induced liver disease. In: Hepatotoxicity, the adverse effects of drugs and other chemicals on the liver, 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1999. p. 428-33.

R07-2456 Arnold AM, Kronmal RA. Multiple Imputation of Baseline Data in the Cardiovascular Health Study. Am J Epidemiol 2003; 157 (1): 74-84.

R08-1486 Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. Ann Intern Med. 2002; 137 (8): 693-5.

R08-1494 Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol 2003; 158 (3): 280-7.

R10-6632 Shah BR, Hux JE. Quantifying the risk of infectious diseases for people with diabetes. Diabetes Care 2003; 26 (2): 510-3.

R11-2162 Jick SS, Kaye JA, Vasilakis-Scaramozza C, García Rodríguez LA, Ruigómez A, Meier CR, et al. Validity of the general practice research database. Pharmacotherapy 2003; 23 (5): 686-9.

R11-2259 ICH. Pharmacovigilance planning. E2E. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2004. Available at: website: ich.org/products/guidelines/efficacy/efficacy-single/article/pharmacovigilance-planning.html. Accessed 7 February 2013.

R11-4318 ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Revision 2. International Society for Pharmacoepidemiology; 2007. Available at: website: pharmacoepi.org/resources/guidelines_08027.cfm. Accessed 07 February 2013.

R11-5210 Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010; 60 (572): e128-36.

R11-5211 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010; 69 (1): 4-14.

R11-5319 Girman CJ, Kou TD, Brodovicz K, Alexander CM, O'Neill EA, Engel S, et al. Risk of acute renal failure in patients with type 2 diabetes mellitus. Diabet Med 2012; 29 (5): 614-21.

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.

R11-5329 Xue JL, Daniels F, Star RA, Kimmel PL, Eggers PW, Molitoris BA, et al. Incidence and mortality of acute renal failure in Medicare beneficiaries, 1992 to 2001. J Am Soc Nephrol 2006; 17 (4): 1135-42.

R12-1080 Benfield T, Jensen JS, Nordestgaard BG. Influence of diabetes and hyperglycaemia on infectious disease hospitalisation and outcome. Diabetologia 2007; 50 (3): 549-54.

R12-1083 Boyko EJ, Fihn SD, Scholes D, Abraham L, Monsey B. Risk of urinary tract infection and asymptomatic bacteriuria among diabetic and nondiabetic postmenopausal women. Am J Epidemiol 2005; 161 (6): 557-64.

R12-1088 Joshi N, Caputo GM, Weitekamp MR, Karchmer AW. Infections in patients with diabetes mellitus. N Engl J Med 1999; 341 (25): 1906-12.

R12-1100 Muller LM, Gorter KJ, Hak E, Goudzwaard WL, Schellevis FG, Hoepelman AI, et al. Increased risk of common infections in patients with type 1 and type 2 diabetes mellitus. Clin Infect Dis 2005; 41 (3): 281-8.

R12-1105 Venmans LM, Hak E, Gorter KJ, Rutten GE. Incidence and antibiotic prescription rates for common infections in patients with diabetes in primary care over the years 1995 to 2003. Int J Infect Dis 2009; 13 (6): e344-51.

R12-1392 Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, Kusek JW, Eggers P, Lente F van, Greene T, Coresh J, Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009; 150 (9): 604-612.

R12-1479 Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR, Group US. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. Diabetes 2006; 55 (6): 1832-9.

R12-1620 EMA. European Medicines Agency recommends suspension of Avandia, Avandamet and Avaglim (press release). 23 September 2010. Available at: website: https://www.ema.europa.eu/en/news/european-medicines-agency-recommends-suspension-avandia-avandamet-avaglim. Accessed 09 October 2014.

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

R12-2432 de Leon EM, Jacober SJ, Sobel JD, Foxman B. Prevalence and risk factors for vaginal Candida colonization in women with type 1 and type 2 diabetes. BMC Infect Dis 2002; 2: 1.

R12-3632 El-Serag HB, Everhart JE. Diabetes increases the risk of acute hepatic failure. Gastroenterology. 2002 Jun; 122 (7): 1822-8.

R12-3639 Hirji I, Andersson SW, Guo Z, Hammar N, Gomez-Caminero A. Incidence of genital infection among patients with type 2 diabetes in the UK General Practice Research Database. J Diabetes Complications 2012; 26 (6): 501-5.

R12-5226 Hirji I, Guo Z, Andersson SW, Hammar N, Gomez-Caminero A. Incidence of urinary tract infection among patients with type 2 diabetes in the UK General Practice Research Database (GPRD). J Diabetes Complications 2012; 26 (6): 513-6.

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

R13-1395 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.

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.

R13-2485 von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008; 61 (4): 344-9.

R13-3433 Holden SH, Barnett AH, Peters JR, Jenkins-Jones S, Poole CD, Morgan CL, et al. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. Diabetes Obes Metab. 2013 Sep;15(9):844-52.

R13-3589 Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. Med Care 2004; 42 (4): 355-60.

R13-3591 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998 Jan; 36 (1): 8-27.

R13-4387 Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013; 3(1): 1-150.

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

R13-5418 ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors; August 2013. Available at: website: icmje.org/urm_main.html. Accessed 25 February 2014.

R13-5420 EMA. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies. European Medicines Agency; 9 July 2012. Available at: website:

emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf. Accessed 7 February 2013.

R14-1393 Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990 Jan; 1 (1): 43-6.

R14-1409 Brunelli SM, Gagne JJ, Huybrechts KF, Wang SV, Patrick AR, Rothman KJ, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. Pharmacoepidemiol Drug Saf. 2013 May;22(5):542-50.

R14-1933 Aithal GP, Watkins PB, Andrade RJ, Larrey D, Molokhia M, Takikawa H, et al. Case definition and phenotype standardization in drug-induced liver injury. Clin Pharmacol Ther. 2011 Jun; 89 (6): 806-15.

R14-2938 Rothman KJ. Six persistent research misconceptions. J Gen Intern Med. 2014 Jul; 29 (7): 1060-4.

R14-3272 Wang ZH, Kihl-Selstam E, Eriksson JW. Ketoacidosis occurs in both type 1 and type 2 diabetes—a population-based study from northern Sweden. Diabet Med. 2008 Jul;25(7):867-70.

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

R14-4617 Boehringer Ingelheim International GmbH. Jardiance (empagliflozin) summary of product characteristics. 2014. Available at: website: https://www.ema.europa.eu/en/medicines/human/EPAR/jardiance#product-information-section. Accessed 09 October 2014.

R14-5236 Boyko EJ, Fihn SD, Scholes D, Chen CL, Normand EH, Yarbro P. Diabetes and the risk of acute urinary tract infection among postmenopausal women. Diabetes Care 2002; 25 (10): 1778-83.

R14-5237 Bohannon NJ. Treatment of vulvovaginal candidiasis in patients with diabetes. Diabetes Care 1998; 21 (3): 451-6.

R14-5241 Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. Am J Epidemiol 2011; 174 (5): 613-20.

R14-5242 Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P, Acute Dialysis Quality Initiative working group. Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second

International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. Crit Care 2004; 8 (4): R204-12.

R14-5243 Edwards S. Balanitis and balanoposthitis: a review. Genitourin Med. 1996 Jun; 72 (3): 155-9.

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

R14-5246 European Commission. Commission implementing Regulation (EU) No 520/2012 of 19 June 2012 on the performance of pharmacovigilance activities provided for in Regulation (EC) No 726/2004 of the European Parliament and of the Council and Directive 2001/83/EC of the European Parliament and of the Council. 2012. Available at: website: eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri = OJ:L:2012:159:0005:0025:EN:PDF. Accessed 25 February 2014.

R14-5247 Faculty of Sexual and Reproductive Healthcare. Management of vaginal discharge in non-genitourinary medicine settings. London (UK): Faculty of Sexual and Reproductive Healthcare, British Association for Sexual Health and HIV; 28 February 2012. Available at: website: guideline.gov/content.aspx?id = 36683. Accessed 09 October 2014.

R14-5248 Lisboa C, Santos A, Dias C, Azevedo F, Pina-Vaz C, Rodrigues A. Candida balanitis: risk factors. J Eur Acad Dermatol Venereol 2010; 24 (7): 820-6.

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.

R14-5251 Nicolle LE, Friesen D, Harding GK, Roos LL. Hospitalization for acute pyelonephritis in Manitoba, Canada, during the period from 1989 to 1992; impact of diabetes, pregnancy, and aboriginal origin. Clin Infect Dis 1996; 22 (6): 1051-6.

R14-5252 Patel DA, Burnett NM, Curtis KM. Reproductive tract infections. Reproductive Health Epidemiology Series—Module 3. Atlanta (GA): Centers for Disease Control and Prevention; 2003. Available at: website: cdc.gov/reproductivehealth/ProductsPubs/PDFs/Epi_Module_03A_Tag508.pdf. Accessed 09 October 2014.

R14-5253 Patkar NM, Curtis JR, Teng GG, Allison JJ, Saag M, Martin C, et al. Administrative codes combined with medical records based criteria accurately

identified bacterial infections among rheumatoid arthritis patients. J Clin Epidemiol. 2009 Mar; 62 (3): 321-7, 7 e1-7.

R14-5254 Peer AK, Hoosen AA, Seedat MA, van den Ende J, Omar MA. Vaginal yeast infections in diabetic women. S Afr Med J 1993; 83 (10): 727-9.

R14-5255 Rubino A, McQuay LJ, Gough SC, Kvasz M, Tennis P. Delayed initiation of subcutaneous insulin therapy after failure of oral glucose-lowering agents in patients with type 2 diabetes: a population-based analysis in the UK. Diabet Med 2007; 24 (12): 1412-8.

R14-5256 Temple R. Hy's law: predicting serious hepatotoxicity. Pharmacoepidemiol Drug Saf 2006; 15 (4): 241-3.

R14-5258 Foxman B, Klemstine KL, Brown PD. Acute pyelonephritis in US hospitals in 1997: hospitalization and in-hospital mortality. Ann Epidemiol. 2003; 13 (2): 144-50.

R14-5259 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Am J Med 2002; 113 Suppl 1A: 5S-13S.

R14-5260 Fakjian N, Hunter S, Cole GW, Miller J. An argument for circumcision. Prevention of balanitis in the adult. Arch Dermatol 1990; 126 (8): 1046-7.

R14-5279 Gelfand JM, Margolis DJ, Dattani H. The UK General Practice Research Database. In: Strom BL, editor. Pharmacoepidemiology, 4th ed: John Wiley & Sons; 2005. p. 337-46.

R14-5280 Van Staa T-P, Abenhaim L. The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycemia and other conditions. Pharmacoepidemiol Drug Saf 1994; 3, 15-21.

R14-5281 Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338: b2393.

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.

R14-5283 Brown JS, Wessells H, Chancellor MB, Howards SS, Stamm WE, Stapleton AE, et al. Urologic complications of diabetes. Diabetes Care 2005; 28 (1): 177-85.

R14-5284 D'Agostino RB, Jr. Propensity scores in cardiovascular research. Circulation 2007; 115 (17): 2340-3.

R14-5285 Waikar SS, Curhan GC, Wald R, McCarthy EP, Chertow GM. Declining mortality in patients with acute renal failure, 1988 to 2002. J Am Soc Nephrol 2006; 17 (4): 1143-50.

R14-5286 Wagenlehner FM, Pilatz A, Naber KG, Weidner W. Therapeutic challenges of urosepsis. Eur J Clin Invest 2008; 38 Suppl 2: 45-9.

R14-5287 Gaieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. Crit Care Med 2013; 41 (5): 1167-74.

R14-5373 Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. Dordrecht; New York: Springer; 2009.

R14-5389 Segal JB, Griswold M, Achy-Brou A, Herbert R, Bass EB, Dy SM, et al. Using propensity scores subclassification to estimate effects of longitudinal treatments: an example using a new diabetes medication. Med Care 2007; 45 (10 Supl 2): S149-S

R15-0067 Edwards S, Bunker C, Ziller F, van der Meijden WI. 2013 European guideline for the management of balanoposthitis. Int J STD AIDS. 2014 May 14; 25 (9): 615-26.R13-3591

R15-2053 Barski L, Nevzorov R, Harman-Boehm I, Jotkowitz A, Rabaev E, Zektser M, et al. Comparison of diabetic ketoacidosis in patients with type-1 and type-2 diabetes mellitus. Am J Med Sci. 2013 Apr;345(4):326-30.

R15-2058 Kitabchi AE, Umpierrez GE, Murphy MB, Kreisberg RA. Hyperglycemic crises in adult patients with diabetes: a consensus statement from the American Diabetes Association. Diabetes Care. 2006 Dec;29(12):2739-48.

R15-3133 Bolíbar B, Fina Avilés F, Morros R, Garcia-Gil Mdel M, Hermosilla E, Ramos R, et al. [SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research]. Med Clin (Barc) 2012; 138 (14): 617-21.

R15-3134 Aitken GR, Roderick PJ, Fraser S, Mindell JS, O'Donoghue D, Day J, et al. Change in prevalence of chronic kidney disease in England over time: comparison of nationally representative cross-sectional surveys from 2003 to 2010. BMJ Open 2014; 4 (9): e005480.

R15-3135 Danish Data Protection Agency. Introduction to the Danish Data Protection Agency. 2014. website: http://www.datatilsynet.dk/english/ (access date: 16 June 2015).

R15-3136 Denburg MR, Haynes K, Shults J, Lewis JD, Leonard MB. Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. Pharmacoepidemiol Drug Saf 2011; 20 (11): 1138-49.

R15-3137 Danish Health and Medicines Authority. Epidemiology. 2013. website: http://www.ssi.dk/English/RandD/Research%20areas/Epidemiology.aspx (access date: 16 June 2015).

R15-3138 Fleet JL, Dixon SN, Shariff SZ, Quinn RR, Nash DM, Harel Z, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. BMC Nephrol 2013; 14: 81.

R15-3139 Jameson K, Jick S, Hagberg KW, Ambegaonkar B, Giles A, O'Donoghue D. Prevalence and management of chronic kidney disease in primary care patients in the UK. Int J Clin Pract 2014; 68 (9): 1110-21.

R15-3140 Johannesdottir SA, Horváth-Puhó E, Ehrenstein V, Schmidt M, Pedersen L, Sørensen HT. Existing data sources for clinical epidemiology: the Danish National Database of Reimbursed Prescriptions. Clin Epidemiol 2012; 4: 303-13.

R15-3141 Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. Scand J Public Health 2011; 39 (7 Suppl): 38-41.

R15-3142 SIDIAP. Database website. General details. 2014. website: http://www.sidiap.org/index.php?option=com_content&view=article&id=61&Itemid=11&lang=en (access date: 3 June 2014).

R15-4888 Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, Timmis A, Hemingway H. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. 2013 May 20;346:f2350.

R15-5270 Levey AS, Becker C, Inker LA. Glomerular filtration rate and albuminuria for detection and staging of acute and chronic kidney disease in adults a systematic review. JAMA 2015; 313(8):837-846.

R15-5707 Erondu N, Desai M, Ways K, Meininger G. Diabetic Ketoacidosis and Related Events in the Canagliflozin Type 2 Diabetes Clinical Program. Diabetes Care. 2015 Sep;38(9):1680-6.

R16-1231 Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827-36.

R16-1371 Gosmanov AR, Gosmanova EO, Dillard-Cannon E. Management of adult diabetic ketoacidosis. Diabetes Metab Syndr Obes. 2014;7:255-64.

R16-1372 Joint British Diabetes Societies, Inpatient Care Group. The management of diabetic ketoacidosis in adults. 2013. Available at: https://www.diabetes.org.uk/Documents/About%20Us/What%20we%20say/Management-of-DKA-241013.pdf. Accessed 16 March 2016.

R16-1373 Savage MW, Dhatariya KK, Kilvert A, Rayman G, Rees JA, Courtney CH, et al. Joint British Diabetes Societies guideline for the management of diabetic ketoacidosis. Diabet Med. 2011 May;28(5):508-15.

R16-1374 The NHS Information Centre. National Diabetes Audit executive summary 2009-2010. 2011. Available at: http://www.hscic.gov.uk/catalogue/PUB02573/nati-diab-audi-09-10-exec-summ.pdf. Accessed 17 March 2016.

R16-1375 Westphal SA. The occurrence of diabetic ketoacidosis in non-insulin-dependent diabetes and newly diagnosed diabetic adults. Am J Med. 1996 Jul;101(1):19-24.

R16-1981 Fishbein H, Palumbo P. Acute metabolic complications in diabetes. Diabetes in America. 1995;2:283-92.

R16-2604 Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014; 29 (8): 541-9.

R16-3197 Zhong VW, Pfaff ER, Beavers DP, Thomas J, Jaacks LM, Bowlby DA, et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. Pediatr Diabetes 2014; 15: 573-84.

R16-4879 Henriksen OM, Røder ME, Prahl JB, Svendsen OL. Diabetic ketoacidosis in Denmark Incidence and mortality estimated from public health registries. Diabetes Res Clin Pract. 2007 Apr; 76 (1): 51-6.

R18-0255 Desai VC, Quinlan SC, Deitz AC, He J, Holick CN, Lanes S. Risk of severe acute liver injury among patients with brain cancer treated with temozolomide: a nested case-control study using the HealthCore Integrated Research Database. J Neurooncol 2017; 134 (1): 89-95.

R18-0256 CDC. National Death Index (factsheet). Atlanta, GA: National Center for Health Statistics, Centers for Disease Control and Prevention; 2015. website: http://www.cdc.gov/nchs/data/factsheets/factsheet_ndi.htm (access date: 02 November 2017).

R18-0259 Thomsen RW, Sørensen HT. Using registries to identify type 2 diabetes patients. Clin Epidemiol. 2014; 7: 1-3.

R18-0262 Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015; 7: 449-90.

R18-0263 Sørensen HT, Christensen T, Schlosser HK, Pedersen L, eds. Use of medical databases in clinical epidemiology. 2nd ed. Department of Clinical Epidemiology, Aarhus University Hospital: Aarhus, Denmark. 2009. website: https://www.prhdn.ca/SiteAssets/Registerbog%202ed.pdf (access date: 29 December 2017).

R18-0264 Grann AF, Erichsen R, Nielsen AG, Frøslev T, Thomsen RW. Existing data sources for clinical epidemiology: the clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. Clin Epidemiol. 2011; 3: 133-8.

R18-0349 CPRD 2017. CPRD Aurum. Frequently asked questions (FAQs). website: https://www.cprd.com/_docs/CPRD%20Aurum%20FAQs.pdf (access date: 02 November 2017).

R18-0350 Lo Re V, Carbonari DM, Saine ME, Newcomb CW, Roy JA, Liu Q, et al. Postauthorization safety study of the DPP-4 inhibitor saxagliptin: a large-scale multinational family of cohort studies of five outcomes. BMJ Open Diabetes Res Care. 2017; 5 (1): e000400.

R18-0353 Eurostat. Population on 1 January. 2017. European Commission. Available at: website: http://ec.europa.eu/eurostat/documents/2995521/8102195/3-10072017-AP-EN.pdf/a61ce1ca-1efd-41df-86a2-bb495daabdab. Accessed 02 November 2017.

R18-067 Thomsen RW, Nicolaisen SK, Hasvold P, Sanchez RG, Pedersen L, Adelborg K, et al. Elevated potassium levels in patients with chronic kidney disease: occurrence, risk factors and clinical outcomes—a Danish population-based cohort study. Nephrol Dial Transplant. 2017 Nov 21. doi: 10.1093/ndt/gfx312.

R18-1213 Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF Project. PLoS One. 2016 Aug 31; 11 (8): e0160648.

R18-1214 Mack CD, Glynn RJ, Brookhart MA, Carpenter WR, Meyer AM, et al. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. Pharmacoepidemiol Drug Saf. 2013 Aug; 22 (8): 10-8.

R18-1215 Fralick M, Schneeweiss S, Patorno E. Risk of diabetic ketoacidosis after initiation of an SGLT2 inhibitor. N Engl J Med. 2017; 376 (23): 2300-2302.

R18-1217 Zhong VW, Juhaeri J, Mayer-Davis EJ. Trends in hospital admission for diabetic ketoacidosis in adults with type 1 and type 2 diabetes in England, 1998-2013: a retrospective cohort study. Diabetes Care. 2018; 41 (9): 1870-7.

R18-1554 Diabetesforeningen. Diabetes in Denmark [in Danish]. 06 April 2018. Available at: website: https://diabetes.dk/presse/diabetes-i-tal/diabetes-i-danmark.aspx. Accessed 13 April 2018.

R18-1556 Thomsen RW, Nicolaisen SK, Svensson E, Hasvold P, Sørensen HT: Hyperkalaemia in patients with diabetes: incidence, risk factors, and clinical outcomes. A Danish population based cohort study. EASD, Lisbon, 2017, Oral presentation # 238.

R18-1557 Liu L, Simon B, Shi J, Mallhi AK, Eisen HJ. Impact of diabetes mellitus on risk of cardiovascular disease and all-cause mortality: evidence on health outcomes and antidiabetic treatment in United States adults. World J Diabetes, 2016. 7(18): 449-461.

R18-1558 Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. JAMA, 2015; 314(10): 1021-9.

R18-1559 Christensen DH, Rungby J, Thomsen RW. Nationwide trends in glucose-lowering drug use, Denmark, 1999-2014. Clin Epidemiol. 2016 Oct 14;8:381-387.

R18-1560 American Diabetes Association. Statistics about diabetes. 22 March 2018. Available at: website: http://www.diabetes.org/diabetes-basics/statistics/. Accessed 13 April 2018.

R18-1561 Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. Am J Epidemiol. 1993 Dec 1;138(11):1007-15.

R18-3569 Spittal MJ, Pirkis J, Gurrin LC. Meta-analysis of incidence rate data in the presence of zero events. BMC Med Res Methodol. 2015; 15: 42.

R18-3575 Saine ME, Gizaw M, Carbonari DM, Newcomb CW, Roy JA, Cardillo S, Esposito DB, Bhullar H, Gallagher AM, Strom BL, Lo Re V 3rd. Validity of diagnostic codes to identify hospitalizations for infections among patients treated with oral anti-diabetic drugs. Pharmacoepidemiol Drug Saf. 2018 Oct;27(10):1147-1150. doi: 10.1002/pds.4368. Epub 2017 Dec 18. PubMed PMID: 29250905.

R18-3576 Southern DA, Roberts B, Edwards A, Dean S, Norton P, Svenson LW, et al. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. Can J Public Health 2010;101:61-64.

R18-3577 Rothnie KJ, Müllerová H, Thomas SL, Chandan JS, Smeeth L, Hurst JR, Davis K, Quint JK. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. Clin Epidemiol. 2016 Nov 21;8:771-782.

R18-3665 Sundhedsdatastyrelsen. Laboratoriedatabasen website [Laboratory database; in Danish]. 21 June 2018. website: https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/doedsaarsager-og-biologisk-materiale/laboratoriedatabasen (accessed 10 November 2018).

R18-3666 CPRD. Set 16 linkage release [e-mail communication]. 25 June 2018.

R18-3667 Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. Am J Med Qual 2006;21:238-245.

R19-1734 CPRD. Set 17 linkage release [e-mail communication]. 22 May 2019.

R19-1781 Clinical Practice Research Datalink (CPRD). Release notes: CPRD GOLD April 2019 (data on file). 2019.

R19-1782 Clinical Practice Research Datalink (CPRD). Release notes: CPRD Aurum April 2019 (data on file). 2019.

R19-1785 Dokumentation af Laboratoriedatabasens Forskertabel – Version 1.3 [in Danish]. 23 Nov 2018. Available at: https://sundhedsdatastyrelsen.dk/-/media/sds/filer/registre-og-services/nationale-sundhedsregistre/doedsaarsager-og-biologisk-materiale/laboratoriedatabasen/dokumentation-af-labdatabasens-forskertabel.pdf?la=da. Accessed 13 Feb 2018.

R99-1044 Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ 1991; 302 (6779): 766-8.

13.2 UNPUBLISHED REFERENCES

c10503662-01 Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, and the risk of genital infections, among patients treated with empagliflozin compared to patients treated with other SGLT2 inhibitors or DPP-4 inhibitors – First Interim Monitoring Report. Document number c10503662-01, June 2016.

c17493314-01 Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated with empagliflozin compared to patients treated with other SGLT2 inhibitors or DPP-4 inhibitors – Second Interim Monitoring Report. Document number c17493314-01, June 2017.

c19231123-01 A 5-year enhanced pharmacovigilance surveillance initiative to survey and characterise spontaneous occurrence and experience of ketoacidotic events in patients treated with empagliflozin–containing products (EUPAS21696), Interim Report 1. Document number c19231123-01, December 2017.

c24036731-01 PASS: Evaluation of the feasibility of a post-authorisation safety study of empagliflozin in additional data sources from Europe and the United States of America. Document number c24036731-01, June 2018.

c24037075-01 Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated with empagliflozin compared to patients treated with other SGLT2 inhibitors or DPP-4 inhibitors – Third Interim Monitoring Report. Document number c24037075-01, June 2018.

ANNEX 1. LIST OF STAND-ALONE DOCUMENTS

PASS: Evaluation of the feasibility of a post-authorisation safety study of empagliflozin in additional data sources from Europe and the United States of America. Document number: c24036731-01, June 2018.

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

European Network of Centres for Pharmacoepidemiology and Pharmacovigilance

Doc.Ref. EMEA/540136/2009
Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 3)

Adopted by the ENCePP Steering Group on 01/07/2016

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

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

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

Study title:

Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated

with empagliflozin compared to patients treated with DPP-4 inhibitors

Study reference number:

ENCEPP/SDPP/13413; http://www.encepp.eu/encepp/viewResource.htm?id=13414

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection‡	☒	☐	☐	6
1.1.2 End of data collection§	☒	☐	☐	6
1.1.3 Study progress report(s)	☒	☐	☐	6
1.1.4 Interim progress report(s)	☒	☐	☐	6
1.1.5 Registration in the EU PAS Register	☒	☐	☐	6
1.1.6 Final report of study results	☒	☐	☐	6

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:				
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	7
2.1.2 The objective(s) of the study?	☒	☐	☐	8
2.1.3 The target population? (i.e., population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	9.2.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no a priori hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, new or alternative design)	☒	☐	☐	9.1

‡ Date from which information on the first study is first recorded in the study data set or, in the case of secondary use of data, the date from which data extraction starts.

§ Date from which the analytical data set is completely available.

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9.1
3.3	Does the protocol specify measures of occurrence? (e.g. incidence rate, absolute risk)	☒	☐	☐	9.1
3.4	Does the protocol specify measure(s) of association? (e.g. relative risk, odds ratio, excess risk, incidence rate ratio, hazard ratio, number needed to harm (NNH) per year)	☒	☐	☐	9.1
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☒	☐	☐	11

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	9.1, 9.2.1
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period?	☒	☐	☐	9.2.2
	4.2.2 Age and sex?	☒	☐	☐	9.2.1
	4.2.3 Country of origin?	☒	☐	☐	9.1
	4.2.4 Disease/indication?	☒	☐	☐	9.2.1
	4.2.5 Duration of follow-up?	☒	☐	☐	9.2.9
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	9.2.7, 9.2.8

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	9.3.1
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	9.3.1
5.3	Is exposure classified according to time windows? (e.g. current user, former user, non-use)	☒	☐	☐	9.3.1
5.4	Is exposure classified based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☒	☐	

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	9.3.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	9.3.2
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, prospective or retrospective ascertainment, use of validation sub-study)	☒	☐	☐	9.3.2
6.4 Does the protocol describe specific endpoints relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease, disease management)	☐	☒	☐	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol describe how confounding will be addressed in the study?	☒	☐	☐	9.9.1, 9.9.2
7.1.1. Does the protocol address confounding by indication if applicable?	☒	☐	☐	9.9.1
7.2 Does the protocol address:				
7.2.1. Selection biases (e.g. healthy user bias)	☒	☐	☐	9.9.1, 9.9.2
7.2.2. Information biases (e.g. misclassification of exposure and endpoints, time-related bias)	☒	☐	☐	9.9.2
7.3 Does the protocol address the validity of the study covariates?	☒	☐	☐	9.3.3

Comments:

Section 8: Effect modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, subgroup analyses, anticipated direction of effect)	☒	☐	☐	9.9

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.1.3 Covariates?	☒	☐	☐	9.4, Annex 1 , Annex 4
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.2.3 Covariates? (e.g. age, sex, clinical and drug use history, comorbidity, comedications, lifestyle)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	9.4, Annex 1 , Annex 6
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD)-10, Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	9.4, Annex 1 , Annex 6
9.3.3 Covariates?	☒	☐	☐	9.4, Annex 1 , Annex 6
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	9.4, Annex 1 , Annex 6

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Is the choice of statistical techniques described?	☒	☐	☐	9.7.1, 9.7.2, 9.7.7
10.2 Are descriptive analyses included?	☒	☐	☐	9.7.2
10.3 Are stratified analyses included?	☒	☐	☐	9.7.2, 9.7.3
10.4 Does the plan describe methods for adjusting for confounding?	☒	☐	☐	9.7.1, 9.7.2, 9.9.1
10.5 Does the plan describe methods for handling missing	☒	☐	☐	9.7.5

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
data?				
10.6 Is sample size and/or statistical power estimated?	☒	☐	☐	9.5

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	9.8
11.2 Are methods of quality assurance described?	☒	☐	☐	9.8
11.3 Is there a system in place for independent review of study results?	☒	☐	☐	9.8

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☒	☐	☐	9.9
12.1.2 Information bias?	☒	☐	☐	9.9.2
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods)	☒	☐	☐	9.9.1
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure, duration of follow-up in a cohort study, patient recruitment)	☒	☐	☐	9.9.5, Annex 1

Comments:

Section 13: Ethical issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	10
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☒	☐	
13.3 Have data protection requirements been described?	☒	☐	☐	10

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	5

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	12
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	12

Comments:

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

Date:19 Jul 2021

Signature:

ANNEX 3. CODES TO BE USED FOR EXCLUSION CRITERIA

ICD-10 codes are used in the three data sources. The list of Read codes (and SNOMED, and EMIS local codes, if needed) to be used in the CPRD and ICD-9 codes to be used in the HIRD will be generated based on the common ICD-10 list and described in the statistical analysis plan.

Annex 3 Table 1 Liver injury exclusion criteria: ICD-10 code

ICD-10 code	Description
History of acute liver injury (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
K71	Toxic liver disease
K72	Hepatic failure, not elsewhere classified (includes acute and subacute hepatic failure, chronic hepatic failure, and hepatic failure, unspecified)
Liver transplant (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
Z94.4	Liver transplant status
T86.4	Liver transplant failure and rejection
Pregnancy (For ALI primary and secondary outcomes, algorithms including different time windows for different codes will be developed and described in detail in the statistical plan)	
O00-O99	Pregnancy, childbirth, and the puerperium
Chronic liver disease and alcoholism (For ALI primary outcome, the lookback period is ever before or at the index date)	
K70	Alcoholic liver disease
K73	Chronic hepatitis, not elsewhere classified
K74	Fibrosis and cirrhosis of liver
K75	Other inflammatory liver disease
K76	Other disease of liver
K77	Liver disorders in disease classified elsewhere
F10.1-F10.9	Alcohol dependence syndrome/Mental and behavioural disorders due to use of alcohol: dependence syndrome
I85.00	Oesophageal varices with bleeding
I86.4A	Gastric varices with bleeding
I85.9	Oesophageal varices without mention of bleeding
I86.4	Gastric varices without mention of bleeding
I98.2 I98.3	Oesophageal varices in diseases classified elsewhere/Secondary oesophageal varices with bleeding
K29.2	Alcoholic gastritis

ICD-10 code	Description
Z65.8	Personal history of alcoholism/Other specified problems related to psychosocial circumstances
K85.2	Alcohol-induced acute pancreatitis
K86.0	Alcohol-induced chronic pancreatitis
E24.4	Alcohol-induced pseudo-Cushing's syndrome
G31.2	Degeneration of nervous system due to alcohol
G62.1, G72.1	Alcoholic polyneuropathy; Alcoholic myopathy
I42.6	Alcoholic cardiomyopathy
Z50.2	Alcohol rehabilitation
Z71.4	Alcohol abuse counselling and surveillance
Infectious hepatitis (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
B15—B19, B26.8, A51.4, Z22.5, B25.1, (B00.8+K77.0), B58.1, B65.x, A52.7	Acute infectious hepatitis; Mumps hepatitis; Secondary syphilitic hepatitis; Other symptomatic late syphilis; Schistosomiasis; Toxoplasma hepatitis; Carrier or suspected carrier of viral hepatitis; Personal history of hepatitis/Other secondary hepatitis; Cytomegaloviral hepatitis; Herpes viral hepatitis
Chronic disease involving the liver or causing hyperbilirubinaemia (For ALI primary outcome, the lookback period is ever before or at the index date)	
E83.1	Haemochromatosis
E83.0	Wilson's disease
E88.0	Deficit of alpha-1-antitrypsin
I82.0	Budd-Chiari syndrome
E80.4, E80.5, E80.6, E80.7	Disorders of bilirubin excretion (Gilbert syndrome)
Biliary disease (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
K80	Cholelithiasis
K81	Cholecystitis
K82	Other diseases of gallbladder
K83	Other diseases of biliary tract
K87.0	Disorders of gallbladder and biliary tract in diseases classified elsewhere
Pancreatic diseases (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
K85	Acute pancreatitis
K86	Other diseases of pancreas
K87.1	Disorders of pancreas in diseases classified elsewhere

ICD-10 code	Description
Hepatobiliary and pancreatic neoplasms (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
C22	Malignant neoplasm of liver and intrahepatic bile ducts
C23	Malignant neoplasm of gallbladder
C24	Malignant neoplasm of other and unspecified parts of biliary tract
Heart failure (For ALI primary outcome, the lookback period is ever before or at the index date. For ALI secondary outcome, the lookback period is 6 months before or at the index date)	
I50.x, I13.0, I13.2, I11.0, I09.81,	Heart failure, hypertensive heart, and chronic kidney disease with heart failure

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 3 Table 2 Kidney injury exclusion criteria: ICD-10 codes

ICD-10 code	ICD-10 term
N00	Acute nephritic syndrome
N10	Acute tubulo-interstitial nephritis
N12	Tubulo-interstitial nephritis, not specified as acute or chronic
N14	Drug- and heavy-metal-induced tubule-interstitial and tubular conditions
N17	Acute kidney injury
N19	Unspecified kidney failure
T86.1	Kidney transplant failure and rejection
Y84.1	Kidney dialysis
Z49	Care involving dialysis
Z94.0	Kidney transplant status
Z99.2	Dependence on renal dialysis

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 3 Table 3 Prior chronic kidney disease exclusion criteria: ICD-10 codes

Read code	Description
N18	Chronic kidney disease
Q61	Cystic kidney disease
Z94.0	Kidney transplant status
C64	Malignant neoplasm of kidney, except renal pelvis
C65	Malignant neoplasm of renal pelvis

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 3 Table 4 Prior chronic pyelonephritis exclusion criteria: ICD-10 codes

Read code	Description
N11	Chronic tubulo-interstitial nephritis (including N11.0 Nonobstructive reflux-associated chronic pyelonephritis, N11.1 Chronic obstructive pyelonephritis, N11.8 Other chronic tubulo-interstitial nephritis, and N11.9 Chronic tubulo-interstitial nephritis, unspecified)

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

ANNEX 4. CODES TO DEFINE STUDY OUTCOMES

ICD-10 codes are used in the three data sources. The list of Read codes (and SNOMED, and EMIS local codes, if needed) to be used in the CPRD and ICD-9 codes to be used the HIRD will be generated based on the common ICD-10 list and described in the statistical analysis plan.

Annex 4 Table 1 Acute liver injury, ICD-10 codes

ICD-10 code	ICD-10 term
K71.1	Toxic liver disease with hepatic necrosis
K71.2	Toxic liver disease with acute hepatitis
K71.6	Toxic liver disease with hepatitis, not elsewhere classified
K71.9	Toxic liver disease, unspecified
K72.0	Acute and subacute hepatic failure
K72.9	Hepatic failure, unspecified
K76.8	Other specified diseases of liver
K76.9	Liver disease, unspecified
R17	Unspecified jaundice, excludes neonatal
Z94.4	Liver transplant

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: http://apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 4 Table 2 Acute liver injury and biochemistry tests, Read codes

Read code/ biochemistry test (enttype)	Description
Enttype=155	Alanine aminotransferase
Enttype=156	Aspartate aminotransferase
Enttype=158	Bilirubin
1675	Yellow/jaundiced colour
1675.11	Jaundice - symptom
2274	O/E - jaundiced colour
2274.11	O/E - jaundiced
7806	Therapeutic endoscopic operations on liver using laparoscope
7807	Diagnostic endoscopic examination of liver using laparoscope
7800111	Auxiliary liver transplant

Read code/ biochemistry test (enttype)	Description
7800112	Piggy back liver transplant
7800500	Orthotopic transplantation of liver NEC
7804200	Open wedge biopsy of lesion of liver
7805211	Exploration of liver transplant
7807000	Diagnostic laparoscopic examination and biopsy liver lesion
7807100	Laparoscopic ultrasound examination liver biopsy lesion liver
7807200	Laparoscopic ultrasound examination of liver NEC
44CU.00	Plasma alkaline phosphatase level
44D2.00	Liver function tests abnormal
44E.00	Serum bilirubin level
44E2.00	Serum bilirubin raised
44E6.00	Serum bilirubin borderline
44G2.00	Liver enzymes abnormal
44G3100	ALT/SGPT level abnormal
44H5100	AST/SGOT level abnormal
44H5200	AST/SGOT level raised
46R5.11	Bilirubin in urine
7800z00	Transplantation of liver NOS
7807y00	Diagnostic laparoscopic examination of liver OS
7807z00	Diagnostic laparoscopic examination of liver NOS
780A.00	Diagnostic percutaneous operations on liver
780A000	Percutaneous transvascular biopsy of lesion of liver
780A100	Percutaneous biopsy of lesion of liver NEC
780A111	Menghini needle biopsy of liver
780A112	Needle biopsy of liver NEC
780A113	Sheeba needle biopsy of liver
780Az00	Diagnostic percutaneous operation on liver NOS
780B000	Biopsy of liver NEC
780B011	Biopsy of lesion of liver NEC
780F000	Endoscopic ultrasound examination liver biopsy lesion liver
9N0v.00	Seen in liver clinic
J60.00	Acute and subacute liver necrosis

Read code/ biochemistry test (enttype)	Description
J600.00	Acute necrosis of liver
J600000	Acute hepatic failure
J600011	Acute liver failure
J600100	Acute hepatitis – non-infective
J600200	Acute yellow atrophy
J600z00	Acute necrosis of liver NOS
J601.00	Subacute necrosis of liver
J601000	Subacute hepatic failure
J601100	Subacute hepatitis – non-infective
J601200	Subacute yellow atrophy
J601z00	Subacute necrosis of liver NOS
J60z.00	Acute and subacute liver necrosis NOS
J622.00	Hepatic coma
J622.11	Encephalopathy - hepatic
J625.00	[X] Hepatic failure
J625.11	[X] Liver failure
J62y.11	Hepatic failure NOS
J62y.12	Liver failure NOS
J62y.13	Hepatic failure
J63.00	Other liver disorders
J633.00	Hepatitis unspecified
J633000	Toxic hepatitis
J633z00	Hepatitis unspecified NOS
J635.00	Toxic liver disease
J635000	Toxic liver disease with cholestasis
J635100	Toxic liver disease with hepatic necrosis
J635200	Toxic liver disease with acute hepatitis
J635700	Acute hepatic failure due to drugs
J635X00	Toxic liver disease, unspecified
J636.00	Central haemorrhagic necrosis of liver
J63y.00	Other specified liver disorder
J63y100	Non-specific reactive hepatitis

Read code/ biochemistry test (enttype)	Description
J63yz00	Other specified liver disorder NOS
J63z.00	Liver disorder NOS
J66y600	Obstructive jaundice NOS
R024.00	[D] Jaundice (not of newborn)
R024000	[D] Cholemia NOS
R024100	[D] Icterus NOS
R024111	[D] Jaundice
R024z00	[D] Jaundice (not of newborn) NOS
R104000	[D] Transaminase or lactic acid dehydrogenase raised
R104013	[D] Transaminase raised
R104200	[D] Alkaline phosphatase raised
R148.00	[D] Abnormal liver function test
R148.11	[D] LFTs abnormal
R148z00	[D] Abnormal liver function test NOS
ZV42700	[V] Liver transplanted
ZV7C000	[V] Assessment for liver transplant

ALT = alanine aminotransferase; AST = aspartate aminotransferase; LFT = liver function tests; NEC = not elsewhere classified; NOS = not otherwise specified; O/E = on examination; OS = otherwise specified; SGOT = serum glutamic oxaloacetic transaminase; SGPT = serum glutamic pyruvic transaminase.

Sources: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 3 Acute kidney injury, ICD-10 codes

ICD-10 code	ICD-10 term
N00	Acute nephritic syndrome
N10	Acute tubulo-interstitial nephritis
N12	Tubulo-interstitial nephritis, not specified as acute or chronic
N14	Drug- and heavy-metal-induced tubule-interstitial and tubular conditions
N17	Acute renal failure
N19	Unspecified kidney failure
T86.1	Kidney transplant failure and rejection
Y84.1	Kidney dialysis
Z49	Care involving dialysis
Z94.0	Kidney transplant status
Z99.2	Dependence on renal dialysis

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 4 Table 4 Acute kidney injury, Read codes

Read code	Description
K00.00	Acute glomerulonephritis
K00.11	Acute nephritis
K000.00	Acute proliferative glomerulonephritis
K001.00	Acute nephritis with lesions of necrotising glomerulitis
K00y100	Acute exudative nephritis
K00y200	Acute focal nephritis
K00y300	Acute diffuse nephritis
K00z.00	Acute glomerulonephritis NOS
K04.00	Acute renal failure
K040.00	Acute renal tubular necrosis
K041.00	Acute renal cortical necrosis
K042.00	Acute renal medullary necrosis
K042.11	Necrotising renal papillitis
K043.00	Acute drug-induced renal failure
K044.00	Acute renal failure due to urinary obstruction
K04y.00	Other acute renal failure
K04z.00	Acute renal failure NOS

Read code	Description
14V2.00	H/O: renal dialysis
14V2.11	H/O: kidney dialysis
7L1A.00	Compensation for renal failure
7L1A.11	Dialysis for renal failure
7L1A000	Renal dialysis
7L1A011	Thomas intravascular shunt for dialysis
7L1Ay00	Other specified compensation for renal failure
7L1Az00	Compensation for renal failure NOS
ZV45100	[V] Renal dialysis status
ZV56.00	[V] Aftercare involving intermittent dialysis
ZV56000	[V] Aftercare involving extracorporeal dialysis
ZV56011	[V] Aftercare involving renal dialysis NOS
ZV56100	[V] Preparatory care for dialysis
ZV56y00	[V] Other specified aftercare involving intermittent dialysis
ZV56y11	[V] Aftercare involving peritoneal dialysis
ZV56z00	[V] Unspecified aftercare involving intermittent dialysis
ZVu3G00	[X] Other dialysis
K0D.00	End-stage renal disease
7L1A200	Haemodialysis NEC
7L1A300	Haemofiltration
7L1A700	Haemoperfusion
K06.00	Renal failure unspecified
K06.11	Uraemia NOS
K060.00	Renal impairment
K060.11	Impaired renal function
Kyu2000	[X] Other acute renal failure
G500400	Acute pericarditis - uraemic
7L1A100	Peritoneal dialysis
7L1A400	Automated peritoneal dialysis
7L1A500	Continuous ambulatory peritoneal dialysis
7L1A600	Peritoneal dialysis NEC
14S2.00	H/O: kidney recipient
7B00.00	Transplantation of kidney
7B00000	Autotransplant of kidney

Read code	Description
7B00100	Transplantation of kidney from live donor
7B00111	Allotransplantation of kidney from live donor
7B00200	Transplantation of kidney from cadaver
7B00211	Allotransplantation of kidney from cadaver
7B00300	Allotransplantation of kidney from cadaver, heart-beating
7B00400	Allotransplantation kidney from cadaver, heart non-beating
7B00500	Allotransplantation of kidney from cadaver NEC
7B00y00	Other specified transplantation of kidney
7B00z00	Transplantation of kidney NOS
ZV42000	[V] Kidney transplanted

H/O = history of; NEC = not elsewhere classified; NOS = not otherwise specified.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 5 Chronic kidney disease, ICD-10 codes

ICD-10 code	ICD-10 term
E132	Other specified diabetes mellitus with incipient diabetes nephropathy adequately or inadequately controlled by insulin, diet, or oral agents
I12	Hypertensive renal disease
I13	Hypertensive renal and heart disease
N08	Glomerular disorders in diseases classified elsewhere
N18	Chronic renal failure

Source: Fleet et al. (2013) [R15-3138].

Annex 4 Table 6 Chronic kidney disease, Read codes

Read code	Read term
1Z12.00	Chronic kidney disease stage 3
K05.00	Chronic renal failure
1Z13.00	Chronic kidney disease stage 4
1Z1.00	Chronic renal impairment
9hE0.00	Except chronic kidney disease qual indic: Patient unsuitable
K060.00	Renal impairment
66i.00	Chronic kidney disease monitoring
K06.00	Renal failure unspecified
1Z14.00	Chronic kidney disease stage 5
9Ot0.00	Chronic kidney disease monitoring first letter
K08.00	Impaired renal function disorder
6AA.00	Chronic kidney disease annual review
K060.11	Impaired renal function
1Z1C.00	Chronic kidney disease stage 3 without proteinuria
9Ot.00	Chronic kidney disease monitoring administration
9hE1.00	Exc chronic kidney disease quality indicators: Inform dissen
K050.00	End-stage renal failure
7L1A200	Haemodialysis NEC
7L1A.11	Dialysis for renal failure
9hE.00	Exception reporting: chronic kidney disease quality indicator
7A60100	Creation of arteriovenous fistula NEC
7L1A100	Peritoneal dialysis
K08z.00	Impaired renal function disorder NOS
K06.11	Uraemia NOS
14V2.00	H/O: renal dialysis
4519.00	Deteriorating renal function
D215000	Anaemia secondary to chronic renal failure
D215.00	Anaemia secondary to renal failure
7L1B000	Insertion of ambulatory peritoneal dialysis catheter
1Z1B.00	Chronic kidney disease stage 3 with proteinuria
1Z15.00	Chronic kidney disease stage 3A
8L50.00	Renal transplant planned

Read code	Read term
7L1A000	Renal dialysis
K05.11	Chronic uraemia
K0D.00	End-stage renal disease
G22.11	Nephrosclerosis
ZV45100	[V] Renal dialysis status
7L1B100	Removal of ambulatory peritoneal dialysis catheter
1Z16.00	Chronic kidney disease stage 3B
1Z1E.00	Chronic kidney disease stage 3A without proteinuria
1Z1J.00	Chronic kidney disease stage 4 without proteinuria
1Z1H.00	Chronic kidney disease stage 4 with proteinuria
SP08300	Kidney transplant failure and rejection
1Z1G.00	Chronic kidney disease stage 3B without proteinuria
SP05613	[X] Peritoneal dialysis associated peritonitis

H/O = history of; NEC = not elsewhere classified; NOS = not otherwise specified.
Source: Denburg et al. (2011) [R15-3136].

Annex 4 Table 7 Pyelonephritis and sepsis, ICD-10 codes

ICD-10 code	Description
A40	Streptococcal sepsis
A41	Other sepsis
N10	Acute tubule-interstitial nephritis
N13.6	Pyonephrosis
N20	Calculus of kidney and ureter

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: http://apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 4 Table 8 Pyelonephritis and sepsis, Read codes

Read code	Description
Pyelonephritis	
K101.00	Acute pyelonephritis
K10y000	Pyelonephritis unspecified
K101z00	Acute pyelonephritis NOS
K100600	Calculous pyelonephritis
K10yz00	Unspecified pyelonephritis NOS
K10y.00	Pyelonephritis and pyonephrosis unspecified
K10y000	Pyelonephritis unspecified
K10y100	Pyelitis unspecified
K10y200	Pyonephrosis unspecified
K10y400	Pyelitis in diseases EC
K101000	Acute pyelonephritis without medullary necrosis
K101200	Acute pyelitis
K101300	Acute pyonephrosis
K10y300	Pyelonephritis in diseases elsewhere classified
K103.00	Pyeloureteritis cystica
K106.00	Candida pyelonephritis
Sepsis	
A38z.11	Sepsis
A3C.00	Sepsis
K190600	Urosepsis
J666.00	Biliary sepsis
L090y00	Sepsis NOS following abortion/ectopic/molar pregnancy
Q404z00	Umbilical sepsis NOS
L40.11	Sepsis - puerperal
A3Cy.00	Other specified sepsis
A3Cz.00	Sepsis NOS
A3C3.00	Sepsis due to Gram-negative bacteria
A3C0100	Sepsis due to *Streptococcus* group B
AB2y500	Candidal sepsis
A3C0300	Sepsis due to *Streptococcus pneumoniae*
A3C1.00	Sepsis due to *Staphylococcus*

Read code	Description
A3C2.11	Sepsis due to anaerobes
A270611	Listerial sepsis
A3C1000	Sepsis due to *Staphylococcus aureus*
A3C0000	Sepsis due to *Streptococcus* group A
A3C0.00	Sepsis due to *Streptococcus*
A023.00	Salmonella sepsis
A3C0y00	Other streptococcal sepsis
A396.00	Sepsis due to *Actinomyces*
A3C2.00	Sepsis due to anaerobic bacteria
A3C3y00	Sepsis due to other Gram-negative organisms
A270600	Sepsis due to *Listeria monocytogenes*
A3C0z00	Streptococcal sepsis, unspecified
AB2y511	Sepsis due to *Candida*

EC = elsewhere classified; NOS = not otherwise specified.

Source: Medical and product dictionary browsers, version 1.3. London: General Practice Research Database (now the CPRD); March 2010. Accessed 15 October 2014.

Annex 4 Table 9 Urinary tract infections, ICD-10 codes

ICD-10 code	Description
N30	Cystitis
N34	Urethritis and urethral syndrome
N37.0	Urethritis in diseases classified elsewhere
N39.0	Urinary tract infection, site not specified
N41	Inflammatory diseases of prostate
O23.2	Infections of urethra in pregnancy
O23.4	Unspecified infection of urinary tract in pregnancy
O23.5	Infections of the genital tract in pregnancy
O23.9	Other and unspecified genitourinary tract infections in pregnancy
O86.2	Urinary tract infection following delivery

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 4 Table 10 Urinary tract infections, Read codes

Read code	Description
K190z00	Urinary tract infection, site not specified NOS
K15.00	Cystitis
K190.00	Urinary tract infection, site not specified
1AG.00	Recurrent urinary tract infections
K150.00	Acute cystitis
K190.11	Recurrent urinary tract infection
K17y000	Urethritis unspecified
A994.00	Non-specific urethritis
K190300	Recurrent urinary tract infection
K190500	Urinary tract infection
K15z.00	Cystitis NOS
L166800	Urinary tract infection complicating pregnancy
L166.00	Genitourinary tract infections in pregnancy
K17.00	Urethritis due to non-venereal causes
K171.00	Postmenopausal atrophic urethritis
L166z11	UTI - urinary tract infection in pregnancy
K190400	Chronic urinary tract infection
L166600	Urinary tract infection following delivery
L166300	Genitourinary tract infection in pregnancy - not delivered
L166.11	Cystitis of pregnancy
L166z00	Genitourinary tract infection in pregnancy NOS
K17z.00	Urethritis due to non-venereal cause NOS
K17y.00	Other urethritis
L166000	Genitourinary tract infection in pregnancy unspecified
K17yz00	Other urethritis NOS
Kyu5500	[X] Other urethritis
L166100	Genitourinary tract infection in pregnancy - delivered
Kyu5100	[X] Other cystitis
K21.11	Prostatitis and other inflammatory diseases of prostate

Read code	Description
K211.00	Chronic prostatitis
K210.00	Acute prostatitis
K21z.00	Prostatitis NOS

NOS = not otherwise specified; UTI = urinary tract infection.

Source: Medical and product dictionary browsers, version 1.3. London: General Practice Research Database (now the CPRD); March 2010. Accessed 15 October 2014.

Annex 4 Table 11 Genital infections, ICD-10 codes

ICD-10 code	Description
Vulvovaginitis	
B37.3	Candidiasis of vulva and vagina
B37.4	Candidiasis of other urogenital sites
N77.1	Vaginitis, vulvitis and vulvovaginitis in infectious and parasitic diseases classified elsewhere
N76	Other inflammation of vagina and vulva (includes vaginitis, vulvitis, etc…) Additional codes (B95-B98) are used to identify infectious agent
Balanitis	
N48.1	Balanoposthitis (additional codes (B95-B98) are used to identify infectious agent)
N51.2	Balanitis in diseases classified elsewhere
B37.4	Candidiasis of other urogenital sites
Complications of GI or severe consequences of GI	
N77.0	Ulceration of vulva in infectious and parasitic diseases classified elsewhere
N77.8	Vulvovaginal ulceration and inflammation in other diseases classified elsewhere
N48.2	Other inflammatory disorders of penis, including the following conditions: • Abscess of corpus cavernosum and penis • Boil of corpus cavernosum and penis • Carbuncle of corpus cavernosum and penis • Cellulitis of corpus cavernosum and penis • Cavernitis (penis) Additional codes (B95-B98) are used to identify infectious agent.
N49.8	Inflammatory disorders of other specified male genital organs Inflammation of multiple sites in male genital organs (includes Fournier's gangrene)

ICD-10 code	Description
Exclude case of vulvovaginitis or balanitis if any of the following is present 30 days before or after the index diagnosis code:	
A06.8	Amoebic infection of other sites
A50-A64	Infections with a predominantly sexual mode of transmission (includes syphilis, gonococci, chlamydia, trichomonas, herpetic, and other)

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: website: apps.who.int/classifications/icd10/browse/2010/en. Accessed 15 October 2014.

Annex 4 Table 12 Vulvovaginitis, Read codes

Read code	Description
"Specific diagnosis" of bacterial vaginosis or *Candida* vulvovaginitis	
A3By700	Gardnerella vaginalis
AB21.00	Candidal vulvovaginitis
AB21.11	Monilial vulvovaginitis
AB21000	Candidiasis of vulva
AB21100	Candidiasis of vagina
AB21111	Vaginal thrush
AB22.00	Other urogenital candidiasis
AB21z00	Candidal vulvovaginitis NOS
K421900	Bacterial vaginitis
K421911	Bacterial vaginosis
Specific microbiology results	
4JK2300	HVS culture - *Gardnerella vaginalis*
4JK2400	High vaginal swab: fungal organism isolated
4J74.11	Fungus on microscopy
4KE0.00	Clue cells present
4KE.00	Clue cells
Non-specific positive microbiology results	
4JK2500	High vaginal swab: white cells seen
4JK7.00	Vaginal swab culture positive
4JK2000	High vaginal swab culture positive
4KA2.00	Vaginal vault smear inadequate
4KA4.00	Vaginal vault smear abnormal

Read code	Description
Non-specific diagnosis of vulvovaginitis	
K421.00	Vaginitis and vulvovaginitis
K421000	Vaginitis unspecified
K421100	Vulvitis unspecified
K421200	Vulvovaginitis unspecified
K421400	Vaginitis in diseases EC
K421500	Vulvitis in diseases EC
K421600	Vulvovaginitis in diseases EC
K421A00	Acute vulvitis
K421z00	Vaginitis and vulvovaginitis NOS
Complications of GI or severe consequences of GI	
K421111	Vulval sores
K423.00	Abscess of Bartholin's gland
K423.11	Vulvovaginal gland abscess
K424.00	Other abscess of vulva
K424000	Abscess of vulva
K424011	Abscess of labia
K424100	Carbuncle of vulva
K424111	Boil of vulva
K424200	Furuncle of vulva
K424z00	Other abscess of vulva NOS
K425.00	Ulceration of vulva
K425000	Ulceration of vulva unspecified
K425200	Ulceration of vulva in Behcet's disease
K425z00	Ulceration of vulva NOS
K42y000	Carbuncle of vagina
K42y100	Carbuncle of labium
K42y200	Ulcer of vagina
Kyu8500	[X]Vaginits,vulvits+vulvovaginitis/infect+parasitc diseas CE
Kyu8600	[X]Vulvovaginal ulceration+inflammation in other diseases CE

EC = elsewhere classified; HVS = high vaginal swab; NOS = not otherwise specified.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 13 List of diagnoses and corresponding Read codes and terms suggestive of sexually transmitted infection to be used to exclude vulvovaginitis cases

Read code	Read term
4JK2200	HVS culture - trichomonas vaginalis
43U8.00	Chlamydia test positive
K5A3.11	Senile (atrophic) vaginitis
A78A.00	Chlamydial infection
A98z.11	Gonorrhoea
K420900	Chlamydia cervicitis
A541100	Herpetic vulvovaginitis
AD10111	Trichomonal vaginitis
AD10100	Trichomonal vulvovaginitis
Ayu4L00	[X] Vulval warts
43U1.00	Chlamydia antigen ELISA positive
A78A000	Chlamydial infection of lower genitourinary tract
A78AX00	Chlamydial infection of genitourinary tract, unspecified
K5A5.00	Perimenopausal atrophic vaginitis
A78A500	Chlamydial infection of genital organs NEC
A541200	Herpetic ulceration of vulva
4JQA.00	Gonorrhoea test positive
A980.00	Acute gonorrhoea of lower genitourinary tract
Ayu6200	[X] Chlamydial infection, unspecified
A980z00	Acute gonorrhoea of lower genitourinary tract NOS
Ayu4D00	[X] Sexually transmitted chlamydial infection of other sites
A980200	Acute gonococcal vulvovaginitis
Ayu4K00	[X] Chlamydial infection of genitourinary tract, unspecified
A78A300	Chlamydial inf of pelviperitoneum oth genitourinary organs
Kyu8500	[X] Vaginitis, vulvitis + vulvovaginitis/infect + parasitic diseas CE
A982200	Chronic gonococcal vulvovaginitis
A982.00	Chronic gonorrhoea lower genitourinary tract
Ayu6100	[X] Other chlamydial diseases
A981z00	Acute gonorrhoea upper genitourinary tract NOS
A981.00	Acute gonorrhoea of upper genitourinary tract

Read code	Read term
A913500	Secondary syphilis of vulva
Kyu8400	[X] Ulceration of vulva in infectious + parasitic diseases CE
A983.00	Chronic gonorrhoea of upper genitourinary tract
A982z00	Chronic gonorrhoea of lower genitourinary tract NOS
K421300	Postirradiation vaginitis
K5A3.00	Postmenopausal atrophic vaginitis
A980200	Acute gonococcal vulvovaginitis
Ayu4K00	[X] Chlamydial infection of genitourinary tract, unspecified
A78A300	Chlamydial inf of pelviperitoneum oth genitourinary organs
Kyu8500	[X] Vaginitis, vulvitis + vulvovaginitis/infect + parasitic diseas CE
A982200	Chronic gonococcal vulvovaginitis

CE = classified elsewhere, EC = elsewhere classified; ELISA = enzyme amplified immunoassay; HVS = high vaginal swab; NEC = not elsewhere classified; NOS = not otherwise specified.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 14 Balanitis, Read codes

Read code	Description
K272.11	Infection of penis
K271z00	Balanoposthitis NOS
K271100	Posthitis
K271000	Balanitis
K271.11	Balanitis
K271.00	Balanoposthitis
AB22000	Candidal balanitis
2663.11	O/E - discharge - penis
Kyu6A00	[X] Balanitis in diseases classified elsewhere
AB22011	Penile candidiasis (thrush)
4JK8000	Penile swab culture positive
Complications of GI or severe consequences of GI	
K272000	Penile abscess
K272100	Penile boil

Read code	Description
K272200	Penile carbuncle
K272300	Cellulitis of penis
K284600	Fournier's gangrene of scrotum

GI = genital infections; NOS = not otherwise specified; O/E = on examination.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 15 List of diagnoses and corresponding Read codes and terms suggestive of sexually transmitted infection that will be used to exclude balanitis cases

Read code	Description
K271200	Zoon's balanitis
A78A000	Chlamydial infection of lower genitourinary tract
A78AX00	Chlamydial infection of genitourinary tract, unspecified
A78A500	Chlamydial infection of genital organs NEC
A541300	Herpetic infection of penis
K274.11	Balanitis xerotica obliterans
A060.00	Balantidiasis
A980.00	Acute gonorrhoea of lower genitourinary tract
A980z00	Acute gonorrhoea of lower genitourinary tract NOS
A05y100	Amoebic balanitis
A982z00	Chronic gonorrhoea of lower genitourinary tract NOS
A781212	Penile warts
A78A.00	Chlamydial infection
A98z.11	Gonorrhoea
4JQA.00	Gonorrhoea test positive
Ayu4D00	[X] Sexually transmitted chlamydial infection of other sites
A78A300	Chlamydial inf of pelviperitoneum oth genitourinary organs
Ayu4K00	[X] Chlamydial infection of genitourinary tract, unspecified

NEC = not elsewhere classified; NOS = not otherwise specified.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, April 2014. Accessed 15 October 2014.

Annex 4 Table 16 Diabetic ketoacidosis, ICD-10 codes

ICD-10 code	Description
E11.1	Type 2 diabetes mellitus, with ketoacidosis
E12.1	Malnutrition-related diabetes mellitus, with ketoacidosis
E13.1	Other specified diabetes mellitus, with ketoacidosis
E14.1	Unspecified diabetes mellitus, with ketoacidosis
E11.0	Type 2 diabetes mellitus, with coma
E12.0	Malnutrition-related diabetes mellitus, with coma
E13.0	Other specified diabetes mellitus, with coma
E14.0	Unspecified diabetes mellitus, with coma

Source: International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available at: http://apps.who.int/classifications/icd10/browse/2010/en. Accessed 16 March 2016.

Annex 4 Table 17 Diabetic ketoacidosis, Read codes

Read code	Description
C362700	Ketoacidaemia NEC
C362600	Metabolic ketoacidaemia
C10FP11	Type II diabetes mellitus with ketoacidotic coma
C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
C10FN11	Type II diabetes mellitus with ketoacidosis
C10FN00	Type 2 diabetes mellitus with ketoacidosis
C10A100	Malnutrition-related diabetes mellitus with ketoacidosis
C103z00	Diabetes mellitus NOS with ketoacidotic coma
C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
C103.00	Diabetes mellitus with ketoacidotic coma
C101z00	Diabetes mellitus NOS with ketoacidosis
C101y00	Other specified diabetes mellitus with ketoacidosis
C101100	Diabetes mellitus, adult onset, with ketoacidosis
C101.00	Diabetes mellitus with ketoacidosis

NEC = not elsewhere classified; NOS = not otherwise specified.

Source: CPRD Medical and Product Dictionary Browsers, version 1.4.0. Clinical Practice Research Datalink, Medicines and Health Products Regulatory Agency, United Kingdom, February 2016. Accessed 16 March 2016.

ANNEX 5. COVARIATES TO BE CONSIDERED FOR INCLUSION IN THE PROPENSITY SCORE MODEL, BY OUTCOME

HOSPITALISATION, ED VISIT, OR SPECIALIST VISIT FOR ACUTE LIVER INJURY

Annex 5 Table 1 Acute liver injury outcome: variables of interest to be collected for propensity score development

Demographic or lifestyle		
Age	Smoking history	
Sex	Alcohol consumption	
Calendar year of index date	History of alcohol abuse	
Duration of lookback time	Socioeconomic status: Index of multiple socioeconomic deprivation, quintiles: first least deprived, fifth most deprived	
Body mass index > 30 or obesity surgery		
Medications		
Drugs associated with liver injury[1]		
Acarbose	Estrogens	
Acetaminophen (prescription)	Fluoxetine	Phenytoin
Allopurinol	Flutamide	Pyrazinamide
Amiodarone	HAART drugs	Rifampicin
Amitriptyline	Irbesartan	Risperidone
Amoxicillin + clavulanic acid	Isoniazid	Sertraline
Anabolic steroids	Ketoconazole	Statins
Azathioprine	Lisinopril	Sulfonamides
Baclofen	Losartan	Terbinafine
Bupropion	Methotrexate	Tetracyclines
Captopril	Mirtazapine	Trazodone
Carbamazepine	Nitrofurantoin	Trazodone
Chlorpromazine	NSAIDs	Tricyclics
Clindamycin	Omeprazole	Trimethoprim-sulfamethoxazole
Clopidogrel	Oral contraceptives	Trovafloxacin
Cyproheptadine	Paroxetine	Valproic acid
Enalapril	Phenobarbital	Verapamil
Erythromycins	Phenothiazines	

Other medications		
Cardiovascular system drugs	Antiepileptics	Methotrexate
Lipid-modifying agents	Drugs for asthma and obstructive airways disease	Cyclosporin
Other antirheumatic agents	Systemic corticosteroids	Other immunosuppressants excluding systemic tacrolimus
Hormone-replacement therapy	Systemic tacrolimus	Systemic antivirals
Insulins	Azathioprine	Other antimicrobials
Other oral antidiabetic drugs (including specification of add-on or switch)		
Medical comorbidities		
Ischaemic heart disease	Diffuse diseases of connective tissue	Pancreatitis
Hypertensive disease	Rheumatoid arthritis	Urinary infections (chronic or recurring)
Heart failure	Osteoarthrosis	Immunosuppressive diseases such as human immunodeficiency virus infection/AIDS
Peripheral vascular disease	Polymyalgia rheumatica	Being hospitalised, especially for a serious condition that requires intensive care
Other cardiovascular disease	Renal insufficiency	Length of hospitalisation
Cerebrovascular disease	Other malignancies	Chronic disease score[2]
Hyperlipidaemia	Dementia	
Autoimmune disease	Peptic ulcer disease	
Asthma	Colon polyps	
Chronic obstructive pulmonary disease, emphysema, respiratory insufficiency	Crohn's disease	
	Ulcerative colitis	
Indicators of diabetes severity		
Renal insufficiency or diabetic nephropathy	Coronary heart disease	
Retinopathy	Cerebrovascular disease	
Neuropathy	Amputations	
Peripheral vascular disease	Time since first diagnosis of type 2 diabetes mellitus	

HAART = highly active antiretroviral therapy; NSAIDs = non-steroidal anti-inflammatory drugs.

1. Source: Navarro and Senior, 2006.
2. For example, scores like the ones developed by Elixhauser et al. [R13-3591] or Charlson/Deyo [R13-3589] to be specified in analysis plan.

HOSPITALISATION, ED VISIT, OR SPECIALIST VISIT FOR ACUTE KIDNEY INJURY

Annex 5 Table 2 Acute kidney injury outcome: variables of interest to be collected for propensity score development

Demographic or lifestyle	**Medications**
Age Sex Calendar year of index date Duration of lookback time Body mass index > 30 or obesity surgery Smoking history History of alcohol abuse Socioeconomic status: index of multiple socioeconomic deprivation, quintiles–first least deprived to fifth most deprived	Antihypertensives, diuretics including angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers, beta blockers, calcium-channel blockers, other antihypertensives, antiarrhythmics, digoxin, nitrates NSAIDs (non-steroidal anti-inflammatory drugs) Statins, fibrates Oral steroids Zoledronic acid Lipid-modifying agents Other: acetaminophen, antibiotics (penicillins, sulfa), anticonvulsants, antifungals, antituberculars, chemotherapeutic agents, methotrexate, aspirin and other antiplatelets (e.g., clopidogrel, ticlopidine, prasugrel), systemic antivirals, anticoagulants Concomitant antidiabetics (including specification of add-on or switch)

Medical comorbidities		**Indicators of diabetes severity**
Prior history of acute kidney injury > 6 months before or at the index date Being hospitalised, especially for a serious condition that requires intensive care Length of hospitalisation High blood pressure Heart failure Chronic renal disease or renal dialysis Liver disease Peripheral artery disease Chronic disease score[1] Other cardiovascular disease Autoimmune disease Chronic obstructive pulmonary disease, emphysema, respiratory insufficiency Diffuse diseases of connective tissue	Rheumatoid arthritis Osteoarthrosis Polymyalgia rheumatica Urinary infections (chronic or recurring) Kidney stones Crohn's disease Ulcerative colitis Pancreatitis Immunosuppressive diseases such as HIV/AIDS Peptic ulcer disease Dementia Asthma	Renal insufficiency or diabetic nephropathy, peripheral Retinopathy Neuropathy Peripheral vascular disease Coronary heart disease Cerebrovascular disease Amputations Time since first diagnosis of type 2 diabetes mellitus, if available

1. For example, scores like the ones developed by Elixhauser et al. [R13-3591] or Charlson/Deyo [R13-3589], to be specified in analysis plan.

SEVERE COMPLICATIONS OF URINARY TRACT INFECTION (INPATIENT AND OUTPATIENT PYELONEPHRITIS AND UROSEPSIS)

Annex 5 Table 3 Urinary tract infection outcome: variables of interest to be collected for propensity score development

Demographic or lifestyle	Medications
Age Sex Calendar year of index date Duration of lookback time Body mass index > 30 or obesity surgery Smoking history History of alcohol abuse Socioeconomic status: Index of multiple socioeconomic deprivation, quintiles—first least deprived to fifth most deprived	Antihypertensives/diuretics including angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers, beta blockers, calcium-channel blockers, other antihypertensives, antiarrhythmics, digoxin, nitrates NSAIDs (non-steroidal anti-inflammatory drugs) Statins, fibrates Oral steroids Zoledronic acid Lipid-modifying agents Other: acetaminophen, antibiotics (e.g., penicillins, sulfa), anticonvulsants, antifungals, antituberculars, chemotherapeutic agents, methotrexate, aspirin and other antiplatelets (e.g., clopidogrel, ticlopidine, prasugrel), anticoagulants, systemic antivirals Concomitant antidiabetics (including specification of add-on or switch)

Medical comorbidities		Indicators of diabetes severity
Prior history of UTI leading to hospitalisation or acute pyelonephritis > 6 months before or at the index date Being hospitalised, especially for a serious condition that requires intensive care Length of hospitalisation Kidney diseases Kidney and genitourinary stones and disease Pregnancy High blood pressure Heart failure Liver disease Other cardiovascular disease Autoimmune disease Chronic obstructive pulmonary disease, emphysema, respiratory insufficiency Diffuse diseases of connective tissue	Rheumatoid arthritis Osteoarthrosis Polymyalgia rheumatica Urinary infections (chronic or recurring) Crohn's disease Ulcerative colitis Pancreatitis Immunosuppressive diseases such as HIV/AIDS Peptic ulcer disease Dementia Asthma Chronic disease score[1]	Renal insufficiency or diabetic nephropathy, peripheral Retinopathy Neuropathy Peripheral vascular disease Coronary heart disease Cerebrovascular disease Amputations Time since first diagnosis of type 2 diabetes mellitus, (CPRD only)

HIV = human immunodeficiency virus.

1. For example, scores like the ones developed by Elixhauser et al. (1998) [R13-3591] or Charlson/Deyo [R13-3589], to be specified in analysis plan.

GENITAL INFECTIONS (INPATIENT AND OUTPATIENT)

Annex 5 Table 4 Genital infections outcome: variables of interest to be collected for propensity score development

Demographic or lifestyle	Medications
Age Sex Calendar year of index date Duration of lookback time Body mass index > 30 or obesity surgery Smoking history History of alcohol abuse Socioeconomic status: Index of multiple socioeconomic deprivation, quintiles—first least deprived to fifth most deprived	Antihypertensives/diuretics including angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers, beta blockers, calcium-channel blockers, other antihypertensives, antiarrhythmics, digoxin, nitrates NSAIDs (non-steroidal anti-inflammatory drugs) Statins, fibrates Oral steroids Zoledronic acid Lipid-modifying agents Other: acetaminophen, antibiotics (e.g., penicillins, sulfa), anticonvulsants, antifungals, antituberculars, chemotherapeutic agents, methotrexate, aspirin and other antiplatelets (e.g., clopidogrel, ticlopidine, prasugrel), anticoagulants, systemic antivirals Concomitant antidiabetics (including specification of add-on or switch)

Medical comorbidities		Indicators of diabetes severity
Prior history of genital infection Being hospitalised, especially for a serious condition that requires intensive care Length of hospitalisation Kidney diseases Kidney and genitourinary stones and disease Pregnancy High blood pressure Heart failure Liver disease Other cardiovascular disease Autoimmune disease Chronic obstructive pulmonary disease, emphysema, respiratory insufficiency Diffuse diseases of connective tissue	Rheumatoid arthritis Osteoarthrosis Polymyalgia rheumatica Urinary infections (chronic or recurring) Crohn's disease Ulcerative colitis Pancreatitis Immunosuppressive diseases such as HIV/AIDS Peptic ulcer disease Dementia Asthma Chronic disease score[1]	Renal insufficiency or diabetic nephropathy, peripheral Retinopathy Neuropathy Peripheral vascular disease Coronary heart disease Cerebrovascular disease Amputations Time since first diagnosis of type 2 diabetes mellitus, (CPRD only)

HIV = human immunodeficiency virus.

1. For example, scores like the ones developed by Elixhauser et al. (1998) [R13-3591] or Charlson/Deyo [R13-3589], to be specified in analysis plan.

HOSPITALISATION OR ED VISIT FOR DIABETIC KETOACIDOSIS

Annex 5 Table 5 Diabetic ketoacidosis outcome: variables of interest to be collected for propensity score development

Demographic or lifestyle	**Medications**
Age	
Sex	Clozapine or olanzapine
Calendar year of index date	Lithium
Duration of lookback time	Terbutaline
Body mass index > 30 or obesity surgery	Corticosteroids
Smoking history	Thiazides
History of alcohol abuse	Pentamidine
Socioeconomic status: Index of multiple socioeconomic deprivation, quintiles—first least deprived to fifth most deprived	Concomitant non-insulin antidiabetics (including specification of add-on or switch)
Cocaine	Insulin therapy, change in insulin dose
Medical comorbidities	**Indicators of diabetes severity**
Prior history of DKA	
Acute illness: infections (such as UTI, gastroenteritis, urosepsis, influenza), recent surgery, or trauma	
Thyroid problems (e.g., thyroid storm and thyrotoxicosis)	Renal insufficiency or diabetic nephropathy, peripheral
Myocardial ischaemia/infarction	Retinopathy
Pancreatitis	Neuropathy
Psychological stress	Peripheral vascular disease
Reduced caloric or fluid intake	Coronary heart disease
Hypovolemia	Cerebrovascular disease
Alcohol intake	Amputations
Hypoxemia	Time since first diagnosis of type 2 diabetes mellitus
Acute renal failure	
Heart Failure	
Cerebrovascular accident	
Chronic disease score[1]	

UTI = urinary tract infection.

1. For example, scores like the ones developed by Elixhauser et al. (1998) [R13-3591] or Charlson/Deyo [R13-3589], to be specified in the analysis plan.

ANNEX 6. OVERVIEW OF CHARACTERISTICS OF THE DATA SOURCES

Annex 6 Table 1 Characteristics of the data sources

Database feature	**CPRD**	**Danish Population Registries**	**HIRD**
Population of country[1]	United Kingdom: 62,435,709	Denmark: 5,748,800	US: 323.1 million
Database population	5,79 million	Whole country	40 million
Database type	Electronic medical records	Population-based registries and databases; link between all databases through Civil Personal Registration Number	Insurance claims records for health care services
Primary care data available	Yes	No on diagnoses, Yes on dispensings	Yes
Specialist outpatient visits available	Only if the GP decided to include these in the medical record	Only hospital clinic visits	Yes
Hospital discharge data available	Partial linkage to HES (~54%) in CPRD GOLD, ~93% in CPRD Aurum	Yes	Yes
Data on medications and type of prescriptions	GP prescriptions issued	Reimbursed-pharmacy-filled prescriptions	Reimbursed pharmacy-filled prescriptions
Drug dictionary codes/therapeutic classification[2]	Multilex/British National Formulary in CPRD GOLD, the Dictionary of Medicines and Devices (DM+D) in CPRD Aurum	ATC	National Drug Codes, which can be mapped to other coding systems
Disease and procedure coding system(s)	Read codes in CPRD GOLD, Read codes, SNOMED codes, and EMIS local codes in CPRD Aurum	ICD-10 Surgical procedures: NOMESCO	ICD-9, ICD-10, CPT, HCPCS
Laboratory (requests, results)	Outpatient laboratory results available electronically for 100% of the population. Inpatients laboratory results may be available in the hospital discharge letters and obtained through GP questionnaires	Results from hospital laboratories available electronically for 25% of the population; otherwise, via abstraction of medical records	Outpatient laboratory results available electronically for 30% of the population; otherwise, via abstraction of medical records

Database feature	CPRD	Danish Population Registries	HIRD
Data availability	Since 1987	Patient register, since 1977; prescription registries, since 1995 (all); since 2004 (reimbursed); laboratory data since ca. 1996	Since January 2006
Approximate time lag (updates per year)	1 month (monthly)	Varies by data source: from 3 to 12 months (1 per year)	3 months (monthly updates for the pharmacy data)
Access to source records for validation (requires special approval)	GPs can be sent questionnaires via the CPRD for validation	Yes, with an approval from the Patient Safety Board	Yes, for about 60% of the subset of around 50% of the patients for whom HealthCore has permission to access their protected health information

ATC = Anatomical Therapeutic Chemical (classification system); CPRD = Clinical Practice Research Datalink; GP = general practitioner or general practice; CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedural Coding System; HES = Hospital Episode Statistics; HIRD = HealthCore Integrated Research Database℠; ICD-9 = International Classification of Diseases, 9th Revision; ICD-9-CM = International Classification of Diseases, 9th Revision, Clinical Modification; ICD-10 = International Statistical Classification of Diseases and Related Health Problems, 10th Revision; ICD-10-CM = International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Clinical Modification; ICD-9-CM = International Classification of Diseases, 9th Revision, Clinical Modification;. NOMESCO = Nordic Medico-Statistical Committee.

1 Eurostat population at 01 January 2017 [R18-0353].

2 Duration derived from prescription/dispensing data (i.e., formulation strength, quantity prescribed/dispensed) and defined daily dose.

APPROVAL / SIGNATURE PAGE

Document Number: c03270726 **Technical Version Number:8.0**

Document Name: 1245-96-protocol-amendment-v0-8-final

Title: Post-authorisation safety study in patients with type 2 diabetes mellitus to assess the risk of acute liver injury, acute kidney injury and chronic kidney disease, severe complications of urinary tract infection, genital infections, and diabetic ketoacidosis among patients treated with empagliflozin compared to patients treated with DPP-4 inhibitors

Signatures (obtained electronically)

Meaning of Signature	Signed by	Date Signed
Verification-Paper Signature Completion	[redacted]	20 Jul 2021 12:34 CEST
Approval-Team Member Medicine	[redacted]	20 Jul 2021 12:42 CEST
Approval-Other	[redacted]	20 Jul 2021 13:44 CEST
Approval-On behalf of Head or VP or Director	[redacted]	20 Jul 2021 17:39 CEST
Approval-Head Safety Evaluation Therapeutic Area	[redacted]	21 Jul 2021 17:56 CEST
Approval	[redacted]	22 Jul 2021 06:54 CEST
Approval-EU Qualified Person Pharmacovigilance	[redacted]	22 Jul 2021 08:28 CEST

(Continued) Signatures (obtained electronically)

Meaning of Signature	**Signed by**	**Date Signed**