PASS INFORMATION

Title	Post-authorization safety study to assess the risk of diabetic ketoacidosis among type 2 diabetes mellitus patients treated with ertugliflozin compared to patients treated with other antihyperglycemic agents
Protocol Version identifier	8835-062/000.V3
Date of last version of protocol	5/7/2018
EU PAS Register No	Not yet registered
Active substance	Ertugliflozin Ertugliflozin + Metformin Hydrochloride Ertugliflozin + Sitagliptin
Medicinal product(s)	STEGLATRO™ (ertugliflozin; ATC code A10BK04) SEGLUROMET™ (ertugliflozin/metformin; ATC code A10BD23) STEGLUJAN™ (ertugliflozin/sitagliptin; ATC code A10BD24)
Product reference	EMEA/H/C/004315 EMEA/H/C/004314 EMEA/H/C/004313
Procedure number	EMEA/H/C/4315/MEA/002 - Steglatro EMEA/H/C/4313/MEA/002 - Steglujan EMEA/H/C/4314/MEA/002 - Segluromet
Marketing authorisation holder(s) (MAH)	Merck Sharp & Dohme, Corp., a subsidiary of Merck & Co., Inc. One Merck Drive P.O. Box 100 Whitehouse Station, NJ, 08889-0100, U.S.A.
Joint PASS	No

Research question and objectives	This study will address the research question of whether new use of ertugliflozin is associated with an increased risk of diabetic ketoacidosis (DKA), compared to new use of other non-sodium-glucose cotransporter 2 (non-SGLT2) inhibitor antihyperglycemic agents (AHAs) among type 2 diabetes mellitus (T2DM) patients. The primary objectives of the study are: 1) To assess the risk of DKA among new users of ertugliflozin relative to new users of sulfonylureas (SUs) or thiazolidinediones (TZDs); 2) To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs [i.e. dipeptidyl peptidase 4 (DPP-4) inhibitors or glucagon-like peptide-1 (GLP-1) receptor agonists]. The secondary objectives of the study are: 1) To assess the risk of DKA among new users of ertugliflozin relative to new users of SUs or TZDs, separately in insulin users and non-insulin users at baseline; 2) To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs, separately in insulin users and non-insulin users at baseline.
Country of study	United States

Author	Jeffrey S. Brown, PhD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD Ting-Ying Huang, BSPharm, PhD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD Aaron B. Mendelsohn, PhD, MPH Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD Sengwee Toh, ScD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD PPD

	Gregory W. Daniel, PhD, MPH, RPh Duke-Robert J. Margolis Center for Health Policy 1201 Pennsylvania Ave NW, Suite 500 Washington DC, USA 20004 PPD
MAH Contact Person	[redacted]
Merck Final Repository (RCAM) Date	14-AUG-2019
Date of Health Authority Approval of Protocol	19-SEP-2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

AE	Adverse event
AHA	Antihyperglycemic agents
CI	Confidence interval
CPT	Current Procedural Terminology
DKA	Diabetic ketoacidosis
DPP-4	Dipeptidyl peptidase 4
DSUR	Development Safety Update Reports
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EV	EudraVigilance system
FAERS	FDA Adverse Event Reporting System
FDA	Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act
FDAMA	Food and Drug Administration Modernization Act
FISMA	Federal Information Security Management Act
GVP	Good PharmacoVigilance Practice
GLP-1	Glucagon-like peptide-1
HbA_{1c}	Haemoglobin A_{1c}
HCPCS	Healthcare Common Procedure Coding System
HPHCI	Harvard Pilgrim Health Care Institute
HR	Hazard ratio
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
ICD-10-CM	International Classification of Diseases, 10th Revision, Clinical Modification
ICD-10-PCS	International Classification of Diseases, 10th Revision, Procedure Coding System
IEC	Independent Ethics Committee
IMEDS	Innovation in Medical Evidence and Development Surveillance
IMEDS-DD	Innovation in Medical Evidence and Development Surveillance Distributed Database
IRB	Institutional Review Board
MSD	Merck Sharp & Dohme, Corp.
NDC	National Drug Codes
PASS	Post-Authorization Safety Study
PBRER	Periodic Benefit Risk Evaluation Report
PCORnet	The National Patient-Centered Clinical Research Network
PSUR	Periodic Safety Update Report
QA	Quality assurance

QC	Quality control
QPPV	Qualified Person for PharmacoVigilance
RMP	Risk Management Plan
SU	Sulfonylurea
SGLT2	Sodium-glucose cotransporter 2
SOP	Standard operating procedure
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TZD	Thiazolidinedione
US	United States
ZIP	Zone Improvement Plan

1 RESPONSIBLE PARTIES

Principal investigator	Jeffrey S. Brown, PhD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD
Coordinating investigator for each country in which the study is to be performed	Not applicable
Sponsor contacts	PPD
Other contacts	Not applicable
Vendor/Collaborator	Innovation in Medical Evidence and Development Surveillance
Investigators	Jeffrey S. Brown, PhD[1] Ting-Ying Huang, BSPharm, PhD[1] Aaron B. Mendelsohn, PhD, MPH[1] Sengwee Toh, ScD[1] Gregory W. Daniel, PhD, MPH, RPh[2] [1] Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA [2] Duke-Robert J. Margolis Center for Health Policy, Washington DC, USA
Shared responsibilities	Not applicable

2 ABSTRACT

Title	Post-authorization safety study to assess the risk of diabetic ketoacidosis among type 2 diabetes mellitus patients treated with ertugliflozin compared to patients treated with other antihyperglycemic agents
Protocol Number / Version	8835-062/000.V3
Date	8/6/2019
Author	Jeffrey S. Brown, PhD Ting-Ying Huang, BSPharm, PhD Aaron B. Mendelsohn, PhD, MPH Sengwee Toh, ScD PPD Gregory W. Daniel, PhD, MPH, RPh
Rationale & Background	Sodium–glucose cotransporter 2 (SGLT2) inhibitors decrease plasma glucose by blocking the reabsorption of glucose at the proximal tubule. Safety concerns have been raised about the potential increased risk of diabetic ketoacidosis (DKA) by use of SGLT2 inhibitors among patients with type 2 diabetes mellitus (T2DM). Up to date, there are no non-interventional safety studies assessing the risk of DKA associated with the use of ertugliflozin.
Research Question(s) & Objective(s)	This study will address the research question of whether new use of ertugliflozin is associated with an increased risk of DKA, compared to new use of other non-SGLT2 inhibitor antihyperglycemic agents (AHAs) among T2DM patients. The primary objectives of the study are: 1) To assess the risk of DKA among new users of ertugliflozin relative to new users of sulfonylureas (SUs) or thiazolidinediones (TZDs); 2) To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs [i.e. dipeptidyl peptidase 4 (DPP-4) inhibitors or glucagon-like peptide-1 (GLP-1) receptor agonists).

	The secondary objectives of the study are: 1) To assess the risk of DKA among new users of ertugliflozin relative to new users of SUs or TZDs, separately in insulin users and non-insulin users at baseline; 2) To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs, separately in insulin users and non-insulin users at baseline.
Study Design	A non-interventional cohort study will be conducted using the Reagan-Udall Foundation for the Food and Drug Administration (FDA)'s Innovation in Medical Evidence and Development Surveillance Distributed Database (IMEDS-DD), a subset of the FDA Sentinel Distributed Database. The study will adopt a new user design and compare patients who initiate ertugliflozin to those who initiate a non-SGLT2 inhibitor comparator AHA. Propensity scores based on baseline characteristics will be used to account for potential confounding.
Population	Eligible patients will meet the following criteria: **Inclusion criteria** • New users of ertugliflozin or new users of a comparator AHA beginning on 1 July 2018; • Age 18 years or older on the new initiation date (i.e., index date) of ertugliflozin or a comparator AHA; • 6 or more months of continuous enrollment (maximum allowable enrollment gap of 45 days) in medical and prescription drug insurance plans before the index date; • T2DM, evidenced by at least one qualifying diagnosis recorded in claims of any encounter type any time before or on the index date.

	Exclusion criteria • Type 1 diabetes (T1DM) or gestational diabetes, evidenced by at least one qualifying diagnosis recorded in claims of any encounter type any time before or on the index date; • Initiation of insulin on the index date (insulin use that discontinues before index date and ongoing insulin use will be allowed); • History of DKA, evidenced by at least one qualifying discharge diagnosis recorded at any diagnosis position in claims of inpatient encounters, any time before the index date.
Variables	Exposure: new use of (1) ertugliflozin; (2) SUs or TZDs; and (3) incretin-based drugs (i.e. DPP-4 inhibitors or GLP-1 receptor agonists). Outcome: hospitalization for DKA identified from principal discharge diagnosis of inpatient claims. Covariates: demographics (age, sex, calendar year), use of antihyperglycemic agents by class (metformin, SU, TZD, alpha glucosidase inhibitor, meglitinides, DPP-4 inhibitor, GLP-1 receptor agonist, SGLT2 inhibitor, insulin), use of medications associated with DKA (clozapine or olanzapine, lithium, terbutaline, oral corticosteroids, thiazides, pentamidine), comorbidity burden (combined comorbidity index), pre-existing comorbidities (acute illness, acute renal failure, cerebrovascular disease, myocardial infarction, stroke, coronary heart disease, heart failure, hypertension, hyperlipidaemia, pancreatitis, hypovolemia, hypoxemia, thyroid disorders), diabetic complications (moderate to severe renal insufficiency or diabetic nephropathy, neuropathy, retinopathy, peripheral vascular disease, amputation), lifestyle (obesity, alcohol use, tobacco use, cocaine abuse), and health services utilization.
Data Sources	The study will be conducted using the IMEDS-DD, a subset of the FDA Sentinel Distributed Database. The Sentinel Distributed Database is a national

	electronic system for active surveillance of the safety of drugs, biologics, vaccines, and medical devices in the US. It is expected that patient population in the IMEDS-DD will be largely representative of the commercially-insured population in US. Feasibility assessment to examine the comparability of the T2DM population in the IMEDS-DD to the general T2DM population in the US and data availability relevant to this study will be conducted upon the approval of this protocol and before study execution.
Study Size	In order to detect a hazard ratio of 2.0 or above for DKA in ertugliflozin users relative to comparator AHA users, with targeted power of 80% and significance level of 0.05 in a two-sided test, a total of 66 DKA events in ertugliflozin and comparator AHA groups combined is required. It is expected that this can be achieved by 8,819 person-years of ertugliflozin new users matched to comparator AHA new users in a 1:1 ratio on propensity score, assuming DKA incidence rate is 2.5 per 1,000 person-years among T2DM patients treated with comparator AHAs. When DKA incidence rate in T2DM patients treated with comparator AHAs increases, the required sample size to achieve the same power is expected to decrease.
Data Analysis	Propensity score matching in 1:1 ratio will be used for confounding adjustment as a primary analysis. Two sets of propensity scores, one for the comparison of ertugliflozin versus SU/TZD and one for the comparison of ertugliflozin versus incretin-based drugs, will be generated. Baseline demographic and clinical characteristics will be described by exposure group before and after propensity score matching. Incidence rates (and 95% confidence interval) of DKA will be calculated by exposure group.

	The differences between the exposure groups in terms of time to DKA will be assessed using Kaplan-Meier survival curves with log rank test. Cox proportional hazards models will be used separately to compare the risk of DKA among new users of ertugliflozin to that among new users of SU/TZD; and to compare the risk of DKA among new users of ertugliflozin to that among new users of incretin-based drugs. Subgroup analysis will be further conducted by concomitant insulin use on the index date. Sensitivity analyses pre-defined in the protocol will be conducted to assess the robustness of the study results.
Milestones	
Start of data collection:	01 July 2018
End of data collection:	30 June 2023
Feasibility Assessment report:	31 December 2020
Interim report(s) of study results:	Interim report 1: 31 December 2021 Interim report 2: 31 December 2022
Study progress report(s):	Not applicable
Final report of study results:	31 December 2023

3 AMENDMENTS AND UPDATES

Number	Date	Section of Study Protocol	Amendment or update	Reason
2	July 29 2019	2. Abstract	Amended Study Progress Reports into Interim Reports.	Address comments in the Preliminary Assessment Report
		4. Milestones	Amended Study Progress Reports into Interim Reports.	Address comments in the Preliminary Assessment Report
1	April 25 2019	4. Milestones	Clarified Study Progress Report	Address PRAC's comments
		5. Rationale and Background	Added the US approval date of ertugliflozin.	Address PRAC's comments
		7.2 Setting	Clarified continuous enrollment	Address PRAC's comments
		7.2 Setting	Added sentence to define "narrow T2DM population"	Address PRAC's comments
		7.3.1 Exposure	Added a paragraph to clarify "duration use for AHAs"	Address PRAC's comments
		7.3.2 Outcome	Provided a precise definition of the outcome	Address PRAC's comments
		7.3.3 Covariates	Added Table 2 Approaches to handling concomitant AHAs	Address PRAC's comments
		7.4.3 Feasibility assessment	Added sentence and Table 3 to describe characteristics to be compared between IMEDS-DD and general T2DM population	Address PRAC's comments

Number	Date	Section of Study Protocol	Amendment or update	Reason
		7.4.3.3.2 Case definition of DKA hospitalization	Provided the precise definition of the outcome	Address PRAC's comments
		7.5 Study size	Updated the background event rate	Address PRAC's comments and improve clarity
		7.7.2 Descriptive analysis	Added one descriptive analysis	Address PRAC's comments
		7.7.3.2 Sensitivity analysis	Added sensitivity analyses and updated Table 5.	Address PRAC's comments
		7.7.3.3 Model specification	Added advantages of propensity score matching on a quarterly basis	Address PRAC's comments
		9.1 Adverse Event Reporting	Updated the AE reporting Fax number	Updated the AE reporting Fax #

4 MILESTONES

Milestone	Planned Date
Start of data collection	1 July 2018
End of data collection	30 June 2023
Interim report 1	31 December 2021
Interim report 2	31 December 2022
Registration in the EU PAS register	Not yet registered
Feasibility assessment report	31 December 2020
Final report of study results	31 December 2023

5 RATIONALE AND BACKGROUND

Sodium-glucose co-transporter 2 (SGLT2) inhibitors are used together with diet and exercise in patients with type 2 diabetes mellitus (T2DM), either alone or in combination with other antihyperglycemic agents (AHAs). SGLT2 is expressed in the proximal renal tubules and is responsible for the majority of the reabsorption of filtered glucose from the tubular lumen. By blocking the action of SGLT2, these substances cause more glucose to be removed via the urine, thereby reducing the levels of glucose in the blood via an insulin-independent mechanism [1].

In 2015, both the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) issued safety warnings describing the risk of diabetic ketoacidosis (DKA) among SGLT2 inhibitor users [2] [3]. Searches in their respective pharmacovigilance system identified DKA cases linked to the use of individual SGLT2 inhibitors. Assessments from the regulatory agencies, along with other research, determined that the safety concern is class-wide and suggested that the associated DKA may present atypically with lower-than-anticipated glucose levels [3] [4] [5] [6] [7] [8] [9] [10] [11]. As post-market safety data are still accruing, evidence remains inconclusive, and the nature of the risk remains uncertain. Ongoing surveillance and continuous monitoring of DKA incidence among SGLT2 users are recommended by EMA, FDA, and various clinical associations worldwide [3] [11] [12] [13].

Ertugliflozin is a SGLT2 inhibitor. Ertugliflozin products (including ertugliflozin, ertugliflozin/sitagliptin and ertugliflozin/metformin HCl) were first approved in the US on 19 December 2017 and approved in Europe in March 2018 for T2DM treatment to improve glycemic control in adults. As its marketing authorization holder, MSD has committed to EMA to conduct a post-authorisation safety study (PASS) to investigate the association of ertugliflozin use with DKA among T2DM patients. This study is also included in the ertugliflozin Risk Management Plan (RMP) as one of the required pharmacovigilance activities.

5.1 Diabetic Ketoacidosis among Patients with Type 2 Diabetes Mellitus

DKA is a serious and sometimes life-threatening condition. Marked as excessive accumulation of acidic ketone bodies in the blood, DKA is triggered by switch of metabolic processing from glucose to fatty acids when insulin levels are too low. DKA most commonly occurs in patients with type 1 diabetes (T1DM) and is usually accompanied by high blood sugar levels (>250 mg/dL) [14]. However, in a number of cases identified in EudraVigilance (EV) database and FDA Adverse Event Reporting System (FAERS), T2DM patients with DKA were reported [3] [11].

The population-based incidence rates of DKA among T2DM patients are not well established and vary internationally. Using a hospital catchment area data in Sweden, Wang et al estimated a crude incidence rate of 0.5 per 1,000 person-year among T2DM patients [15]. Using commercially-insured claims databases in US, Erondu et al estimated a crude

incidence rate of 0.3 to 2.0 per 1,000 person-years among T2DM patients [5]. Using a Danish national registry, Jensen et al estimated a crude incidence rate of 1.3 per 1,000 person-years among T2DM patients [16]. Using a medical claims database in Japan, Takeuchi et al estimated a crude incidence rate of 0.5 per 1,000 person-year among treated T2DM patients [17].

5.2 Diabetic Ketoacidosis among Patients Treated with Sodium-Glucose Co-Transporter 2 Inhibitors

Since the EMA and FDA warnings [2] [3], several non-interventional, pharmacoepidemiologic studies have assessed the DKA incidence rates specifically for T2DM patients exposed to SGLT2 inhibitors or other antihyperglycemic agents. Notably, a range of 0.6 to 4.9 per 1,000 person-years was reported for SGLT2 inhibitor users, whereas 0.5 per 1,000 person-years was reported for any treated T2DM patients, a range of 0.7 to 3.3 per 1,000 person-years was reported for dipeptidyl peptidase 4 (DPP-4) inhibitor users, and a range of 0.6 to 1.8 per 1,000 person-years was reported for users of other antihyperglycemic agents (including sulfonylurea (SU), DPP-4 inhibitors, glucagon-like peptide-1 (GLP-1) receptor agonists, thiazolidinediones (TZDs), insulin, and other miscellaneous medications) [16] [17] [18] [19] [20]. While risk estimates for DKA among T2DM patients treated with SGLT2 inhibitors remain low, data variability and availability continue to be the biggest challenges in safety assessments using health insurance administrative claims (market uptake of SGLT2 inhibitors takes time, in addition to expected data lags).

On average, time from the new initiation of SGLT2 inhibitors to DKA diagnosis among T2DM patients is 108 to 130 days, depending on whether T2DM population is defined with or without patients also diagnosed as T1DM or secondary diabetes after exposure [18]. These observational data-based estimates are consistent with those reported in clinical trials of SGLT2 inhibitors (time-to-onset is around 2 to 4 months in most cases, ranging from 1 day to 12 months in total) [11]. Decreasing risk of DKA over time has been found in separate studies: one study observed incidence rates decreasing from 7.5 to 5.6 to 4.9 per 1,000 person-year when assessed within 30, 60, and 180 days of SGLT2 inhibitor initiation respectively [19], and another study found incidence rates decreasing from 2.5 to 1.6 to 1.1 per 1,000 person-year when assessed within 30, 90, and 180 days of SGLT2 inhibitor initiation, respectively [20]. However, similar decreasing trend in DKA risk was described for DPP-4 users in these studies. In time-to-event analyses comparing new users of SGLT2 versus DPP-4 inhibitors, the hazard ratios (HRs) for DKA did not vary by these assessed risk windows.

When compared with other antihyperglycemic agents (non-SGLT2 inhibitors), SGLT2 inhibitors were estimated to increase the risk of DKA with HRs ranging from 1.1 to 2.0, depending on the T2DM and other antihyperglycemic agent definitions [18]. When compared with DPP-4 inhibitors, SGLT2 inhibitors were estimated with HRs ranging from 1.0 to 2.2 in different studies [19] [20]. The majority of the 95% confidence intervals of these estimates crossed the null.

However, inconsistent DKA diagnosing practice (e.g., misdiagnosis of mild ketosis in absence of clinical workup) has been reported [12]. Potential risk factors identified from the literature include insulin use, prolonged fasting, urinary tract or other severe infection, surgery or severe injury, alcohol use, cocaine use, hypovolemia, severe metabolic stress-related conditions (e.g., myocardial infarction, stroke, sepsis) and other drug exposures (glucocorticoids, atypical antipsychotic). Per the latest clinical guidelines, peri-operative considerations for temporarily discontinuing SGLT2 inhibitors are recommended [12] [13].

5.3 Diabetic Ketoacidosis in Ertugliflozin Trials and Safety Studies

According to the Steglatro™ Assessment Report published by EMA dated 25 January 2018 [1], in a pooled safety assessment using the last data analysed from 7 phase III studies, three (0.1%) ertugliflozin-treated patients were assessed to have met the case definition of ketoacidosis with either certain or possible likelihood compared to no cases in the comparator group (placebo or active control). Twenty-two other cases were determined either unlikely to represent ketoacidosis or unclassifiable. All events of ketoacidosis resolved, two after discontinuation of study medication and one resolved on treatment.

Up to date, there are no non-interventional safety studies assessing the risk of DKA associated with the use of ertugliflozin.

6 RESEARCH QUESTION AND OBJECTIVES

This study will address the research question of whether new use of ertugliflozin is associated with an increased risk of DKA, compared to new use of other non-SGLT2 inhibitor antihyperglycemic agents (AHAs), among T2DM patients.

The primary objectives of the study are:

- To assess the risk of DKA among new users of ertugliflozin relative to new users of sulfonylureas (SUs) or thiazolidinediones (TZDs);
- To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs [i.e. dipeptidyl peptidase 4 (DPP-4) inhibitors or glucagon-like peptide-1 (GLP-1) receptor agonists).

The secondary objectives of the study are:

- To assess the risk of DKA among new users of ertugliflozin relative to new users of SUs or TZDs, separately in insulin users and non-insulin users at baseline;
- To assess the risk of DKA among new users of ertugliflozin relative to new users of incretin-based drugs, separately in insulin users and non-insulin users at baseline.

7 RESEARCH METHODS

7.1 Study Design

A non-interventional cohort study using electronic healthcare data will be conducted to compare the risk of DKA between new users of ertugliflozin and new users of other non-SGLT2 inhibitor AHAs among T2DM patients. Eligible drug classes include SUs, TZDs, DPP-4 inhibitors, and GLP-1 receptor agonists (hereinafter referred to as "comparator AHAs"). These drug classes are chosen based on their similar place in therapy to ertugliflozin as the second-line treatment for T2DM and as the recommended 'add-on' medications to metformin [21] [22].

Metformin as a mono-therapy will not be included in the comparator group to reduce confounding by disease severity because new users of metformin are likely to be at an earlier stage of T2DM progressions and treatment. However, use of metformin in combination with ertugliflozin or a comparator AHA will be allowed, as the number of treatment-naïve patients before new initiation of ertugliflozin and a comparator AHA (recommended as second-line treatment for T2DM [21] [22]) are expected to be low.

Since the reduction of glucagon as a component of the mechanism of action of incretin-based therapies could in theory mitigate against the development of DKA, exposure to SUs or TZDs is included as a separate comparison group. Therefore, this study includes two head-to-head comparisons: ertugliflozin versus SU/TZD; and ertugliflozin versus incretin-based drugs (i.e. DPP-4 inhibitors or GLP-1 receptor agonists).

Propensity score matching will be used for confounding adjustment, followed by Cox proportional hazards models for risk estimation. Propensity score matching is an effective confounding adjustment approach widely used in pharmacoepidemiology studies [23] [24] [25]. The propensity score reduces large numbers of variables by summarizing baseline characteristics into a single score. Compared with standard multivariable regression modeling, this summary score approach makes risk adjustment and estimation feasible when the number of patients with health outcome of interest is low in relation to the number of covariates (a pattern commonly observed during early phase on newly-marketed medical products). On average, matching eligible members on propensity score balances the exposure cohorts with respect to distributions of baseline characteristics. Matched members can then enter the risk estimation model with minimal further adjustment [23].

Unless otherwise specified, this study will use pharmacy claims to define drug utilization and medical encounter claims to define existing conditions, medical history, or outcomes. National Drug Code (NDC) will be used to identify individual medications. Diagnosis and procedure codes encoded in the following coding systems will be used to identify individual medical conditions: International Classification of Diseases, 9th and 10th Revision, Clinical Modification (ICD-9-CM, ICD-10-CM), International Classification of Diseases, 10th

Revision, Procedure Coding System (ICD-10-PCS), Healthcare Common Procedure Coding System (HCPCS), and Current Procedural Terminology (CPT) codes.

7.2 Setting

This study will be conducted using existing administrative claims data in the Reagan-Udall Foundation for the FDA's Innovation in Medical Evidence Development and Surveillance System (IMEDS) Distributed Database (IMEDS-DD), a subset of the FDA Sentinel Distributed Database. The Sentinel Distributed Database is a national electronic system for active surveillance of the safety of drugs, biologics, vaccines, and medical devices in the US, established under the Sentinel Initiative [26] [27]. IMEDS-DD and Sentinel Distributed Database deploy the Sentinel Common Data Model [28] [29] for standardization of demographic and clinical data elements from various data partners.

The IMEDS-DD includes data from regional and national health insurers in US, some of which are integrated health care systems. Health plan members are predominately commercially-insured, community dwelling individuals, as such, the IMEDS-DD is expected to be largely representative of the commercially-insured population in US. However, if health plan members use any institutional services, the IMEDS-DD also includes data of such health services utilization. Members enrolled in Medicare-Advantage health and/or drug plans through IMEDS data partners may also be included.

Eligible patients will meet the following inclusion and exclusion criteria:

Inclusion criteria

- New users of ertugliflozin or new users of a comparator AHA beginning on 1 July 2018
- Age 18 years or older on the new initiation date (referred to as "index date") of ertugliflozin or a comparator AHA
- 6 or more months of continuous enrollment (maximum allowable enrollment gap of 45 days) in medical and prescription drug insurance plans before the index date
- T2DM, evidenced by at least one qualifying diagnosis recorded in claims of any encounter type any time before or on the index date. Qualifying diagnoses include ICD-9-CM *250.x0 Type II*, *Diabetes Mellitus* or ICD-10-CM *E11.x Type 2 Diabetes Mellitus*.

Exclusion criteria

- Type 1 diabetes (T1DM) or gestational diabetes, evidenced by at least one qualifying diagnosis recorded in claims of any encounter type any time before or on the index date. Qualifying diagnoses include ICD-9-CM *250.x1 and 250.x3 Type I Diabetes Mellitus*, *648.8x Abnormal Glucose Tolerance of Mother Complicating Pregnancy Childbirth or the Puerperium*, ICD-10-CM *E10.x Type 1 Diabetes Mellitus*, and *O24.4x Gestational Diabetes Mellitus in Pregnancy*

- Initiation of insulin on the index date, defined as insulin initiation on the index date with no prior use any time before the index date (note: history of insulin use that discontinues before index date and ongoing insulin use will be allowed)
- History of DKA, evidenced at least one qualifying discharge diagnosis of any position recorded in claims of inpatient encounters, any time before the index date. Qualifying diagnoses include ICD-9-CM *250.10 Type II Diabetes Mellitus with Ketoacidos*is, and ICD-10-CM *E11.1x Type II Diabetes Mellitus with Ketoacidosis*.

As a diagnosis of T1DM following T2DM may indicate that the T2DM diagnosis was incorrect, a sensitivity analysis will be conducted using a "narrow T2DM population" definition by excluding patients with T1DM diagnosis any time during the study (i.e., before, on or post-index date).

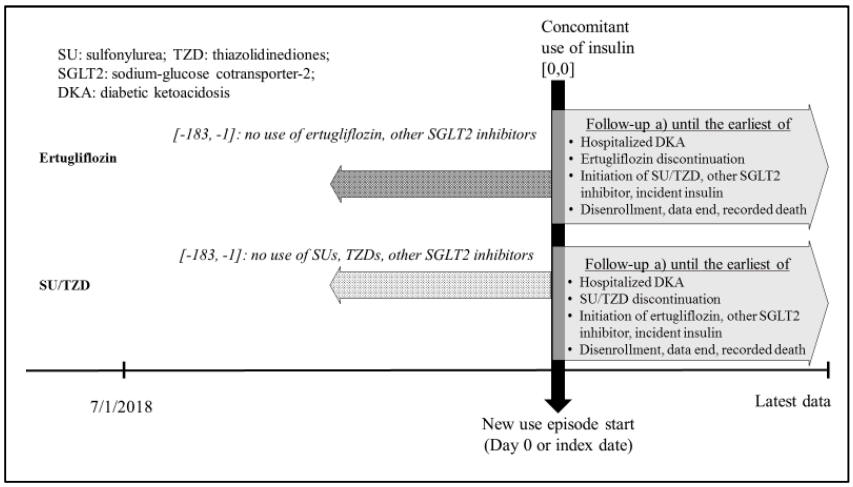
7.3 Variables

7.3.1 Exposure

The study population will be classified into three new user groups based on exposure: 1) ertugliflozin, 2) SUs or TZDs, and 3) incretin-based drugs (i.e., DPP-4 inhibitors, or GLP-1 receptor agonists). These exposure groups will be identified via outpatient pharmacy claims.

New user is defined as having a first exposure of the cohort-defining drug(s) (referred to as "index exposure"), but no prior use of the index exposure nor other non-ertugliflozin SGLT2 inhibitors in the 6 months before the index date. This is considered the "primary new user" definition.

Figure 1 Primary New User Definition (Sulfonylurea/Thiazolidinedione Comparator Example)



As antihyperglycemic agents are likely to be dispensed with 90-day supplies, a 6-month evaluation period prior to the index date is considered long enough to account for extended use of the discontinued exposure prior to the index date and a potential delay in effect.

A patient will be allowed to contribute to more than one exposure group or to the same exposure group more than once, as long as he/she qualifies as a new user of that exposure category (i.e., index exposure). For example, if a TZD new user starts on ertugliflozin right after the end of the last dispensing's days supply for the TZD, that patient will qualify as a new user of TZD and new user of ertugliflozin at the different time points.

Patients who switch exposure groups will have their follow-up censored for the original exposure group to which they were contributing exposure data. However, if they later become new users of the opposite exposure group (i.e., ertugliflozin new users become comparator AHA new users, and vice versa), then they contribute exposure data to the opposite exposure group until they discontinue or switch again. Patients who developed DKA over the course of the study will be censored and won't re-enter the study cohort. This approach was selected because diabetes is a progressive disease and subjects can be exposed to various antidiabetic medications over the course of this disease. The intent is to capture the various medication exposure categories over time and to attribute the outcome to the appropriate category at the time when the outcome occurs, thus accurately capturing exposure-outcome associations in the study. The limitation of this approach is that subjects will be captured in multiple exposure categories over time, and creating non-mutually exclusive categories of subjects, which can be more difficult to interpret.

A sensitivity analysis will be conducted with a more restrictive new user definition requiring no prior use of SGLT2 inhibitors (including ertugliflozin), a comparator AHA (i.e. SU/TZD

when comparing ertugliflozin with SU/TZD; or DPP-4 inhibitors/GLP-1 receptor agonists while comparing ertugliflozin with incretin-based medicines) in the 6 months before the index date. This is considered the "incident new user" definition.

Duration of use for AHAs, including ertugliflozin and comparator AHAs, will be established using the days supply per dispensing recorded in pharmacy claims and the specified grace period. The study considers days supply as evidence of the period in which a patient is covered for the dispensed medication. In the event of early refills, days supply will be stockpiled, and sum of days supply of the two overlapping dispensings will be newly assigned as the covered period. In the event of late refills, dispensing with a gap shorter than the grace period will be bridged, and the exact number of days in the gap will be considered by duration of use. For the last refill, the grace period will be considered by duration of use to account for potential medication overstock or residual biologic effect remaining in the system. Duration of use will be the total number of days summing the covered period, if any dispensing gap(s), and one grace period.

7.3.2 Outcome

The study outcome will be hospitalization for DKA, identified from principal discharge diagnosis of inpatient claims. Qualifying diagnoses include ICD-10-CM *E11.1x Type 2 Diabetes Mellitus with Ketoacidosis*. The admission date will be used as the diagnosis date.

A sensitivity analysis will be conducted by using the hospitalization for DKA identified from first- or second-listed inpatient diagnosis to capture as many DKA cases as possible, reflecting the DKA cases in the real-world settings.

7.3.3 Covariates

Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between two exposure groups in baseline covariates. The propensity score is the probability of a patient becoming an ertugliflozin versus comparator AHA new user, given a set of observed covariates [23] [24] [25] [30].

This study will include baseline demographics, Hemoglobin A_{1c} (HbA_{1c}, if available), use of antihyperglycemic agents, use of medications associated with DKA, comorbidity burden, pre-existing comorbidities, diabetic complications, lifestyle, and health services utilization as covariates in the propensity score estimation model. Two sets of propensity scores, one for the comparison of ertugliflozin versus SU/TZD and one for the comparison of ertugliflozin versus incretin-based drugs, will be generated. Ertugliflozin and comparator AHA new use will be 1:1 matched on propensity score by the nearest neighbor approach. Unless otherwise specified, all covariates listed in Table 1 will be evaluated within the 6 months prior to the index date, and medical conditions will be assessed using medical encounter claims from any care setting.

Table 1 List of Covariates

Category	Covariates
Demographics	Age, sex, calendar year of cohort entry
Lab data	HbA_{1c} (most recent, if available)
Use of antihyperglycemic agents	Distribution of diabetes treatment by class* (i.e., metformin, SU, TZD, alpha glucosidase or meglitinides, DPP-4 inhibitor, GLP-1 receptor agonist, SGLT2 inhibitor, and insulin)
Use of medications associated with DKA	Clozapine or olanzapine, lithium, terbutaline, oral corticosteroids, thiazides, pentamidine
Comorbidity burden	Combined comorbidity index [31]
Pre-existing comorbidities	Acute illness (i.e., serious infection, surgery, trauma, acute febrile illness, or sepsis), acute renal failure, cerebrovascular disease, myocardial infarction, stroke, coronary heart disease, heart failure, hypertension, hyperlipidemia, pancreatitis, hypovolemia, hypoxemia, thyroid disorders
Diabetic complications	Moderate to severe renal insufficiency (i.e. stage 3-5 chronic kidney disease or end stage renal disease) or diabetic nephropathy, neuropathy, retinopathy, peripheral vascular disease, amputation
Lifestyle	Obesity surgery, alcohol use, tobacco use, cocaine abuse
Health services utilization	Number of generic medication, unique pharmacological classes, dispensing, inpatient encounters, non-acute institutional encounters, emergency department encounters, ambulatory encounters, and other ambulatory† encounters

HbA_{1c}: haemoglobin A1c; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2; DKA: diabetic ketoacidosis.

† *Other ambulatory encounters include other non-overnight ambulatory encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations*

The approaches to handling concomitant AHAs in the analyses are summarized in Table 2.

Table 2 Approaches to Handling Concomitant Antihyperglycemic Agents (AHAs)

Timing and type of AHA Dispensing	Analysis Approach
If any AHA taken during the baseline period that is not eligible to be a study exposure	Include as covariate in propensity score estimation model
If any AHA taken during the baseline period that is eligible to be a study exposure	Include as covariate in propensity score estimation model when the "primary new user" definition as stated in Section 7.3.1 is used.
At the index date, patients will be classified according to their treatment complexity as receiving mono vs. dual vs. triple therapy.	Include as covariate in propensity score estimation model
If insulin is used at the index date	Conduct stratified analyses by insulin use at index date (Yes or No)
If non-ertugliflozin SGLT2 inhibitors are added during follow-up	End follow-up and censor
If any AHA that is not a study exposure have been added during follow-up	Allow during follow-up; and may include in the propensity score quarterly-based re-estimation if patients later become new users of a study exposure

7.4 Data Sources

7.4.1 Data source

This study will be conducted using existing, administrative claims data in the Innovation in Medical Evidence Development and Surveillance System (IMEDS) Distributed Database (IMEDS-DD). IMEDS is a public-private partnership launched in 2017 by the *Reagan-Udall Foundation for the Food and Drug Administration*, an independent, non-for-profit organization created by the US Congress, to advance the US FDA's mission by promoting regulatory science. IMEDS provides a framework for private-sector entities (e.g. regulated industry, academic institutes) to leverage the FDA Sentinel Distributed Database, a national electronic system for active surveillance of the safety of drugs, biologics, vaccines, and medical devices in the US, established under the Sentinel Initiative [26] [27]. The IMEDS-DD works with selected data partners from the Sentinel Distributed Database, as well as the

Harvard Pilgrim Health Care Institute (HPHCI; functioning as the IMEDS Analytic Center) and the Reagan-Udall Foundation, to provide real-world healthcare information on large patient populations in a timely manner, by facilitating efficient analyses of medical product safety evaluations.

As a subset of the Sentinel Distributed Database, the IMEDS-DD is expected to be largely representative of the commercially-insured population in US. At present, the IMEDS-DD have claims data available for research for over 95 million health plan members who have overlapping medical and pharmacy insurance coverage. The average enrollment length is similar to other claims databases of members with medical and pharmacy coverage - about 25% of patients have over three years of enrollment, and patients with chronic conditions such as diabetes and older members typically have longer than average enrollment periods within these databases.

As of December 2018, IMEDS-DD includes eight data partners, including national and regional health insurers in the US. All listed data partners have access to claims data and provide input and feedback for the study.

Brief descriptions of the data partners are provided below:

Aetna is one of the nation's leading healthcare benefits companies, serving 38 million people with information and resources to help them make better-informed decisions about their health care. Aetna became a FDA Sentinel data partner in 2010 and continues to be one of the largest contributors of data for public health purposes.

Harvard Pilgrim Health Care is one of the country’s premier health plans. It is large non-profit health plan with diverse enrollees across New England. HPHCI is a research and academic partnership between Harvard Medical School and Harvard Pilgrim Health Care. HPHCI also participates in the IMEDS program as the IMEDS Analytic Center.

HealthCore, Inc., a wholly-owned subsidiary of Anthem, Inc., uses real-world data to conduct outcomes, health economics, pharmacoepidemiologic, and late phase research. The HealthCore Integrated Research Database is a proprietary, fully integrated, longitudinal claims database that combines medical, pharmacy, and laboratory information drawn from 66 million unique individuals with medical coverage and nearly 48 million researchable lives with medical and pharmacy claims information since 2006. In addition, The HealthCore Integrated Research Environment has the ability to link the claims data in the HealthCore Integrated Research Database to complementary data sources, including inpatient and outpatient medical records, national vital statistics records, cancer and vaccine registries (state-by-state), disease and device registries, member and provider surveys, and point of care clinical data. Using these resources, HealthCore conducts a range of real world research designed to meet client needs, including retrospective database studies, medical record review studies, cross-sectional and longitudinal patient and provider surveys, and prospective site-based studies, including pragmatic clinical trials.

HealthPartners Institute is a 501(c)(3) nonprofit organization dedicated to conducting high-quality, public-domain health research, often in collaboration with other academic and research organizations throughout the world. It employs 33 career research investigators and more than 400 clinician researchers and encompasses vast and varied areas of research. The Institute is linked to an integrated health care system that provides health insurance for more than 1.5 million members and health care for more than 1 million patients.

Humana/Comprehensive Health Insights is a health economics and outcomes research subsidiary of Humana, which focuses on treatment effectiveness, drug safety, adherence, medical and pharmacy benefit design, disease management programs, and other healthcare services. Humana/CHI has been an active collaborator and data partner in the FDA Sentinel System, the Patient-Centered Outcomes Research Institute's National Patient-Centered Research Network (PCORnet), and several Distributed Research Network initiatives.

Marshfield Clinic Health System is an integrated health system serving mostly rural central, northern, and northwestern Wisconsin with more than 10,000 employees and more than 1,200 providers comprising 86 specialties, a health plan, and research and education programs. The Clinic had nearly 3.5 million patient encounters for the year ended in 30 September 2017, and reported approximately 328,000 unique patients in the health system during this same period. The Marshfield Clinic Research Institute is a division of the Marshfield Clinic Health System and is affiliated with the Security Health insurance plan. As a data partner, Marshfield has access to claims data from members of Security Health plan linked to electronic health record data for members who are cared for at the Marshfield Clinic Health System.

Meyers Primary Care Institute was established in 1996, as a joint endeavor of the University of Massachusetts Medical School, Reliant Medical Group (previously Fallon Clinic), and Fallon Health. The Institute is committed to research across the lifespan, from childhood to advanced age and conducts population-based research to inform policy and practice, promoting evidence-based care for the benefit of our community, and beyond. As a data partner, Meyers has access to claims data from members of Fallon Health which can be linked to electronic health record data for patients who are cared for at Reliant Medical Group practices.

Vanderbilt University Medical Center is a comprehensive health care facility dedicated to patient care, research, and the education of health care professionals. Its reputation for excellence in each of these areas has made Vanderbilt a major patient referral center for the South, Southeast and Midwest. The clinical enterprise has 12 adult specialties ranked among the nation's best according to the 2018 US News and World Report. The hospital has 758 licensed beds and conducts over 70,000 emergency department visits and 1.6 million ambulatory visits per year. Vanderbilt University Medical Center has been named among the nation's 100 "most-wired" hospitals and health systems for its efforts in

innovative medical technology 10 times. Vanderbilt University Medical Center researchers have access to Tennessee Medicaid (TennCare) data and have a long history of using these data for pharmacoepidemiologic research, with review and approval by the Division of TennCare.

The IMEDS-DD and Sentinel Distributed Database use the Sentinel Common Data Model [28] [29] for standardization of demographic and clinical data elements and have routine analytical tools (i.e., reusable, modular SAS programs) in place to permit rapid queries, including descriptive analyses and complex methodologies (e.g., comparative analyses), across data partners. Data partners contributing to the IMEDS-DD maintain their data in the Sentinel Common Data Model format. Specific information in the Sentinel Common Data Model includes, but is not limited to, the following types of data:

- *Enrollment* data: One record per covered individual per unique enrollment span is included in the Sentinel Common Data Model. Individuals are assigned a unique identifier by their insurer, which is linkable to all other data in the Sentinel Common Data Model. Due to changes in employment status, individuals may be enrolled multiple times with the same insurer, and the length of each given enrollment "span" may vary substantially. Each record in the enrollment file indicates the patient identifier, enrollment start and end dates, and whether the patient was enrolled in medical coverage, pharmacy coverage, or both during that range. Likewise, a final field indicates whether the data partner can request medical charts for a given patient during a given enrollment span.

- *Demographic* data, including birth date, sex, race/ethnicity, and the Zone Improvement Plan (ZIP) code of their most recently recorded primary residence.

- *Pharmacy dispensing* data, including the date of each prescription dispensing, the NDC identifier associated with the dispensed product, the nominal days supply, and the number of individual units (pills, tablets, vials, etc.) dispensed. Note that products purchased over the counter, or at some cash-only retail locations selling prescription drug products (e.g., through the Walmart Prescription Program) are not captured.

- *Medical encounter* data, including the healthcare provider most responsible for the encounter as well as the facility in which the encounter occurred and its ZIP code. Admission and discharge dates (if applicable) are also included, as is the encounter type (either an ambulatory visit, an emergency department visit, an inpatient hospital, a non-acute inpatient, or an otherwise unspecified ambulatory visit). Discharge disposition (alive, expired, or unknown) as well as discharge status (to where a patient was discharged) are also included for inpatient hospital stays and non-acute inpatient stays. Finally, laboratory data, are available for some, but not all, of the data partners; and the level of completeness for laboratory information for those data partners with such data varies [32].

- *Diagnosis* data, including the date of diagnosis, its associated encounter identifier, admission date, provider identifier, and encounter type. Diagnoses are recorded with ICD-9-CM and ICD-10-CM codes. For inpatient hospital and non-acute inpatient stay encounters, the Sentinel Common Data Model includes the principal discharge diagnosis.

- *Procedure* data, including the procedure date, its associated encounter identifier, admission date, provider identifier, and encounter type. Procedures are coded as ICD-9-CM and ICD-10-PCS procedure codes, CPT categories II, III, or IV codes, as well as HCPCS levels II and III codes.

- *Death* data, including the date of death, source of death information, whether the death month and day were imputed, and the degree of confidence in the record (excellent, fair, poor). Among the eight IMEDS-DD data partners, seven have death data and five have cause of death data [33]. Both death and cause of death information is substantially lagged (at least 2 years). Cause of death is coded as ICD-10-CM diagnosis codes.

Since its official launch in 2017, the Reagan-Udall Foundation for the FDA has completed two IMEDS projects, with several other projects currently in progress or planned. A pilot study demonstrated that a class-wide labeling change for proton pump inhibitors to address the increased risk of bone fracture for these products may have impacted subsequent prescribing pattern [34] [35]. Another important conclusion of this study was that large, distributed data networks such as the IMEDS-DD, could be a useful approach for measuring the effectiveness of risk minimisation strategies. A second recently performed investigation in the IMEDS-DD evaluated the safety of biologic and non-biologic disease modifying anti-rheumatic drugs and estimated the incidence of venous thromboembolism in patients diagnosed with rheumatoid arthritis [36]. The results of this study supported a regulatory submission to the FDA and has been presented in several international conferences focusing on rheumatology [37].

7.4.2 Study procedures

As described in 7.4.1, the IMEDS-DD is a subset of the Sentinel Distributed Database and shares the same data management, privacy protection methods, and quality assurance procedures with the Sentinel Distributed Database [38] [39] [40]. The Sentinel Distributed Database is compliant to the security requirements of the US Federal Information Security Management Act of 2002 (FISMA, specifically Moderate Risk Security Controls, as specified in the National Institute of Standards and Technology Special Publication 800-53) and has implemented policies and procedures to ensure the utmost data security, including an annual assessment process to ensure compliance.

The IMEDS-DD operates on a minimum necessary basis [28] [41] and implements a secure distributed querying environment to enable safe distribution of analytic queries, data transfer,

and document storage. In the IMEDS-DD, queries are sent securely by the IMEDS Analytic Center at HPHCI, and data partner responses are securely returned using a web-based distributed querying application (PopMedNet) [42] [43] administered by HPHCI. In this approach, data remain behind each data partner's local firewall, and data partners maintain physical and operational control of their data. In most cases, query results are returned to the web portal in aggregate form. All communications between the web portal and the application use HTTP/SSL/TLS connections to securely transfer queries and results.

As described in more detail in 8. Protection of Human Subjects, this study is subject to research ethics review by an Institutional Review Board/Independent Ethics Committee (IRB/IEC). Non-interventional studies using administrative claims data typically post no direct risk of harm to patients. Therefore, this study is expected to be determined as exempt from IRB/IEC review. Data used in this study will be anonymized and no personal identifiers will be available to maintain patient confidentiality.

7.4.3 Feasibility assessment

Feasibility assessment will be conducted to assess data relevancy and data quality of the IMEDS-DD. The study period for the feasibility assessment will be between 1 April 2016 and latest data available. The objective of this feasibility assessment is three-fold:

1. To assess the availability of key data elements (such as demographic, comorbidities, and HbA_{1c}) relevant to conduct this study;

2. To evaluate comparability of the IMEDS-DD T2DM population to the general T2DM population;

3. To assess the impact of various cohort definitions on the T2DM population and to assess DKA case definitions using existing ICD-10-CM codes

7.4.3.1 To assess the availability of key data elements in the IMEDS-DD

The study population for this feasibility assessment will be T2DM patients 18 years of age or older with a diagnosis of T2DM (ICD-10-CM codes *E11.x Type 2 Diabetes Mellitus*), evidenced by at least one qualifying diagnosis recorded in claims of any encounter type between 1 April 2016 and latest data available. Patients will be excluded if they have a diagnosis code for T1DM (ICD-10-CM *E10.x Type 1 Diabetes Mellitus)* or gestational diabetes (ICD-10-CM codes: *O24.4x Gestational Diabetes Mellitus in Pregnancy)*.

A descriptive analysis of the T2DM population will be performed. This will include age, sex, obesity (surgery), tobacco use, alcohol use, and distribution of diabetes treatment by class.

Presence of comorbid conditions will be assessed using all available data during the study period (i.e., cross-sectional assessment). The following conditions will be evaluated: acute renal failure, cerebrovascular disease, myocardial infarction, stroke, coronary heart disease,

heart failure, hypertension, hyperlipidemia, pancreatitis, hypovolemia, hypoxemia, thyroid disorders, renal insufficiency or diabetic nephropathy, neuropathy, retinopathy, peripheral vascular disease, amputation, and hospitalized DKA. Hospitalization for DKA will be identified by inpatient claims containing ICD-10-CM *E11.1x Type 2 Diabetes Mellitus with Ketoacidosis*, as proposed in the study protocol.

In addition, descriptive data will also include most recent lab value for Hemoglobin A_{1c} (HbA_{1c}) to explore the availability and quality of HbA_{1c} in the IMEDS-DD.

Distributions of each characteristic will be reported for the overall T2DM population using means and standard deviations for continuous variables and in counts and percentages for categorical variables.

7.4.3.2 To evaluate comparability of the IMEDS-DD T2DM population to the general T2DM population

Comparability of the IMEDS-DD T2DM population to the general T2DM population in the US will focus on the prevalence of T2DM and basic demographic breakdowns. Comparison will be conducted against the national or international benchmarks (e.g., T2DM prevalence statistics reported by the Centers for Disease Control and Prevention or by the International Diabetes Federation) and/or other published literature describing disease epidemiology.

More specifically, comparison between database T2DM population and the general T2DM population will focus on the prevalence of T2DM, demographic characteristics (age and sex), AHA use by class, most recent HbA_{1c}, prevalence of comorbid conditions among patients with T2DM including cardiovascular diseases, congestive heart failure, chronic kidney disease, neuropathy and retinopathy, and health service utilization (Table 3).

Table 3 Comparability between IMEDS-DD T2DM and General T2DM Population

	IMEDS-DD T2DM	General T2DM population
Prevalence of T2DM (n, %)		
Demographic characteristics		
Age (mean, SD)		
Male (n, %)		
Antihyperglycemic agent use (n, %) *		
Metformin		
SU		
TZD		
DPP-4 inhibitor		
GLP-1 receptor agonist		
SGLT-2 inhibitor		
Insulin		
Lab data		
HbA_{1c} (most recent) (mean, SD)		
Comorbidity burden		
Combined comorbidity index (mean, SD)		
Comorbidities (n, %)		
Cardiovascular disease		
Congestive heart failure		
Chronic kidney disease		
Neuropathy		
Retinopathy		
Health Service Utilization (mean, SD)		
Number of unique non-diabetic drugs		
Number of diabetic drugs		
Number of physician visits		
Number of hospitalizations		

SD: standard deviation; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2

* not mutually exclusive

7.4.3.3 To assess the impact of various cohort definitions on the T2DM population and to assess DKA case definitions using existing ICD-10-CM codes

7.4.3.3.1 Cohort definition of the T2DM population

The T2DM population definition proposed in the study protocol requires diagnoses of T2DM and no diagnosis of T1DM or gestational diabetes any time before or on the new initiation date of a study exposure (referred to as "pre-index period"). However, misdiagnosis of T2DM may occur when a diagnosis code for T2DM or T1DM is recorded only after the initiation of a study exposure (referred to as "post-index period") [44] [45] [46] [47].

To explore the extent of such T2DM misclassification, this feasibility assessment will evaluate among new users of second-line T2DM treatment (i.e. SU, SGLT2 inhibitors, DPP-4 inhibitors, GLP-1 receptor agonists, or TZDs), defined as patients who are 18 years or older and have no use of their index exposure in the 6 months before the new initiation date:

- how many patients receive their first T2DM diagnosis only after the initiation of an eligible drug class (i.e. no T2DM diagnosis any time before or on the new initiation date of a study exposure), and

- how many patients, identified as having T2DM and no T1DM based on diagnoses in the pre-index period, receive their first T1DM diagnosis only after the initiation of an eligible drug class.

Additionally, the T2DM population definition proposed in the study protocol does not require further exclusion of patients receiving both diagnoses of T2DM and unspecified T2DM. Considering misdiagnosis and miscoding of T2DM may also occur when a patient receives both diagnoses of *E11.x Type 2 Diabetes Mellitus* and *E13.x Other Specified Diabetes Mellitus* (including diabetes mellitus due to genetic defects of beta-cell function or in insulin action, postpancreatectomy or postprocedural diabetes mellitus, and secondary diabetes mellitus), this feasibility assessment will also evaluate the impact of including this mixed-diagnosed group of patients on T2DM population size in a separate analysis. Specifically, this analysis will use all available data in the IMEDS-DD during the study period for the feasibility assessment and compare the number of T2DM patients identified with *E11.x Type 2 Diabetes Mellitus* diagnoses to the number of T2DM patients identified with both *E11.x Type 2 Diabetes Mellitus* and *E13.x Other Specified Diabetes Mellitus* diagnoses. Depending on the magnitude of the difference, characteristics listed in 7.4.3.1 may be examined and compared between the two cohorts for qualitative differences in cohort composition, if necessary.

Based on the feasibility assessment results, decisions will be made for the final T2DM population definition regarding whether to 1) include post-index data as source of evidence for T2DM or T1DM eligibility, and/or 2) exclude patients who receive both *E11.x Type 2 Diabetes Mellitus* and *E13.x Other Specified Diabetes Mellitus* diagnoses.

7.4.3.3.2 Case definition of DKA hospitalization

To explore how to best capture DKA cases from inpatient claims using existing ICD-10-CM codes, this feasibility assessment will examine two approaches to identifying hospitalization for DKA: principal discharge diagnosis of 1) ICD-10-CM *E11.1x Type 2 Diabetes Mellitus with Ketoacidosis*, and 2) ICD-10-CM *E11.1x Type 2 Diabetes Mellitus with Ketoacidosis* or *E13.1x Other Specified Diabetes Mellitus with Ketoacidosis*. Difference in the number of DKA cases identified in the IMEDS-DD during the study period for the feasibility assessment will be quantified.

7.5 Study Size

Sample size estimates assuming different combinations of HR, power, and DKA incidence rate in the comparator AHA new users are provided in Table 4. The calculations assumed two-sided tests at a significance level of 0.05 (or type I error of 0.05) for power to be 80% and 90% (or type II error of 0.20 and 0.10, respectively). The number of events and person-years are estimates for the matched sample after 1:1 propensity score matching. These results assume proportional hazards and exponential survival times.

For example, in order to detect a HR of 2.0 or above in ertugliflozin users relative to comparator AHA for DKA, with targeted power of 80% and significance level of 0.05 in a two-sided test, a total of 66 DKA events from ertugliflozin and comparator AHA groups combined is required. This can be achieved by 8,819 person-years of ertugliflozin new users matched to comparator AHA new users in a 1:1 ratio on propensity score, assuming the DKA incidence rate is 2.5 per 1,000 person-years among T2DM patients treated with comparator AHAs [19] [18] [48].

Sample size calculation presented here reflects estimates meeting assumed values of HRs and DKA incidence rate in T2DM patients treated with comparator AHAs. As literature evolves, these assumptions may change over time. In general, when DKA incidence rate in T2DM patients treated with comparator AHAs increases, the required sample size to achieve the same power is expected to decrease, holding constant the total number of DKA cases needed for any pre-specified HR.

Table 4 Sample Size Calculation

Number of ertugliflozin-exposed person-years needed, by hazard ratios and incidence rate of diabetic ketoacidosis (DKA) in Type 2 diabetes mellitus patients treated with comparator AHAs

Hazard Ratio	Power	Total DKA Events	DKA Incidence Rate (per 1,000 Person-Years)				
			0.5	1.0	1.5	2.0	2.5
2.5	80%	38	21,726	10,869	7,250	5,440	4,355
2.0	80%	66	44,019	22,019	14,686	11,019	8,819
1.5	80%	192	153,650	76,850	51,250	38,450	30,770
2.5	90%	51	29,158	14,588	9,730	7,302	5,844
2.0	90%	88	58,692	29,358	19,581	14,692	11,758
1.5	90%	256	204,868	102,468	68,334	51,268	41,028

The number of events and person-years are estimates for the matched sample after 1:1 propensity score matching. These results assume two-sided tests with significance level= 0.05, proportional hazards and exponential survival times.

Time to accrue sufficient sample size will be driven by the commercial launch plan and the uptake of ertugliflozin in the US. After the feasibility assessment is completed, MSD will periodically monitor the amount of person-years of exposure to ertugliflozin that has been accumulated in the database in order to conduct the study and include that in the progress report(s). MSD agrees to perform the final analysis as soon as the study size meets the minimum requirement. However, if sample size requirement is not met upon five years after market entry of ertugliflozin, the MAH plans to perform the final analysis using all available data at this time point, recognizing that the study may have lower power than initially planned.

7.6 Data Management

As described in 7.4.1, the IMEDS-DD is a subset of the Sentinel Distributed Database and shares the same data management, privacy protection methods, and quality assurance procedures with the Sentinel Distributed Database [38] [39] [40]. The Sentinel Distributed Database is compliant to the security requirements of the US Federal Information Security Management Act of 2002 (FISMA, specifically Moderate Risk Security Controls, as specified in the National Institute of Standards and Technology Special Publication 800-53) and has implemented policies and procedures to ensure the utmost data security, including an annual assessment process to ensure compliance.

The IMEDS-DD operates on a minimum necessary basis [28] [41] and implements a secure distributed querying environment to enable safe distribution of analytic queries, data transfer, and document storage. In the IMEDS-DD, queries are sent securely by the IMEDS Analytic Center at HPHCI, and data partner responses are securely returned using a web-based distributed querying application (PopMedNet) [42] [43] administered by HPHCI. In this

approach, data remain behind each data partner's local firewall, and data partners maintain physical and operational control of their data. In most cases, query results are returned to the web portal in aggregate form. All communications between the web portal and the application use HTTP/SSL/TLS connections to securely transfer queries and results.

The IMEDS-DD deploys the Sentinel Common Data Model [28] [29] to allow data standardization across data partners in the network. Only data elements of Sentinel Common Data Model are available for queries, including demographics, health plan enrollment, diagnoses, procedures, and outpatient pharmacy dispensing records. During query execution, analytic programs based on SAS software will be used. Data management and conversion of the Sentinel Common Data Model to analysis variables will be performed using SAS software version 9.4 and above (SAS Institute, Inc., Cary, North Carolina).

For quality assurance of IMEDS-DD data, refer to 7.8 Quality Control.

7.7 Data Analysis

The analyses will be conducted separately for the comparison between new use of ertugliflozin versus new use of SU/TZD and the comparison between new use of ertugliflozin versus new use of incretin-based drugs.

7.7.1 Follow-up

For each comparison, four different methods will be used to define follow-up time in this study:

a) The primary analysis will use the "as-treated" approach. Follow-up for each new use of a given exposure will begin on the index date until the earliest of hospitalized DKA or any of the following censoring criteria met:

- Discontinuation of the index exposure, defined as last refill date plus days supply on the last refill plus 30 days
- Initiation of the opposite exposure (i.e., ertugliflozin new users starting a SU/TZD; ertugliflozin new users starting an incretin-based drugs; or vice versa)
- Initiation of other SGLT2 inhibitor(s)
- Initiation of insulin, defined as no insulin use any time before the date of this initiation
- Disenrollment from either medical or prescription drug insurance plan

- End of data availability
- Recorded death

As a patient will be allowed to contribute to more than one exposure group or to the same exposure group more than once, as long as he/she qualifies as a new user of that exposure category during the course of the study period, he/she can contribute follow-up time to both exposure groups or contribute follow-up time to the same exposure group more than once. Each time, from the initiation of a study exposure (i.e. index exposure) to the end of follow up of the index exposure is defined as one "new use episode". The total person-years for a given study exposure will be the sum of total follow-up time contributed by all qualified new use episodes.

b) A sensitivity analysis will use the "intent-to-treat" approach. Patients will be followed from the earliest index date until the earliest of hospitalized DKA or the end of their observation time, regardless of whether they discontinue the index exposure or, switch to or add on the opposite exposure. DKA diagnosis will be attributed to the index drug exposure category, even if the patient switched or discontinued the use of the index drug.

c) A second sensitivity analysis will be performed using "as-treated with a 90-day grace period" approach for which the censoring criterion for discontinuation of the index exposure is changed to "last refill date plus days supply on the last refill plus **90** days". 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, extended use of the discontinued index drug and a delay in effect.

d) A third sensitivity analysis will be performed using the "as-treated approach with no censoring at switching (overlapping period) or treatment augmentation", to ensure the patient is at risk during the entire time when they are exposed to the index exposure. DKA diagnosis will be attributed to the index drug exposure category, even if they augment therapy with an opposite exposure or during the overlapping period of switching.

7.7.2 Descriptive analysis

Baseline demographic and clinical characteristics by exposure will be reported in mean and standard deviation for continuous variables and in count and percentage for categorical variables before and after propensity score-matching. Cohort size, average follow-up time, and DKA incidence rate before and after propensity score-matching by exposure will also be reported. DKA incidence rate will be summarized in point estimates (per 1,000 person-years) and 95% confidence intervals (CIs). See the expected outputs for baseline characteristics of the study cohort by exposure in Appendix 12.1-12.4 (ertugliflozin versus SU/TZD;

ertugliflozin versus incretin-based drugs; before and after propensity score matching separately). Descriptive data of DKA cases among new users of ertugliflozin, incretin-based drugs, and SU/TZD will be provided in Appendix 12.5-12.6.

7.7.3 Comparative analysis

7.7.3.1 Primary and subgroup analysis

The primary analysis will compare the new users of ertugliflozin versus new users of SU/TZD; and compare new users of ertugliflozin versus new users of incretin-based drugs on risk of DKA, based on "primary new user" definition specified in Section 7.3.1.

The differences between the exposure groups in time to DKA will be illustrated using Kaplan Meier survival curves with log rank test (see the expected outputs Appendix 12.7).

A time-to-event analysis will be conducted separately for two comparisons. Cox proportional hazards models, with adjustment for within-subject correlation, will be used for risk estimation.

HRs and their 95% CIs before and after propensity score-matching will be reported. See the expected outputs for risk estimates in Appendix 12.8 (ertugliflozin versus SU/TZD) and Appendix 12.9 (ertugliflozin versus incretin-based drugs).

Subgroup analyses stratified by concomitant insulin use on the index date will be conducted because insulin use is clinically considered to be associated with a longer history of diabetes or more advanced diabetes. An observational study using the Truven Health MarketScan Commercial Claims and Encounters (CCAE) database has shown that a substantial proportion (37% in SGLT2 inhibitors versus 13% in non-SGLT2 inhibitors) of the DKA cases occurred among patients who were taking insulin before the index date [18]. Concomitant insulin use is defined as any prescription claims whose duration plus a 30-day grace period includes the index date.

7.7.3.2 Sensitivity analysis

The sensitivity analysis with varying T2DM definition (7.2), new user definitions (7.3.1), DKA definition (7.3.2), different approaches to define follow-up time (7.7.1), and propensity score stratification (7.7.3.3) will be conducted to assess the robustness of study results. See the overview of comparative analyses in Table 5 and the expected outputs for risk estimates in Appendix 12.10.

Table 5 Overview of Comparative Analyses

	T2DM definition	DKA definition	New user definition	Follow-up approach	Subgroup analysis	Propensity score analysis type
Primary analysis	Primary T2DM definition	Principal discharge diagnosis	Primary new user definition	As-treated approach with 30-day grace period	Concomitant insulin use at baseline	1:1 matching
Sensitivity analysis 1	Primary T2DM definition	Principal discharge diagnosis	Primary new user definition	Intent-to-treat approach	Not applicable	1:1 matching
Sensitivity analysis 2	Primary T2DM definition	Principal discharge diagnosis	Primary new user definition	As-treated approach with 90-day grace period	Not applicable	1:1 matching
Sensitivity analysis 3	Primary T2DM definition	Principal discharge diagnosis	Primary new user definition	As-treated approach with no censoring at switching (overlapping period) or treatment augmentation	Not applicable	1:1 matching
Sensitivity analysis 4	Primary T2DM definition	Principal discharge diagnosis	Incident new user definition	As-treated approach with 30-day grace period	Not applicable	1:1 matching
Sensitivity analysis 5	Narrow T2DM definition	Principal discharge diagnosis	Primary new user definition	As-treated approach with 30-day grace period	Not applicable	1:1 matching
Sensitivity analysis 6	Primary T2DM definition	First- or second-listed inpatient diagnosis	Primary new user definition	As-treated approach with 30-day grace period	Not applicable	1:1 matching
Sensitivity analysis 7	Primary T2DM definition	Principal discharge diagnosis	Primary new user	As-treated approach with 30-day grace period	Not applicable	Stratification

7.7.3.3 Model specifications for propensity score estimation

Two sets of analyses will be performed: one for the comparison of ertugliflozin versus SU/TZD and one for the comparison of ertugliflozin versus incretin-based drugs. Each of the primary and sensitivity analyses will fit a logistic regression model to estimate the propensity score for new use of ertugliflozin (i.e., probability of newly initiating ertugliflozin versus SU/TZD; and probability of newly initiating ertugliflozin versus incretin-based drugs). Observations with missing information on covariates will be removed from the estimation. Subgroup analyses will use the propensity score estimated in the primary analysis and rematch new use episodes within subgroup levels and within the matched new use episodes of the primary analysis (i.e., subgroup cohorts are subsets of primary analysis cohorts). All baseline covariates, including concomitant medication and antihyperglycemic agent use, will be considered in the propensity score estimation model as independent variables. Ertugliflozin and comparator AHA new use will be 1:1 matched on propensity score by the nearest neighbor approach [49]. Nearest neighbor matching selects for matching to a given ertugliflozin subject that comparator subject whose propensity score is closest to that of the ertugliflozin subject within a pre-specified caliper. The caliper plays an important role in quality control of the matches by placing restrictions upon the maximum acceptable difference between the propensity scores of two matched subjects. Calipers frequently used in propensity score matching are 0.01, 0.025, and 0.05 on the natural scale of the propensity score. A narrow caliper leads to closer matches and therefore reduces bias, although it also increases risk of no match. If multiple comparator subjects have propensity scores that are equally close to that of the ertugliflozin subject, one of these comparator subjects is selected at random [49] [50].

A patient will be allowed to contribute to more than one exposure group at a different time points during the study period whenever he/she qualifies as a new use definition of that exposure category. To allow the changing drug utilization patterns over time and to minimize the impact of self-matching across exposure groups among patients who become eligible for both exposure groups at different time points, the propensity score estimation and 1:1 propensity score matching will be performed repeatedly on a quarterly basis [51]. Based on the new user definition, each patient very likely belongs to only one exposure group in each quarter and will be matched on the refreshed propensity score with an eligible patient of the opposite exposure group during the same quarter. Only new users within the same quarter are allowed to be matched. Each matched new user will be followed from their index dates until outcome occurrence or censoring criteria are met. If a matched new user is censored, this patient then becomes eligible as a new user for either exposure groups again, as long as s/he satisfies the new user definition.

The re-estimation of propensity scores on a quarterly basis will essentially accommodate potential patient risk profile updates driven by their most recent characteristic changes (e.g., newly diagnosed comorbidities, newly added AHAs). Compared with one-time propensity score analytic procedures, this proposed recurring technique offers an opportunity not only to observe real-world changing patterns of AHA utilization among T2DM patients over time

but also to distinguish risk for hospitalized DKA associated with various exposure categories using baseline characteristics most relevant in time to the initiation of the index exposure. Additionally, the sequential nature of this recurring technique can also temporally segregate estimation of a patient's probability of receiving new medical products such as ertugliflozin during its adoption period and in turn partially mitigates potential channeling bias commonly observed in beginning of the market entry [51]. Moreover, the recurring technique has statistical advantages. In the event that model convergence issues occur during the propensity score estimation process in one quarter (e.g., due to insufficient new user sample size in one of the two exposure categories), the baseline characteristics will continue to contribute to (or sometimes power) the propensity score re-estimation in the next quarter(s). That is, if the propensity score estimation cannot be completed, the collected covariate distribution information by the exposure category is reserved and later reused, along with new information becoming available in the next quarter, to support the re-estimation in a larger sample. Similar methods have been previously used by the U.S. FDA in several prospective, post-market drug safety surveillance activities [52] [53] [54] [55].

Covariate distributions by exposure will be output before and after propensity score-matching. Multiple propensity score-matching diagnostics per analysis will be applied - between exposure groups, standardized difference will be used, with difference < 0.1 indicating covariate balance [49]; within each data partner site and per monitor quarter, c statistics and propensity score histograms per will be examined [49].

As a sensitivity analysis, propensity score stratification will be used to examine the robustness of study results using the entire data set for the analysis. Propensity score stratification is an alternative propensity score analysis method to achieve confounding adjustment. Unlike propensity score matching, this method runs no risk of losing patients when no match can be found. Propensity score stratification preserves all new users of both ertugliflozin and comparator AHA groups for the analysis, instead of just those who have a match. The increased sample size improves precision of the effect estimates and finding generalizability to the real-world T2DM patients who receive treatment. For each exposure group, individual patients are assigned to a propensity score stratum based on their score ranking. The finer the stratum is (or the larger the number of total strata is), the more similar patients of the two exposure groups within the same stratum become (i.e., balanced covariate distributions), and the more bias can be reduced [49]. Our study will stratify the propensity scores into deciles.

To account for variation in health insurance features among IMEDS-DD data partners, propensity score estimation, matching, and decile assignment will be performed locally within individual data partner sites. Once propensity score procedures are complete, the IMEDS Analytic Center will combine data from all data partners and perform the outcome analysis (i.e., risk estimation). Aggregated results will be presented in the final report. This is the same approach used by the FDA Sentinel Initiative for similar analyses. Should a data partner have very limited (or no) ertugliflozin exposure over the course of the study period; that data partner will be excluded from the data analysis.

7.7.4 Exploratory analysis

Additionally, as our "primary new user" definition allows subjects to be captured in multiple exposure categories over time, this will create non-mutually exclusive categories of subjects, which can be more difficult to interpret. To help address this concern and explore the potential impact to the validity of primary analysis results, the follow-up time for each defined new user will be further divided into the following mutually exclusive categories, wherever applicable:

- Ertugliflozin exposure person-time excluding any time overlapping with comparator AHA group person-time;
- Comparator AHA exposure person-time excluding any time overlapping with ertugliflozin person-time;
- Multiple exposure person-time of ertugliflozin overlapping with a comparator AHA.

For example, if a TZD new user starts on 9-month treatment of ertugliflozin right after the end of 6-month use of TZD, that patient will qualify as a new user of TZD and new user of ertugliflozin at the different time points. In the exploratory analysis, the follow-up time for this patient will be:

- TZD exposure person-time excluding any time overlapping with ertugliflozin person-time: 6 months (i.e. Months 1-6);
- Multiple exposure person-time of ertugliflozin overlapping with TZD: 30 days after the end of the last dispensing's days supply for the TZD (i.e. Month 7);
- Ertugliflozin exposure person-time excluding any time overlapping with TZD: 8 months (i.e. Months 8-15).

The exploratory analysis will be conducted separately, one for the comparison of ertugliflozin versus SU/TZD and one for the comparison of ertugliflozin versus incretin-based drugs. Number of DKA cases will be reported in all 3 categories as an exploratory analysis.

7.8 Quality Control

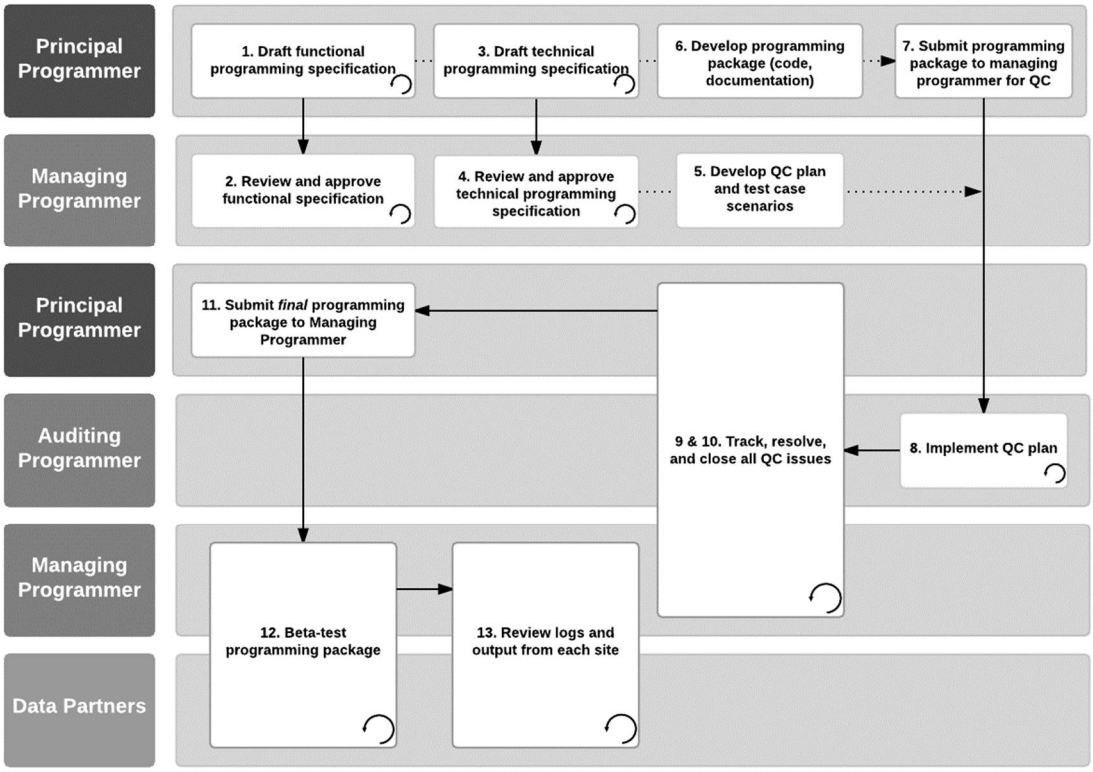
As described in 7.4.1, the IMEDS-DD is a subset of Sentinel Distributed Database and shares the same data quality assurance procedures with Sentinel Distributed Database. The quality assurance approach assesses consistency with the Sentinel Common Data Model, evaluates adherence to data model requirements and definitions, evaluates logical relationships between data model tables, and reviews trends in medical and pharmacy services use within and across data partners. Full quality assurance process and details on the Sentinel data

curation approach are documented on the Sentinel website [40] [56]. The data curation approach is consistent with guidance set forth by the US FDA in its current recommendations for data quality assurance, specifically - “Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data” (Guidance), section IV.E “Best Practices – Data Sources: Quality Assurance (QA) and Quality Control (QC)”, published in May 2013 [57]. This Guidance describes best practices that particularly apply to observational studies designed to assess the risk associated with a drug exposure using electronic healthcare data.

In addition to quality assurance of data elements, the IMEDS Analytic Center adopts standard SAS programming quality assurance and quality control processes used by the Sentinel System to check SAS programs and deliverables. Figure 2 illustrates the standard operating procedures (SOPs) for SAS programming quality assurance and quality control in the Sentinel System.

By signing this protocol, the investigators agree to be responsible for implementing and maintaining a quality management system with written development procedures and functional area SOPs to ensure that studies are conducted and data are generated, documented, and reported in compliance with the protocol, accepted standards of Good Clinical and Pharmacoepidemiology Practice, and all applicable federal, state, and local laws, rules and regulations relating to the conduct of the study.

Figure 2 Standard Operating Procedure for SAS Programming Quality Assurance and Quality Control in the Sentinel System



7.9 Limitations of the Research Methods

This study will have several limitations due to the proposed study design, data source, and analytic methods.

Firstly, as with any other non-interventional studies, this study is susceptible to the confounding and bias related to its non-randomized design. Unlike randomized controlled trials, non-randomized studies have less direct control over the study sample. Specifically, for newly-marketed medications such as ertugliflozin, channeling bias is prone to occur. Channeling occurs when pharmacotherapies with similar indications are differentially prescribed to patients with varying baseline risk profiles. This pattern can be triggered by either patient or prescriber preference – patients who are further advanced in indicated disease are more likely to try newly-marketed products, whereas practitioners who tend to early adopt new technologies are more likely to prescribe newly-marketed products. Consequently, risk estimates may be biased toward or away from the null [58]. Propensity score matching mitigates this issue by creating cohorts with balanced observed risk profiles, especially among pharmacotherapies recommended for similar place in therapy (both

ertugliflozin and comparator AHAs are recommended as second-line treatment for T2DM). However, the balance comes with a cost of the representativeness of the final analytic cohorts. Depending on the percentage of ertugliflozin users that we are able to match to comparator AHA, generalizability of study results may need to be interpreted with caution. Findings from proposed analyses in this protocol may not apply to all T2DM patients or to those treated with different glycemic control regimens.

Secondly, in database studies using health insurance administrative claims, evidence of patient medical history, treatment exposure, and outcome occurrence is typically captured by health services utilization. This type of data is collected and maintained for billing or record-keeping purposes. Most of the time, only services during medical encounters are recorded, excluding those not covered by health plans (e.g., over-the-counter medications and free drug samples) or not itemized under coverage (e.g., bundle payment for inpatient encounters). Clinical details (e.g., severity indices or laboratory results) and death are often missing or incomplete. Both systematic and random errors during cohort identification phase may occur. As a result, misclassification and residual or unmeasured confounding, is possible. However, proxy measures using pharmacy claims or condition algorithms combining diagnosis/procedure codes, encounter type, and diagnosis position may sometimes suffice [59]. To the extent possible, diagnosis or procedure codes recorded in medical encounter claims will be used as indicators for medical conditions, which will then serve as surrogates for various lifestyles and existing conditions. Underreporting for substance use or abuse (such as alcohol and cocaine abuse) is another common limitation shared by most claims-based database studies. To the extent possible, we will still capture these behaviors, using diagnosis and procedure codes, to characterize the study cohort.

Thirdly, exposure is inferred from prescription claims in this study, and days supply information from outpatient pharmacy claims will be used to approximate on-treatment time for individual T2DM patients exposed to ertugliflozin and comparator AHAs, both of which are self-administered oral medications. A prescription claims do not necessarily mean that patient consumed the drug and is a surrogate measure of drug exposure. This could result in exposure misclassification, if subjects who did not consume the drug are categorized as exposed to the drug, the results will be biased towards the null, i.e. missing an association if one exists. Furthermore, polypharmacy and switching of medications inherent to this patient population complicates efforts of each drug on the outcome of interest when subjects are exposed to multiple drugs simultaneously.

Fourthly, death data are used as one of the censoring criteria to calculate the follow-up time. However, death data is not available for every subject in the study. Only seven of the eight IMEDS-DD data partners have death data, which are substantially lagged (assume two or more years).

Fifthly, as a distributed data network, the IMEDS-DD does not guarantee data uniqueness at the patient or patient-time level. During the study period, a patient may contribute data to multiple data partners. However, given that the same health service utilization is

uncommonly covered by multiple health plans, repeated observation of the same patient-time in the IMEDS-DD is minimized.

Finally, the IMEDS DD is expected to be largely representative of the commercially-insured population and may not be representative of the entire treated population, such as the uninsured or elderly people of US population with Medicare insurance. As such, the study results will be generalizable to the commercial health insurance population from which the study population will be derived as well as others with similar characteristics.

8 PROTECTION OF HUMAN SUBJECTS

The proposed study is a secondary data analysis using an existing database IMEDS-DD. As such, there is no direct risk of harm to patients. Data used in this study will be anonymized and no personal identifiers will be available to maintain patient confidentiality. Studies of this type are typically determined to be exempt from IRB review.

The Reagan-Udall Foundation for the FDA has the responsibility to obtain approval of the study protocol, protocol amendments, and other relevant documents, if applicable, from an IRB/IEC. Participating data partners can either cede IRB review to the Reagan-Udall Foundation for the FDA or seek approval from their local IRB. All correspondence with the IRB/IEC will be retained in the Investigator File.

The study will be conducted in accordance with all legal and regulatory requirements. Additionally, we will adhere to commonly accepted research practices, including those described in European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology, Guidelines for Good Pharmacoepidemiology Practices issued by the International Society for Pharmacoepidemiology, FDA Guidance for Industry: Good Pharmacovigilance and Pharmacoepidemiologic Assessment, and FDA Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting of Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets [57].

8.1 Informed Consent

Not applicable.

8.1.1 Consent and collection of specimens for future biomedical research

Not applicable.

9 MANAGEMENT AND RESPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

9.1 Adverse Event Reporting

This is a non-interventional database study based on secondary use of data collected for other purposes. No administration of any therapeutic or prophylactic agent is required in this protocol. No reporting of individual adverse events (AEs) to regulatory agencies is planned for this database study because there is no access to individual patient/subject records and it is not possible to assess the causality of individual cases. Pre-specified health outcomes of interest, including any that qualify as adverse events, will be summarized as part of any interim analysis (including safety analysis, if required) and in the final study report, which will be provided to regulatory agencies by the Sponsor as required.

Any relevant safety information will be summarized in the appropriate Periodic Safety Update Report (PSUR)/Periodic Benefit Risk Evaluation Report (PBRER) and/or Development Safety Update Reports (DSUR) if required.

If an investigator elects to spontaneously report any suspected adverse reactions, they should be reported via fax to MNSC FAX PPD in English using an AE form (as specified in Annex 1 List of Stand-Alone Documents) for reporting to worldwide regulatory agencies as appropriate.

10 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

The study progress will be reported by MSD in regulatory communications in line with the EMA risk management plan, Periodic Safety Update Reports, and other regulatory milestones and requirements. Per commitment to the EMA, a feasibility assessment report will be submitted by end of Q4 2020 and a final study report is estimated to be submitted no later than the end of 2023. The study report will provide an overview of the study background, objectives, methods, and findings. Study results, as well as the main methodological components developed as part of this study, will also be disseminated at scientific meetings as oral or poster presentations and as peer-reviewed publications.

The IMEDS reserves the right to submit the results from any of the study analyses for publication. Any publications will follow guidelines, including those for authorship (e.g. guidelines established by the International Committee of Medical Journal Editors [60]) and for reporting of observational studies in epidemiology (e.g., Strengthening the Reporting of Observational Studies in Epidemiology checklists [61]).

11 REFERENCES

[1]	European Medicines Agency. Assessment report: Steglatro. London (England): European Medicines Agency (EMA); 2018. 139 p. EMA/86938/2018.	052V28
[2]	European Medicines Agency. Review of diabetes medicines called SGLT2 inhibitors started. London (England): European Medicines Agency (EMA); 2015. 2 p. EMA/390062/2015.	052TRB
[3]	Food and Drug Administration. Drug Safety Communications: FDA warns that SGLT2 inhibitors for diabetes may result in a serious condition of too much acid in the blood [Safety Announcement]. Silver Spring (MD): Food and Drug Administration (FDA); 2015 May 15. 4 p.	052W83
[4]	European Medicines Agency. Product Information as approved by the CHMP on 25 February 2016, pending endorsement by the European Commission: Amendments to relevant sections of the summary of product characteristics and package leaflet for Forxiga, Xigduo, Invokana, Vokanamet, Jardiance and Synjardy. London (England): European Medicines Agency (EMA). 2016.	052TRX
[5]	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;38:1680-6.	052VSR
[6]	Peters AL, Buschur EO, Buse JB, Cohan P, Diner JC, Hirsch IB. Euglycemic diabetic ketoacidosis: a potential complication of treatment with sodium-glucose cotransporter 2 inhibition. Diabetes Care. 2015 Sep;38:1687-93.	052WPV
[7]	Rosenstock J, Ferrannini E. Euglycemic diabetic ketoacidosis: a predictable, detectable, and preventable safety concern with SGLT2 inhibitors. Diabetes Care. 2015 Sep;38:1638-42. Comment in: Diabetes Care. 2016 Aug;39: e139-40.	052W42
[8]	Fadini GP, Bonora BM, Avogaro A. SGLT2 inhibitors and diabetic ketoacidosis: data from the FDA adverse event reporting system. Diabetologia. 2017 Aug;60(8):1385-1389.	04Q5H8
[9]	Blau JE, Tella SH, Taylor SI, Rother KI. Ketoacidosis associated with SGLT2 inhibitor treatment: analysis of FAERS data. Diabetes Metab Res Rev. 2017;33:e2924.	052TMK

[10] D'Elia JA, Segal AR, Bayliss GP, Weinrauch LA. Sodium-glucose cotransporter-2 inhibition and acidosis in patients with type 2 diabetes: a review of US FDA data and possible conclusions. Int J Nephrol Renovasc Dis. 2017:10:153-8. 052VSC

[11] European Medicines Agency. Assessment report: procedure under Article 20 of Regulation (EC) No 726/2004 resulting from pharmacovigilance data: SGLT 2 inhibitors. London (England): European Medicines Agency (EMA); 2016. 23 p. EMA/PRAC/50218/2016. 052TSG

[12] Handelsman Y, Henry RR, Bloomgarden ZT, Dagogo-Jack S, DeFronzo RA, Einhorn D, et al. American Association of Clinical Endocrinologists and American College of Endocrinology position statement on the association of SGLT-2 inhibitors and diabetic ketoacidosis. Endocr Pract. 2016 Jun;22(6):753-62. 052W98

[13] Dashora U, Gallagher A, Dhatariya K, Winocour P, Gregory R. Association of British Clinical Diabetologists (ABCD) position statement on the risk of diabetic ketoacidosis associated with the use of sodium-glucose cotransporter-2 inhibitors. Br J Diabetes. 2016 Oct-Nov-Dec;16(4):206-9. 052TR2

[14] 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. 052X4B

[15] 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. 04QDG7

[16] Jensen ML, Persson F, Andersen GS, Ridderstrale M, Nolan JJ, Carstensen B, et al. Incidence of ketoacidosis in the Danish type 2 diabetes population before and after introduction of sodium-glucose cotransporter 2 inhibitors - a nationwide, retrospective cohort study, 1995-2014. Diabetes Care. 2017 May;40:e57-8. 052W99

[17] Takeuchi M, Kawamura T, Sato I, Kawakami K. Population-based incidence of diabetic ketoacidosis in type 2 diabetes: medical claims data analysis in Japan. Pharmacoepidemiol Drug Saf. 2018;27:123-26. 052WH7

[18] Wang Y, Desai M, Ryan PB, DeFalco FJ, Schuemie MJ, Stang PE, et al. Incidence of diabetic ketoacidosis among patients with type 2 diabetes mellitus treated with SGLT2 inhibitors and other antihyperglycemic agents. Diabetes Res Clin Pract. 2017;128:83-90. 052Y70

[19] Fralick M, Schneeweiss S, Patorno E. Risk of diabetic ketoacidosis after initiation of an SGLT2 inhibitor [correspondence]. N Engl J Med. 2017 Jun 8;376(23):2300-2. 052V37

[20] Kim YG, Jeon JY, Han SJ, Kim DJ, Lee KW, Kim HJ. Sodium-glucose co-transporter-2 inhibitors and the risk of ketoacidosis in patients with type 2 diabetes mellitus: a nationwide population-based cohort study. Diabetes Obes Metab. 2018;20:1852-8. 052X49

[21] American Diabetes Association. 8. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2018. Diabetes Care. 2018 Jan;41(Suppl 1):S73-85. 04YCLB

[22] National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. London (England): National Institute for Health and Care Excellence (NICE); 2015 Dec 2. 44 p. 052W8Y

[23] Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoep Drug Safety 2010;19(8):1-22. 00W5HN

[24] Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98:253-9. 052V45

[25] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55. 052WSR

[26] Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System - a national resource for evidence development. N Engl J Med. 2011 Feb 10;364(6):498-9. 052TMC

[27] Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf. 2012;21(suppl 1):1-8. 052WPW

[28] Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf. 2012;21(suppl 1):23-31. 052TNG

[29] Sentinel [Internet]. Silver Spring (MD): Food and Drug Administration (FDA); c2010-2018. Distributed database and common data model; [cited 2018 Oct 29]; [about 3 screens]. Available from: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model. 052Y0K

[30] Joffe MM, Rosenbaum PR. Invited commentary: propensity scores [commentary]. Am J Epidemiol. 1999 Aug 15;150(4):327-33. 052X48

[31] Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol. 2011;64:749-59. 052TS6

[32] Raebel MA, Haynes K, Woodworth TS, Saylor G, Cavagnaro E, Coughlin KO, et al. Electronic clinical laboratory test results data tables: lessons from Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2014;23:609-18. 052WSP

[33] Swain RS. Feasibility analysis of mortality outcomes in the sentinel distributed database. Slides presented at: 33rd International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE); 2017 Aug 26-30; Montreal, QC. 052X67

[34] Sobel R. Insights from phase II of the IMEDS evaluation pilot - lessons learned and future needs. Slides presented at: Ninth Annual Sentinel Initiative Public Workshop; 2017 Feb 2; Washington, DC. 052Y0S

[35] Sobel RE, Bate A, Marshall J, Haynes K, Selvam N, Nair V, et al. Do FDA label changes work? Assessment of the 2010 class label change for proton pump inhibitors using the Sentinel System's analytic tools. Pharmacoepidemiol Drug Saf. 2018;27:332-39. 052WGG

[36] Salinas CA, Hornbuckle K, Reynolds L. Venous thromboembolism in rheumatoid arthritis. Slides presented at: 2018 Sentinel Initiative Annual Public Workshop; 2018 Feb 7; Bethesda, MD. 052Y59

[37] Maro JC, Menzin T, Hornbuckle K, Giles JT, Kavanaugh A, Dorner T, et al. Risk of venous thromboembolism in rheumatoid arthritis patients treated with biologic and non-biologic DMARDs [abstract]. Ann Rheum Dis. 2018;77(2 suppl):932. Abstract no. SAT0140. 052WPS

[38] McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. Pharmacoepidemiol Drug Saf. 2012;21(suppl 1):18-22. 052WPT

[39] Rosati K, Jorgensen N, Soliz M, Evans BJ. Sentinel Initiative principles and policies: HIPAA and common rule compliance in the Sentinel Initiative. Silver Spring (MD): U.S. Food and Drug Administration (FDA); 2018 Feb 1. 49 p. 052WWP

[40] Sentinel [Internet]. Silver Spring (MD): Food and Drug Administration (FDA); c2010-2018. Sentinel data quality assurance practices; 2017 Mar 23 [cited 2018 Oct 29]; [about 2 screens]. Available from: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-data-quality-assurance-practices. 052XV9

[41] Toh S, Shetterly S, Powers JD, Arterburn D. Privacy-preserving analytic methods for multisite comparative effectiveness and patient-centered outcomes research. Med Care. 2014 Jul;52(7):664-8. 052WY2

[42] PopMedNet [Internet]. Boston (MA): Department of Population Medicine; PopMedNet; [cited 2018 Oct 29]. Available from: https://www.popmednet.org/. 052X99

[43] Davies M, Erickson K, Wyner Z, Malenfant J, Rosen R, Brown J. Software-enabled distributed network governance: the PopMedNet experience. EGEMS (Wash DC). 2016;4(2):1213. 052VKQ

[44] Cernea S, Buzzetti R, Pozzilli P. Beta-cell protection and therapy for latent autoimmune diabetes in adults. Diabetes Care. 2009 Nov;32(suppl 2):S246-52. 052TMV

[45] Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: a systematic review. Diabet Med. 2010;27:491-7. Erratum in: Diabet Med. 2010;27:732. 052WGZ

[46] Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. Diabetes Care. 2013 Apr;36:914-21. 052X4C

[47] Hou L, Cocoros N, Hampp C, Swain R, Maro JC. Type 1 and type 2 diabetes mellitus identification using ICD-9-CM codes within a cohort of new users of drugs labeled for type 2 diabetes mellitus. Poster session presented at: 33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE); 2017 Aug 26-30; Montreal, QC. 052VZK

[48] Sentinel [Internet]. Silver Spring (MD): Food and Drug Administration (FDA); c2010-2019. SGLT-2 inhibitor use and incidence of Diabetic Ketoacidosis in patients with Type 1 Diabetes Mellitus; 2019 Mar 19 [cited 2019 Apr 25]. Available from: https://www.sentinelinitiative.org/drugs/assessments/sglt-2-inhibitor-use-and-incidence-diabetic-ketoacidosis-patients-diabetes. 0576MY

[49] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3):399-424. 04TF35

[50] Lunt M. Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. Am J Epidemiol. 2014;179(2):226-35. 05789N

[51] Gagne JJ, Wang SV, Rassen JA, Schneeweiss S. A modular, prospective, semi-automated drug safety monitoring system for use in a distributed data environment. Pharmacoepidemiol Drug Saf. 2014;23:619-27. 052Y64

[52] Chrischilles EA, Gagne JJ, Fireman B, Nelson J, Toh S, Shoaibi A, et al. Prospective surveillance pilot of rivaroxaban safety within the US Food and Drug Administration Sentinel system. Pharmacoepidemiol Drug Saf. 2018;27:263-71. 05789G

[53] Fireman B, Toh S, Butler MG, Go AS, Joffe HV, Graham DJ, et al. A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent. Pharmacoepidemiol Drug Saf. 2012;21(S1):282-90. 05789K

[54] Toh S, Reichman ME, Graham DJ, Hampp C, Zhang R, Butler MG, et al. Prospective postmarketing surveillance of acute myocardial infarction in new users of saxagliptin: a population-based study. Diabetes Care. 2018 Jan;41:39-48. 05789P

[55] Toh S, Hampp C, Reichman ME, Graham DJ, Balakrishnan S, Pucino F, et al. Risk for hospitalized heart failure among new users of saxagliptin, sitagliptin, and other antihyperglycemic drugs: a retrospective cohort study. Ann Intern Med. 2016 Jun 7;164(11):705-14. 04Q34K

[56] Sentinel [Internet]. Silver Spring (MD): Food and Drug Administration (FDA); c2010-2018. Data quality review and characterization programs v4.1.0; 2018 Feb 18 [cited 2018 Oct 29]; [about 2 screens]. Available from: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/data-quality-review-and-characterization. 052XSG

[57] Food and Drug Administration. Guidance for industry and FDA staff: best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. Rockville, MD. May 2013. 052W62

[58] Lobo FS, Wagner S, Gross CR, Schommer JC. Addressing the issue of channeling bias in observational studies with propensity scores analysis. Res Social Adm Pharm. 2006;2:143-51. 052X4D

[59] Nordstrom BL, Fraeman KH, Lambrelli D. Post-authorization safety studies of medicinal products: the PASS book. Ali AK, Hartzema AG, editors. London (England): Academic Press; c2018. Chapter 3.1, Health insurance administrative claims; p. 49-62. 05335J

[60] International Committee of Medical Journal Editors [homepage on the Internet]. Philadelphia (PA): International Committee of Medical Journal Editors (ICMJE); Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [Updated 2017 Dec; cited 2018 Oct 31]. Available from: http://www.ICMJE.org. 052VZL

[61] von Elm E, Egger M, Altman DG, Pocock SJ, Gotzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2007 Oct 20;335:808-8. 052X2T

12 APPENDICES

Appendix 12.1 Baseline Characteristics of New Users of Ertugliflozin and New Users of Sulfonylureas/Thiazolidinediones before Propensity Score Matching (Primary Analysis)

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Demographics					
Age	mean	std	mean	std	
Sex, female	n	%	n	%	
Year	n	%	n	%	
2018	n	%	n	%	
2019	n	%	n	%	
2020	n	%	n	%	
2021	n	%	n	%	
2022	n	%	n	%	
2023	n	%	n	%	
Lab data					
HbA_{1c} (most recent)	mean	std	mean	std	
Use of antihyperglycemic agents					
Prior 0-6 month antihyperglycemic agent use					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Concomitant antihyperglycemic agent use on the index date					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Use of medications associated with DKA					
Baseline medication use					

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Concomitant medication use on the index date					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Comorbidity burden					
Combined comorbidity index	mean	std	mean	std	
Pre-existing comorbidities					
Acute illness (i.e., serious infection, surgery, trauma, acute febrile illness, or sepsis)	n	%	n	%	
Acute renal failure	n	%	n	%	
Cerebrovascular disease	n	%	n	%	
Myocardial infarction	n	%	n	%	
Stroke	n	%	n	%	
Coronary heart disease	n	%	n	%	
Heart Failure	n	%	n	%	
Hypertension	n	%	n	%	
Hyperlipidemia	n	%	n	%	
Pancreatitis	n	%	n	%	
Hypovolemia	n	%	n	%	
Hypoxemia	n	%	n	%	
Thyroid disorders	n	%	n	%	
Diabetic complications					
Moderate to severe renal insufficiency (i.e. stage 3-5 chronic kidney disease or end stage renal disease) or diabetic nephropathy	n	%	n	%	
Neuropathy	n	%	n	%	
Retinopathy	n	%	n	%	

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Peripheral vascular disease	n	%	n	%	
Amputation	n	%	n	%	
Lifestyle					
Obesity surgery	n	%	n	%	
Alcohol use	n	%	n	%	
Tobacco use	n	%	n	%	
Cocaine abuse	n	%	n	%	
Health services utilization					
Number of generic medications	mean	std	mean	std	
Number of unique pharmacological classes	mean	std	mean	std	
Number of dispensings	mean	std	mean	std	
Number of inpatient encounters	mean	std	mean	std	
Number of non-acute institutional encounters	mean	std	mean	std	
Number of emergency department encounters	mean	std	mean	std	
Number of ambulatory encounters	mean	std	mean	std	
Number of other ambulatory encounters	mean	std	mean	std	

Std: standard deviation; HbA1c: hemoglobin A1c; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2; DKA: diabetic ketoacidosis. Other ambulatory encounters include other non-overnight ambulatory encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations.

Appendix 12.2 Baseline Characteristics of New Users of Ertugliflozin and New Users of Sulfonylureas/Thiazolidinediones after Propensity Score Matching (Primary Analysis)

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Demographics					
Age	mean	std	mean	std	
Sex, female	n	%	n	%	
Year	n	%	n	%	
2018	n	%	n	%	
2019	n	%	n	%	
2020	n	%	n	%	
2021	n	%	n	%	
2022	n	%	n	%	
2023	n	%	n	%	
Lab data					
HbA_{1c} (most recent)	mean	std	mean	std	
Use of antihyperglycemic agents					
Prior 0-6 month antihyperglycemic agent use					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Concomitant antihyperglycemic agent use on the index date					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Use of medications associated with DKA					
Baseline medication use					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Concomitant medication use on the index date					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Comorbidity burden					
Combined comorbidity index	mean	std	mean	std	
Pre-existing comorbidities					
Acute illness (i.e., serious infection, surgery, trauma, acute febrile illness, or sepsis)	n	%	n	%	
Acute renal failure	n	%	n	%	
Cerebrovascular disease	n	%	n	%	
Myocardial infarction	n	%	n	%	
Stroke	n	%	n	%	
Coronary heart disease	n	%	n	%	
Heart Failure	n	%	n	%	
Hypertension	n	%	n	%	
Hyperlipidemia	n	%	n	%	
Pancreatitis	n	%	n	%	
Hypovolemia	n	%	n	%	
Hypoxemia	n	%	n	%	
Thyroid disorders	n	%	n	%	
Diabetic complications					
Moderate to severe renal insufficiency (i.e. stage 3-5 chronic kidney disease or end stage renal disease) or diabetic nephropathy	n	%	n	%	
Neuropathy	n	%	n	%	
Retinopathy	n	%	n	%	
Peripheral vascular disease	n	%	n	%	
Amputation	n	%	n	%	

	Ertugliflozin		Sulfonylurea/Thiazolidinedione		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Lifestyle					
Obesity surgery	n	%	n	%	
Alcohol use	n	%	n	%	
Tobacco use	n	%	n	%	
Cocaine abuse	n	%	n	%	
Health services utilization					
Number of generic medications	mean	std	mean	std	
Number of unique pharmacological classes	mean	std	mean	std	
Number of dispensings	mean	std	mean	std	
Number of inpatient encounters	mean	std	mean	std	
Number of non-acute institutional encounters	mean	std	mean	std	
Number of emergency department encounters	mean	std	mean	std	
Number of ambulatory encounters	mean	std	mean	std	
Number of other ambulatory encounters	mean	std	mean	std	

Std: standard deviation; HbA1c: hemoglobin A1c; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2; DKA: diabetic ketoacidosis. Other ambulatory encounters include other non-overnight ambulatory encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations.

Appendix 12.3 Baseline Characteristics of New Users of Ertugliflozin and New Users of Incretin-Based Drugs before Propensity Score Matching (Primary Analysis)

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Demographics					
Age	mean	std	mean	std	
Sex, female	n	%	n	%	
Year	n	%	n	%	
2018	n	%	n	%	
2019	n	%	n	%	
2020	n	%	n	%	
2021	n	%	n	%	
2022	n	%	n	%	
2023	n	%	n	%	
Lab data					
HbA_{1c} (most recent)	mean	std	mean	std	
Use of antihyperglycemic agents					
Prior 7-12 month antihyperglycemic agent use					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, Meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Prior 0-6 month antihyperglycemic agent use					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Concomitant antihyperglycemic agent use on the index date					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Use of medications associated with DKA					
Baseline medication use					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Concomitant medication use on the index date					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Comorbidity burden					
Combined comorbidity index	mean	std	mean	std	
Pre-existing comorbidities					
Acute illness (i.e., serious infection, surgery, trauma, acute febrile illness, or sepsis)	n	%	n	%	
Acute renal failure	n	%	n	%	
Cerebrovascular disease	n	%	n	%	
Myocardial infarction	n	%	n	%	
Stroke	n	%	n	%	
Coronary heart disease	n	%	n	%	
Heart Failure	n	%	n	%	
Hypertension	n	%	n	%	
Hyperlipidemia	n	%	n	%	
Pancreatitis	n	%	n	%	
Hypovolemia	n	%	n	%	
Hypoxemia	n	%	n	%	
Thyroid disorders	n	%	n	%	

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Diabetic complications					
Moderate to severe renal insufficiency (i.e. stage 3-5 chronic kidney disease or end stage renal disease) or diabetic nephropathy	n	%	n	%	
Neuropathy	n	%	n	%	
Retinopathy	n	%	n	%	
Peripheral vascular disease	n	%	n	%	
Amputation	n	%	n	%	
Lifestyle					
Obesity surgery	n	%	n	%	
Alcohol use	n	%	n	%	
Tobacco use	n	%	n	%	
Cocaine abuse	n	%	n	%	
Health services utilization					
Number of generic medications	mean	std	mean	std	
Number of unique pharmacological classes	mean	std	mean	std	
Number of dispensings	mean	std	mean	std	
Number of inpatient encounters	mean	std	mean	std	
Number of non-acute institutional encounters	mean	std	mean	std	
Number of emergency department encounters	mean	std	mean	std	
Number of ambulatory encounters	mean	std	mean	std	
Number of other ambulatory encounters	mean	std	mean	std	

Std: standard deviation; HbA1c: hemoglobin A1c; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2; DKA: diabetic ketoacidosis. Other ambulatory encounters include other non-overnight ambulatory encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations.

Appendix 12.4 Baseline Characteristics of New Users of Ertugliflozin and New Users of Incretin-Based Drugs after Propensity Score Matching (Primary Analysis)

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Demographics					
Age	mean	std	mean	std	
Sex, female	n	%	n	%	
Year	n	%	n	%	
2018	n	%	n	%	
2019	n	%	n	%	
2020	n	%	n	%	
2021	n	%	n	%	
2022	n	%	n	%	
2023	n	%	n	%	
Lab data					
HbA_{1c} (most recent)	mean	std	mean	std	
Use of antihyperglycemic agents					
Prior 0-6 month antihyperglycemic agent use					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Concomitant antihyperglycemic agent use on the index date					
Metformin	n	%	n	%	
SU	n	%	n	%	
TZD	n	%	n	%	
Alpha glucosidase, meglitinides	n	%	n	%	
DPP-4 inhibitor	n	%	n	%	
GLP-1 receptor agonist	n	%	n	%	
SGLT2 inhibitor	n	%	n	%	
Insulin	n	%	n	%	
Use of medications associated with DKA					
Baseline medication use					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Concomitant medication use on the index date					
Clozapine or olanzapine	n	%	n	%	
Lithium	n	%	n	%	
Terbutaline	n	%	n	%	
Oral corticosteroids	n	%	n	%	
Thiazides	n	%	n	%	
Pentamidine	n	%	n	%	
Comorbidity burden					
Combined comorbidity index	mean	std	mean	std	
Pre-existing comorbidities					
Acute illness (i.e., serious infection, surgery, trauma, acute febrile illness, or sepsis)	n	%	n	%	
Acute renal failure	n	%	n	%	
Cerebrovascular disease	n	%	n	%	
Myocardial infarction	n	%	n	%	
Stroke	n	%	n	%	
Coronary heart disease	n	%	n	%	
Heart Failure	n	%	n	%	
Hypertension	n	%	n	%	
Hyperlipidemia	n	%	n	%	
Pancreatitis	n	%	n	%	
Hypovolemia	n	%	n	%	
Hypoxemia	n	%	n	%	
Thyroid disorders	n	%	n	%	
Diabetic complications					
Moderate to severe renal insufficiency (i.e. stage 3-5 chronic kidney disease or end stage renal disease) or diabetic nephropathy	n	%	n	%	
Neuropathy	n	%	n	%	
Retinopathy	n	%	n	%	
Peripheral vascular disease	n	%	n	%	
Amputation	n	%	n	%	

	Ertugliflozin		Incretin-based drugs		Standardized Difference
	N/mean	%/std	N/mean	%/std	
Lifestyle					
Obesity surgery	n	%	n	%	
Alcohol use	n	%	n	%	
Tobacco use	n	%	n	%	
Cocaine abuse	n	%	n	%	
Health services utilization					
Number of generic medications	mean	std	mean	std	
Number of unique pharmacological classes	mean	std	mean	std	
Number of dispensings	mean	std	mean	std	
Number of inpatient encounters	mean	std	mean	std	
Number of non-acute institutional encounters	mean	std	mean	std	
Number of emergency department encounters	mean	std	mean	std	
Number of ambulatory encounters	mean	std	mean	std	
Number of other ambulatory encounters	mean	std	mean	std	

Std: standard deviation; HbA1c: hemoglobin A1c; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2; DKA: diabetic ketoacidosis. Other ambulatory encounters include other non-overnight ambulatory encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations.

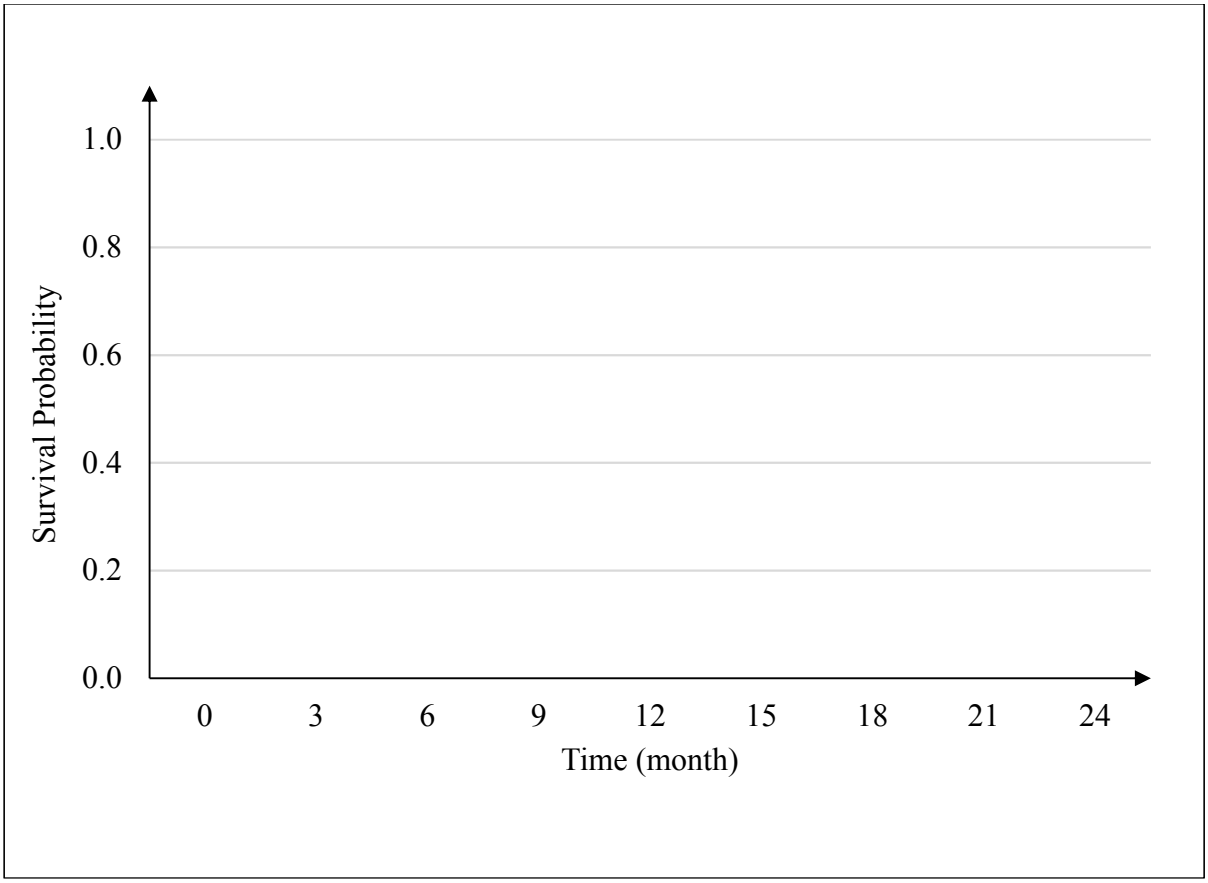
Appendix 12.5 Descriptive data of the diabetic ketoacidosis (DKA) cases, identified by principal discharge diagnosis, among new users of ertugliflozin, Sulfonylurea/Thiazolidinedione (SU/TZD) and incretin-based drugs

	New User of Ertugliflozin versus SU/TZD		New User of Ertugliflozin versus Incretin-Based Drugs	
	Ertugliflozin	SU/TZD	Ertugliflozin	Incretin-based drugs
Total, N				
Male, n(%)				
Age, Mean (SD)				
Insulin use, n (%)				
Number of days between index date and DKA onset, mean (range)				

Appendix 12.6 Descriptive data of the diabetic ketoacidosis (DKA) cases, identified by first- or second-listed inpatient diagnosis, among new users of ertugliflozin, Sulfonylurea/Thiazolidinedione (SU/TZD) and incretin-based drugs

	New User of Ertugliflozin versus SU/TZD		New User of Ertugliflozin versus Incretin-Based Drugs	
	Ertugliflozin	SU/TZD	Ertugliflozin	Incretin-based drugs
Total, N				
Male, n(%)				
Age, Mean (SD)				
Insulin use, n (%)				
Number of days between index date and DKA onset, mean (range)				

Appendix 12.7 Kaplan-Meier Curves Comparing New User of Ertugliflozin versus Sulfonylureas/Thiazolidinediones or versus Incretin-Based Drugs from Time to Diabetic Ketoacidosis (Primary Analysis)



Appendix 12.8 Risk Estimates for Diabetic Ketoacidosis among New Users of Ertugliflozin and New Users of Sulfonylureas/Thiazolidinediones (SU/TZD) -Primary and Subgroup Analysis

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Primary Analysis								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Subgroup Analysis (with Concomitant Insulin Use at Baseline)								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Subgroup Analysis (without Concomitant Insulin Use at Baseline)								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								

Appendix 12.9 Risk Estimates for Diabetic Ketoacidosis among New Users of Ertugliflozin and New Users of Incretin-Based Drugs -Primary and Subgroup Analysis

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Primary Analysis								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								
Subgroup Analysis (with Concomitant Insulin Use at Baseline)								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								
Subgroup Analysis (without Concomitant Insulin Use at Baseline)								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								

Appendix 12.10 Risk Estimates for Diabetic Ketoacidosis - Sensitivity Analysis

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
New user of ertugliflozin versus new user of sulfonylureas/thiazolidinediones (SU/TZD)								
Sensitivity Analysis 1								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Sensitivity Analysis 2								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Sensitivity Analysis 3								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Sulfonylurea/TZD								
Sensitivity Analysis 4								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Sensitivity Analysis 5								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
Sensitivity Analysis 6								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Sensitivity Analysis 7								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Sulfonylurea/TZD								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Sulfonylurea/TZD								
New user of ertugliflozin versus new user of Incretin-Based Drugs								
Sensitivity Analysis 1								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								
Sensitivity Analysis 2								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Sensitivity Analysis 3								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
incretin-based drugs								
Incretin-based drugs								
Sensitivity Analysis 4								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								
Sensitivity Analysis 5								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								

Medical Product	Number of New Users	Number of New Use Episodes	Average Person Years at Risk	Number of Events	Incidence Rate per 1,000 Person Years	Incidence Rate Difference per 1,000 Person Years	Hazard Ratio (95% CI)	Wald p-Value
Sensitivity Analysis 6								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								
Sensitivity Analysis 7								
Unmatched Analysis (site-adjusted only)								
Ertugliflozin								
Incretin-based drugs								
1:1 Propensity Score-Matched Analysis								
Ertugliflozin								
Incretin-based drugs								

Annex 1 List of Stand-Alone Documents

No.	Document Reference No	Date	Title
1.	PV-GLB-01-ER04, V2.0	11/6/2018	Adverse Event Reporting Form

Annex 2 ENCePP Checklist for Study Protocols (Revision 4)

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

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

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

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

Study title: Post-authorization safety study to assess the risk of diabetic ketoacidosis among type 2 diabetes mellitus patients treated with ertugliflozin compared to patients treated with other antihyperglycemic agents

EU PAS Register® number: Not yet registered
Study reference number (if applicable):

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	4
1.1.2 End of data collection[2]	☒	☐	☐	4
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☒	☐	☐	4
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	4
1.1.6 Final report of study results.	☒	☐	☐	4

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

[2] Date from which the analytical dataset is completely available.

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	6
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)	☒	☐	☐	5
2.1.2 The objective(s) of the study?	☒	☐	☐	6
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	6, 7.2
2.1.4 Which hypothesis(-es) is (are) to be tested?	☒	☐	☐	6
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	7.1
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	7.1, 7.2, 7.4.1
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	7.7.2
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☒	☐	☐	7.7.3
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)	☒	☐	☐	9.1

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	7.2, 7.4.1
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	7.2
4.2.2 Age and sex	☒	☐	☐	7.2
4.2.3 Country of origin	☒	☐	☐	7.2
4.2.4 Disease/indication	☒	☐	☐	7.2
4.2.5 Duration of follow-up	☒	☐	☐	7.7.1
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	7.2

Comments:

Only adult patients will be included (18+ years), and both females and males will be studied. The databases to be used are from health insurers in the US, though it is possible that some patients may reside outside of the US.

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)	☒	☐	☐	7.3.1
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	7.3.1, 7.7.3.3
5.3 Is exposure categorised according to time windows?	☒	☐	☐	7.3.1, 7.7.3.3
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	7.7.1, 7.7.2
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	7.3.1

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	7.3.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	7.3.2
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☒	☐	☐	7.3.2, 7.4.3.3.2
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

The primary outcome in this study is hospitalization for DKA.

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☒	☐	☐	7.1, 7.3.3, 7.7.3.3
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	7.9
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	7.4.3.2, 7.4.3.3, 7.9

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☒	☐	☐	7.7.3.1

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)	☒	☐	☐	7.4.1
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	7.4.1
9.1.3 Covariates and other characteristics?	☒	☐	☐	7.4.1
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)	☒	☐	☐	7.4.1
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	7.4.1
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	7.4.1
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	7. 1
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	7. 1
9.3.3 Covariates and other characteristics?	☒	☐	☐	7. 1
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	7.7
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	7.5
10.3 Are descriptive analyses included?	☒	☐	☐	7.7.2
10.4 Are stratified analyses included?	☒	☐	☐	7.7.3.1
10.5 Does the plan describe methods for analytic control of confounding?	☒	☐	☐	7.1, 7.3.3, 7.7.3.3
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☒	☐	☐	7.4.3.3.2
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	7.7.3.3

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	7.3.1, 7.7.1, 7.7.3.2

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)	☒	☐	☐	7.6
11.2 Are methods of quality assurance described?	☒	☐	☐	7.8
11.3 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

Though there is not a system in place for independent review of the study results, study deliverables including data tables and summaries of results (study reports) will be reviewed by all stakeholders including the IMEDS, the participating data partners, and MSD staff (including persons not directly involved in the study). Additionally, we will consult with independent experts for their insight on the study findings as appropriate. Finally, the study will be submitted to the EMA for their review and approval.

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?	☒	☐	☐	7.9
12.1.2 Information bias?	☒	☐	☐	7.9
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☒	☐	☐	7.9
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	7.4.3

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	8
13.2 Has any outcome of an ethical review procedure been addressed?	☒	☐	☐	8
13.3 Have data protection requirements been described?	☒	☐	☐	8

Comments:

This study will employ a distributed analytic approach that uses person-level health information to generate aggregate findings that are combined across data sources. The study protocol will be reviewed by an Institutional Review Board, but studies of this type are typically determined to be exempt from Institutional Review Board review upon submission.

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	3

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)?	☒	☐	☐	10
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	10

Comments:

Name of the main author of the protocol: Jeffrey S. Brown, PhD

Date: August 6, 2019

Signature: PPD

Annex 3 Administrative and Regulatory Details

Confidentiality:

Confidentiality of Data

By signing this protocol, the investigator affirms to the Sponsor that information furnished to the investigator by the Sponsor will be maintained in confidence, and such information will be divulged to the Institutional Review Board, Ethics Review Committee or similar or expert committee; affiliated institution and employees, only under an appropriate understanding of confidentiality with such board or committee, affiliated institution and employees. Data generated by this study will be considered confidential by the investigator, except to the extent that it is included in a publication as provided in the Publications section of this protocol.

Confidentiality of Subject Records

By signing this protocol, the investigator agrees that the Sponsor (or Sponsor representative), Institutional Review Board/Independent Ethics Committee (IRB/IEC), or Regulatory Agency representatives may consult and/or copy study documents in order to verify worksheet/case report form data. By signing the consent form, the subject agrees to this process. If study documents will be photocopied during the process of verifying worksheet/case report form information, the subject will be identified by unique code only; full names/initials will be masked prior to transmission to the Sponsor.

By signing this protocol, the investigator agrees to treat all subject data used and disclosed in connection with this study in accordance with all applicable privacy laws, rules and regulations.

Confidentiality of Investigator Information

By signing this protocol, the investigator recognizes that certain personal identifying information with respect to the investigator, and all subinvestigators and study site personnel, may be used and disclosed for study management purposes, as part of a regulatory submissions, and as required by law. This information may include:

- name, address, telephone number and e-mail address;
- hospital or clinic address and telephone number;
- curriculum vitae or other summary of qualifications and credentials; and
- other professional documentation.

Consistent with the purposes described above, this information may be transmitted to the Sponsor, and subsidiaries, affiliates and agents of the Sponsor, in your country and other countries, including countries that do not have laws protecting such information. Additionally, the investigator's name and business contact information may be included when reporting certain serious adverse events to regulatory agencies or to other investigators. By signing this protocol, the investigator expressly consents to these uses and disclosures.

If this is a multicenter study, in order to facilitate contact between investigators, the Sponsor may share an investigator's name and contact information with other participating investigators upon request.

Administrative:

Compliance with Financial Disclosure Requirements

Financial Disclosure requirements are outlined in the US Food and Drug Administration Regulations, Financial Disclosure by Clinical Investigators (21 CFR Part 54). It is the Sponsor's responsibility to determine, based on these regulations, whether a request for Financial Disclosure information is required. It is the investigator's/subinvestigator's responsibility to comply with any such request.

The investigator/subinvestigator(s) agree, if requested by the Sponsor in accordance with 21 CFR Part 54, to provide his/her financial interests in and/or arrangements with the Sponsor to allow for the submission of complete and accurate certification and disclosure statements. The investigator/subinvestigator(s) further agree to provide this information on a Certification/Disclosure Form, commonly known as a financial disclosure form, provided by the Sponsor or through a secure password-protected electronic portal provided by the Sponsor. The investigator/subinvestigator(s) also consent to the transmission of this information to the Sponsor in the United States for these purposes. This may involve the transmission of information to countries that do not have laws protecting personal data.

Compliance with Law, Audit and Debarment

By signing this protocol, the investigator agrees to conduct the study in an efficient and diligent manner and in conformance with this protocol; generally accepted standards of Good Pharmacoepidemiology Practice and all applicable federal, state and local laws, rules and regulations relating to the conduct of the study.

The investigator also agrees to allow monitoring, audits, Institutional Review Board/Independent Ethics Committee review and regulatory agency inspection of study-related documents and procedures and provide for direct access to all study-related source data and documents.

The investigator agrees not to seek reimbursement from subjects, their insurance providers or from government programs for procedures included as part of the study reimbursed to the investigator by the Sponsor.

The Investigator shall prepare and maintain complete and accurate study documentation in compliance with Good Pharmacoepidemiology Practice, standards and applicable federal, state and local laws, rules and regulations; and, for each subject participating in the study, provide all data, and, upon completion or termination of the study, submit any other reports to the Sponsor as required by this protocol or as otherwise required pursuant to any agreement with the Sponsor.

Study documentation will be promptly and fully disclosed to the Sponsor by the investigator upon request and also shall be made available at the investigator's site upon request for inspection, copying, review and audit at reasonable times by representatives of the Sponsor or any regulatory agencies. The investigator agrees to promptly take any reasonable steps that are requested by the Sponsor as a result of an audit to cure deficiencies in the study documentation and worksheets/case report forms.

The investigator must maintain copies of all documentation and records relating to the conduct of the study in accordance with their institution's records retention schedule which is compliant with all applicable regional and national laws and regulatory requirements. If an institution does not have a records retention schedule to manage its records long-term, the investigator must maintain all documentation and records relating to the conduct of the study for 5 years after final report or first publication of study results, whichever comes later, per GPP guidelines. This documentation includes, but is not limited to, the protocol, worksheets/case report forms, advertising for subject participation, adverse event reports, subject source data, correspondence with regulatory authorities and IRBs/ERCs, consent forms, investigator's curricula vitae, monitor visit logs, laboratory reference ranges, laboratory certification or quality control procedures and laboratory director curriculum vitae. All study documents shall be made available if required by relevant regulatory authorities. The investigator must consult with the Sponsor prior to discarding study and/or subject files.

The investigator will promptly inform the Sponsor of any regulatory agency inspection conducted for this study.

Persons debarred from conducting or working on studies by any court or regulatory agency will not be allowed to conduct or work on this Sponsor's studies. The investigator will immediately disclose in writing to the Sponsor if any person who is involved in conducting the study is debarred or if any proceeding for debarment is pending or, to the best of the investigator's knowledge, threatened.

In the event the Sponsor prematurely terminates a particular study site, the Sponsor will promptly notify that site's IRB/IEC.

According to European legislation, a Sponsor must designate an overall coordinating investigator for a multi-center study (including multinational). When more than one study site is open in an EU country, MSD, as the Sponsor, will designate, per country, a national principal coordinator (Protocol CI), responsible for coordinating the work of the principal investigators at the different sites in that Member State, according to national regulations. For a single-center study, the Protocol CI is the principal investigator. In addition, the Sponsor must designate a principal or coordinating investigator to review the study report that summarizes the study results and confirm that, to the best of his/her knowledge, the report accurately describes the conduct and results of the study in the study's final report. The Sponsor may consider one or more factors in the selection of the individual to serve as the Protocol CI and or CSR CI (e.g., availability of the CI during the anticipated review process, thorough understanding of study methods, appropriate enrollment of subject cohort, timely achievement of study milestones). The Protocol CI must be a participating study investigator.

Compliance with Study Registration and Results Posting Requirements

Under the terms of the Food and Drug Administration Modernization Act (FDAMA) and the Food and Drug Administration Amendments Act (FDAAA), the Sponsor of the study is solely responsible for determining whether the study and its results are subject to the requirements for submission to the Clinical Trials Data Bank, such as ENCePP. MSD, as Sponsor of this study, will review this protocol and submit the information necessary to fulfill these requirements. MSD entries are not limited to FDAMA/FDAAA mandated studies. Information posted will allow subjects to identify potentially appropriate studies for their disease conditions and pursue participation by calling a central contact number for further information on appropriate study locations and site contact information.

By signing this protocol, the investigator acknowledges that the statutory obligations under FDAMA/FDAAA are that of the Sponsor and agrees not to submit any information about this study or its results to the Clinical Trials Data Bank.

Annex 4 PRAC Endorsement

Placeholder for PRAC endorsement letter.

Not applicable.

Annex 5 Qualified Person for PharmacoVigilance (QPPV)

PPD
European Union Qualified Person for Risk Management and Pharmacovigilance
Office of the European Union Qualified Person for Pharmacovigilance (EU QPPV)
Merck Sharp & Dohme (Europe), Inc.
Siège d'exploitation : 5, Clos du Lynx 1200 Bruxelles
Exploitatiezetel : Lynx Binnenhof, 5 1200 Brussel
Tel: PPD
Email: PPD

Emergency/Out of Hours: GSM numbers above or via PPD

Dear Sir/Madam

Re: EU QPPV Signature Page for PASS

INN: Ertugliflozin L-pyroglutamic acid

Ertugliflozin L-pyroglutamic acid/ metformin hydrochloride

Ertugliflozin L-pyroglutamic acid/ sitagliptin phosphate monohydrate

Product: STEGLATRO™ (ertugliflozin)

SEGLUROMET™ (ertugliflozin/metformin)

STEGLUJAN™ (ertugliflozin/sitagliptin)

Protocol No.: 8835-062

Epidemiology No.: EP02039.002

Protocol Date: August 6, 2019

MAH: Merck Sharp & Dohme Ltd.

In line with the Guideline on Good PharmacoVigilance Practice (GVP), Module VIII – Post-Authorisation Safety Studies (PASS) and according to MSD internal SOPs, this study has been reviewed and approved by the European Qualified Person for Pharmacovigilance.

Yours faithfully

PPD

Associate Vice President,
EU Qualified Person for Pharmacovigilance