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: Interim Report 1
Version identifier of the interim study report	002
Date of last version of the final study report	Not applicable
EU PAS register number	EUPAS31718
Active substance	Ertugliflozin Ertugliflozin + Metformin Hydrochloride Ertugliflozin + Sitagliptin
Medicinal product	STEGLATRO™ (ertugliflozin; ATC code A10BK04) SEGLUROMET™ (ertugliflozin/metformin; ATC code A10BD23) STEGLUJAN™ (ertugliflozin/sitagliptin; ATC code A10BD24)
Product reference	EMEA/H/C/004315 - Steglatro™ EMEA/H/C/004314 - Segluromet™ EMEA/H/C/004313 - Steglujan™
Procedure number	EMEA/H/C/004315/MEA/002.3 - Steglatro™ EMEA/H/C/004314/MEA/002.3 - Segluromet™ EMEA/H/C/004313/MEA/002.3 - Steglujan™
Marketing authorisation holder(s)	Merck Sharp & Dohme B.V. Waarderweg 39 2031 BN Haarlem The Netherlands
Joint PASS	No

Research question and objectives	Preliminary analyses on the study population included as of 28 Feb 2021 were conducted to: 1. assess the number of new users of ertugliflozin and comparator antihyperglycemic agents (AHAs, including sulfonylureas, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, and glucagon-like peptide-1 receptor agonists) 2. describe the baseline patient characteristics of new users of ertugliflozin and comparator AHAs 3. discuss the sample size and explore data solutions if findings from Objective 1 indicate sample size, particularly the number of new users of ertugliflozin accumulated in the Innovation in Medical Evidence and Development Surveillance Distributed Database (IMEDS-DD), is not anticipated to reach the target number to perform the final study (Protocol MK8835-062; EUPAS31718).
Country(-ies) of study	United States
Author	PPD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD PPD PPD PPD PPD PPD PPD PPD PPD PPD
Merck Final Repository (RCAM) Date	02-DEC-2022

MARKETING AUTHORISATION HOLDER(S)

Marketing authorisation holder(s)	Merck Sharp & Dohme B.V. Waarderweg 39 2031 BN Haarlem The Netherlands
MAH contact person	PPD Associate Principal Scientist, Epidemiology Biostatistics and Research Decision Sciences Merck Sharp & Dohme LLC PPD PPD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ANNEXES

1 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: Interim Report 1

Keywords

ertugliflozin, antihyperglycemic agents, IMEDS-DD, sample size, type 2 diabetes mellitus

Rationale and background

MSD has committed to the European Medicines Agency (EMA) to conduct *the Post- Authorization Safety Study (PASS) 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 MK8835-062; EUPAS31718) to investigate the association of ertugliflozin use with diabetic ketoacidosis (DKA) among type 2 diabetes mellitus (T2DM) patients using the Innovation in Medical Evidence Development and Surveillance System Distributed Database (IMEDS-DD). As part of the study milestones, two interim reports will be submitted (4Q 2021 and 4Q 2022). As requested in the PRAC PASS Protocol Assessment Report dated 05 September 2019 (Section 11), these interim reports should provide the preliminary analyses on the study population available at their respective reporting time, discuss the sample size, and consider inclusion of other database(s) if the sample size is not anticipated to reach the target number needed for the final analyses. This is the Interim Report 1.

Objectives of Interim Report 1

Preliminary analyses on the study population included as of 28 Feb 2021 were conducted to

1. assess the number of new users of ertugliflozin and comparator antihyperglycemic agents (AHAs)
2. describe the baseline patient characteristics of new users of ertugliflozin and comparator AHAs
3. discuss the sample size and explore data solutions if findings from Objective 1 indicate sample size, particularly the number of new users of ertugliflozin accumulated in the IMEDS-DD, is not anticipated to reach the target number to perform the final study

Study design

A non-interventional cohort study using electronic healthcare data.

Setting

The preliminary analysis was first conducted using the Reagan-Udall Foundation for the Food and Drug Administration (FDA)'s IMEDS-DD, a subset of the FDA Sentinel Distributed Database. The same analysis was then replicated in the IBM® MarketScan® databases to demonstrate the advantageous ability of a distributed data network such as the IMEDS-DD in terms of sample size expansion and deployment of analytics across the included databases.

Subjects and study size, including dropouts

Three primary exposure cohorts of interest were defined as new users of: (1) ertugliflozin; (2) sulfonylureas (SU) or thiazolidinediones (TZD); and (3) incretin- based drugs. Each exposure cohort was defined as the set of patients who: had a first qualifying exposure for the cohort-defining drug(s), with the first exposure date (index date) occurring between 1 July 2018 and most recent available in each database; had at least 1 diagnosis of T2DM without any diagnosis of type 1 diabetes mellitus (T1DM) or gestational diabetes on or any time before the index date; and had no evidence of dispensing of the defining drug in the required 6 months of prior continuous observation.

Variables and data sources

Variables: baseline demographics (age and sex), comorbidity burden (Charlson-Elixhauser combined comorbidity index), pre-existing comorbidities and diabetes-related complications (acute illness, surgical procedures, acute renal failure, cerebrovascular disease, coronary heart disease, heart failure, hyperlipidemia, hypertension, hypoglycemia, hypovolemia, hypoxemia, myocardial infarction, obesity, pancreatitis, peripheral artery disease, stroke, thyroid disorders, moderate or severe renal insufficiency, diabetic nephropathy, neuropathy, retinopathy, and amputation), AHA utilization by class, and health services utilization (number of unique medications, number of inpatient, emergency department, ambulatory department visits).

Data sources: The preliminary analysis utilized electronic health insurance claims data from the Reagan-Udall Foundation for the FDA's IMEDS-DD, a subset of the FDA Sentinel Distributed Database. The Sentinel Distributed Database is a national electronic system for active surveillance of medical product safety in the United States. In addition, the same analysis was then replicated using electronic health insurance claims data from the IBM® MarketScan® Commercial Claims and Encounters Database (CCAE) linked with the IBM® MarketScan® Medicare Supplemental Beneficiaries (MDCR), to demonstrate the feasibility of a proposed data solution to address Objective 3. It is expected that the patient populations in both databases are largely representative of the commercially insured population in US.

Results

A total of 647 patients who initiated ertugliflozin were identified in the IMEDS-DD between 1 July 2018 and 28 February 2021. Among those, 43.0% were female and the mean age was 57.2 years (SD=11.0). Of the 647 patients, 103 (15.9%) patients initiated ertugliflozin as monotherapy. The most concomitantly utilized AHA class at the index date was metformin (61.8%). The most common comorbidities included hyperlipidemia (72.8%), hypertension (69.7%), and obesity (36.2%). Individuals with a history of cardiovascular disease (CVD) represented 19.2% of new users of ertugliflozin.

Replicating the same study design by utilizing the same Sentinel Common Data Model and analytic programming, a total of 2,004 patients who initiated ertugliflozin were identified in the IBM® MarketScan® databases between 1 July 2018 and 30 June 2020. Among these, mean age was 53.3 years (SD=9.2), 43.8% were female, and 12.9% initiated ertugliflozin as monotherapy. The most concomitantly utilized AHA class at the index date was metformin (68.6%), and the most prevalent comorbidities included hyperlipidemia (67.7%), hypertension (65.0%), and obesity (33.7%). History of CVD was reported in 9.6% of new users of ertugliflozin.

During the same study periods, there were 257,316 new users of SU/TZD (161,941 in IMEDS-DD and 95,375 in IBM® MarketScan® databases) and 283,546 new users of incretin-based drugs (157,203 in IMEDS-DD and 126,343 in IBM® MarketScan® databases). The demographic and clinical characteristics were comparable to what have been reported as comparison groups in observational studies of other SGLT2 inhibitors.

Discussion

The preliminary analyses indicate that study sample size, particularly the number of new users of ertugliflozin accumulated in the IMEDS-DD, is not anticipated to reach the target number by the milestone date for the final report, if data are limited to the current 5 regional and national health insurers network partners included in IMEDS-DD. Recognizing this possible issue, the first interim report pursued exploration of data solutions. The IMEDS framework employs the Sentinel Common Data Model for standardization of demographic and clinical data elements from various network partners. By doing so, the IMEDS-DD enables rapid queries across the included databases due to the analytic tools established and actively maintained by the Sentinel System. Successful replication of the planned analysis in the IBM® MarketScan® databases demonstrated feasibility of sample size expansion via this approach. The Applicant will consider including data from additional sources in the data network, and the results will be summarized in the Interim Report 2 that will be submitted to EMA in 2022 as planned. Using the same methods for replication in the IBM® MarketScan® databases, the additional data sources being considered are: 1) Optum© Database, 2) the Centers for Medicare & Medicaid Services (CMS) Medicare Fee-for-Service Research Identifiable Files, and/or 3) multi-state Medicaid database. In this way the sample size will be increased significantly for the next interim report and the final analysis.

Despite the insufficient cohort size of new users of ertugliflozin projected for the final analysis by this preliminary analysis, the baseline characteristics of these identified patients are largely similar to those of new users of other SGLT2 inhibitors reported in large-scale real-world observational studies. These other studies include: EMPagliflozin compaRative effectIveness and SafEty (EMPRISE; EUPAS20677); Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases (OBSERVE-4D); Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors (CVD-REAL); Diabetic Ketoacidosis in Patients with Type 2 Diabetes Treated with Sodium Glucose Co- Transporter 2 Inhibitors versus Other Antihyperglycemic Agents: An Observational Study of Four US Administrative Claims Databases (EUPAS23705); and two Canadian Network for Observational Drug Effect Studies (CNODES) studies assessing the association of SGLT2 inhibitor exposure and various outcomes.

Marketing Authorisation Holder(s)

Merck Sharp & Dohme B.V.
Waarderweg 39
2031 BN Haarlem
The Netherlands.

Names and affiliations of principal investigators

PPD
Department of Population Medicine
Harvard Medical School & Harvard Pilgrim Health Care
Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts,
USA 02215
PPD
PPD

PPD
Department of Population Medicine
Harvard Medical School & Harvard Pilgrim Health Care
Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts,
USA 02215
PPD
PPD

[1] PPD served as the Principal Invesigator (PI) from July 2018 to October 2021, when he left his position at the Department of Population Medicine of Harvard Medical School.

[2] PPD, who has been the co-investigator of the project since July 2018, replaces PPD as the PI starting November 2021.

2 LIST OF ABBREVIATIONS

AHA	Antihyperglycemic agents
CCAE	IBM® MarketScan® Commercial Claims and Encounters Database
CDC	Centers for Disease Control and Prevention
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
CVD	Cardiovascular disease
CVD-REAL	Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors
CNODES	Canadian Network for Observational Drug Effect Studies
DKA	Diabetic ketoacidosis
DPP-4	Dipeptidyl peptidase 4
EHR	Electronic health record
EMA	European Medicines Agency
EMPRISE	EMPagliflozin compaRative effectIveness and SafEty
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
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
GLP-1	Glucagon-like peptide-1
HbA_{1c}	Hemoglobin A_{1c}
HCPCS	Healthcare Common Procedure Coding System
HPHCI	Harvard Pilgrim Health Care Institute
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
IDF	International Diabetes Federation
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
MDCR	IBM® MarketScan® Medicare Supplemental Beneficiaries
MSD	Merck Sharp & Dohme, Corp.
NDC	National Drug Codes
PASS	Post-Authorization Safety Study
PCORnet	National Patient-Centered Clinical Research Network
QA	Quality assurance
QC	Quality control

(EU GUIDANCE: 23 JANUARY 2013 EMA/738724/2012)

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

3 INVESTIGATORS

Principal investigator	PPD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD PPD PPD Department of Population Medicine Harvard Medical School & Harvard Pilgrim Health Care Institute 401 Park Drive, Suite 401 East, Boston, Massachusetts, USA 02215 PPD PPD
Coordinating investigator for each country in which the study is to be performed	Not applicable
Sponsor contacts	PPD Associate Principal Scientist, Epidemiology Biostatistics and Research Decision Sciences Merck Sharp & Dohme LLC PPD PPD
Other contacts	Not applicable
Vendor/Collaborator	Innovation in Medical Evidence and Development Surveillance
Investigators	PPD PPD PPD PPD PPD [1] Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA [2] Reagan-Udall Foundation for the Food and Drug Administration, Washington DC, USA

[3] PPD served as the Principal Invesigator (PI) from July 2018 to October 2021, when he left his position at the Department of Population Medicine of Harvard Medical School.

[4] PPD who has been the co-investigator of the project since July 2018, replaces PPD as the PI starting November 2021.

(EU GUIDANCE: 23 JANUARY 2013 EMA/738724/2012)

4 OTHER RESPONSIBLE PARTIES

Not applicable.

5 MILESTONES OF MK8835-062

Milestone	Planned date	Actual date
Start of data collection of MK8835-062	1 July 2018	1 July 2018
End of data collection of MK8835-062	30 June 2023	
Registration in the EU PAS register	24 October 2019	17 October 2019
Feasibility assessment report	31 December 2020	7 December 2020
Interim report 1	31 December 2021 (version 001) 31 December 2022 (version 002)	9 December 2021 (version 001)
Interim report 2	31 December 2022	
Final report of study results	31 December 2023	

6 RATIONALE AND BACKGROUND

Ertugliflozin is a sodium-glucose co-transporter 2 (SGLT2) inhibitor. Ertugliflozin products (including ertugliflozin, ertugliflozin/sitagliptin and ertugliflozin/metformin hydrochloride) were approved in the United States (US) on 19 December 2017 and approved in Europe on 21 March 2018 for type 2 diabetes mellitus (T2DM) treatment to improve glycemic control in adults. As its marketing authorization holder, Merck Sharp & Dohme, Corp (MSD) has committed to the European Medicines Agency (EMA) to conduct the *Post-Authorization Safety Study (PASS) to Assess the Risk of Diabetic Ketoacidosis among Type 2 Diabetes Mellitus Patients Treated with Ertugliflozin Compared to Patients Treated with Other Antihyperglycemic Agents* (MK8835-062; EUPAS31718) to investigate the association of ertugliflozin use with diabetic ketoacidosis (DKA) among T2DM patients and more specifically, 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 using the Innovation in Medical Evidence Development and Surveillance System (IMEDS) Distributed Database (IMEDS-DD).

The IMEDS-DD is a subset of the Food and Drug Administration (FDA) Sentinel Distributed Database. IMEDS is a public-private partnership launched in 2017 by the Reagan-Udall Foundation for the FDA, 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 [Ref. 5.4: 052TMC, 052WPW].

As requested in the PRAC PASS Protocol Assessment Report dated 05 September 2019 (Section 11), the present report is the first of the two interim reports to provide preliminary analyses on the study population available at the time of report submission and to assess the availability of ertugliflozin exposure in the IMEDS-DD in order to conduct the final analysis planned for MK8835-062.

7 OBJECTIVES OF INTERIM REPORT 1

Preliminary analyses on the study population included as of 28 Feb 2021 were conducted to:

1. assess the number of new users of ertugliflozin and comparator AHAs
2. describe the baseline patient characteristics of new users of ertugliflozin and comparator AHAs
3. discuss the sample size and explore data solutions if findings from Objective 1 indicate sample size, particularly the number of new users of ertugliflozin accumulated in the IMEDS-DD, is not anticipated to reach the target number to perform the final study

8 AMENDMENTS AND UPDATES

Interim Report 1 was revised to correct an error in the originally submitted report. In the Original Interim Report 1, one IMEDS-DD Network Partner executed the study queries on the wrong study population (i.e., the entire population including both the research-eligible populations and Administrative Services Only members, instead of research-eligible population), leading to an overestimation of the sample size. This updated report contains the latest updated analytic results reflective of the research-eligible populations within the IMEDS-DD. The overall conclusion remains the same namely, the targeted sample size will not be reached by the milestone date for the final report if data are limited to those from the IMEDS-DD; additional data sources are needed.

9 RESEARCH METHODS

9.1 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 AHA drug classes include sulfonylureas (SU), thiazolidinediones (TZD), dipeptidyl peptidase-4 (DPP-4) inhibitors, and glucagon-like peptide-1 (GLP-1) receptor agonists (hereinafter collectively referred to as “comparator AHAs”). These drug classes were 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 [Ref. 5.4: 04YCLB, 052W8Y].

In the first interim report, three primary exposure cohorts of interest were defined as new users of: (1) ertugliflozin; (2) SU or TZD; and (3) incretin-based drugs. Each exposure cohort was defined as the set of patients who had a first qualifying exposure for the cohort-defining drug(s), with the first exposure date (index date) occurring between 1 July 2018 and 28 February 2021 (or most recent available).

Unless otherwise specified, this study used outpatient pharmacy claims to define drug utilization and medical encounter claims to define existing conditions, medical history, or outcomes. National Drug Code (NDC) was used to identify individual medications. Diagnosis and procedure codes encoded in the following coding systems were 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.

9.2 Setting

The study utilized data from the Reagan-Udall Foundation for the FDA's 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 [Ref. 5.4: 052TMC, 052WPW]. The IMEDS-DD and Sentinel Distributed Database employ the Sentinel Common Data Model for standardization of demographic and clinical data elements from various network partners [Ref. 5.4: 052TNG, 052Y0K]. This study used data from five regional and national health insurers of the IMEDS-DD in US. Health plan members enrolled with these insurers are predominately commercially insured and community dwelling individuals. 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 or drug plans through IMEDS network partners may also be included.

Additionally, as part of the exploration, this study used existing administrative claims in the IBM® MarketScan® Commercial Claims and Encounters Database (CCAE) linked with the IBM® MarketScan® Medicare Supplemental Beneficiaries (MDCR) to demonstrate the feasibility of including an additional data source to the IMEDS-DD to help reach the target sample size for the final analysis.

9.3 Subjects

This study included eligible patients who met the following inclusion and exclusion criteria in the analysis:

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 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 or 250.x2 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 *O24.2x 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 discontinued before the index date or ongoing insulin use were allowed)
- History of DKA, evidenced by 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*.

9.4 Variables

9.4.1 Exposure

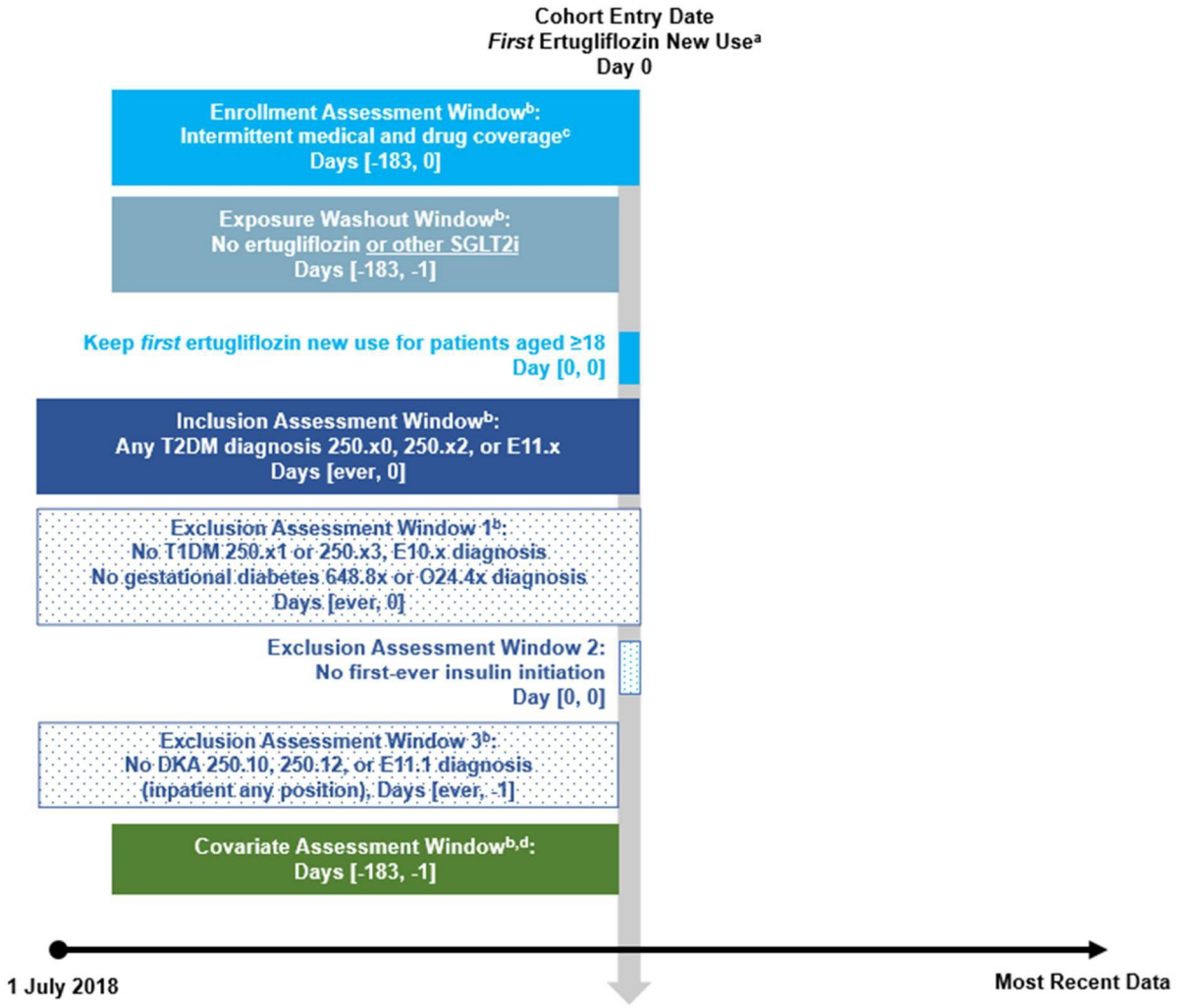
The study population was classified into three new user groups based on exposure: 1) ertugliflozin, 2) SUs or TZDs, and 3) incretin-based drugs (i.e., DPP-4 inhibitors, or GLP-1 receptor agonists).

New user was 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 was considered the "primary new user" definition [Figure 1].

(EU GUIDANCE: 23 JANUARY 2013 EMA/738724/2012)

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

Figure 1. Design Schematic for Primary New Users: Ertugliflozin Example



a. Index date (Day 0) is defined by date of the first valid new use of ertugliflozin. Members are only allowed to enter the same exposure cohort once.

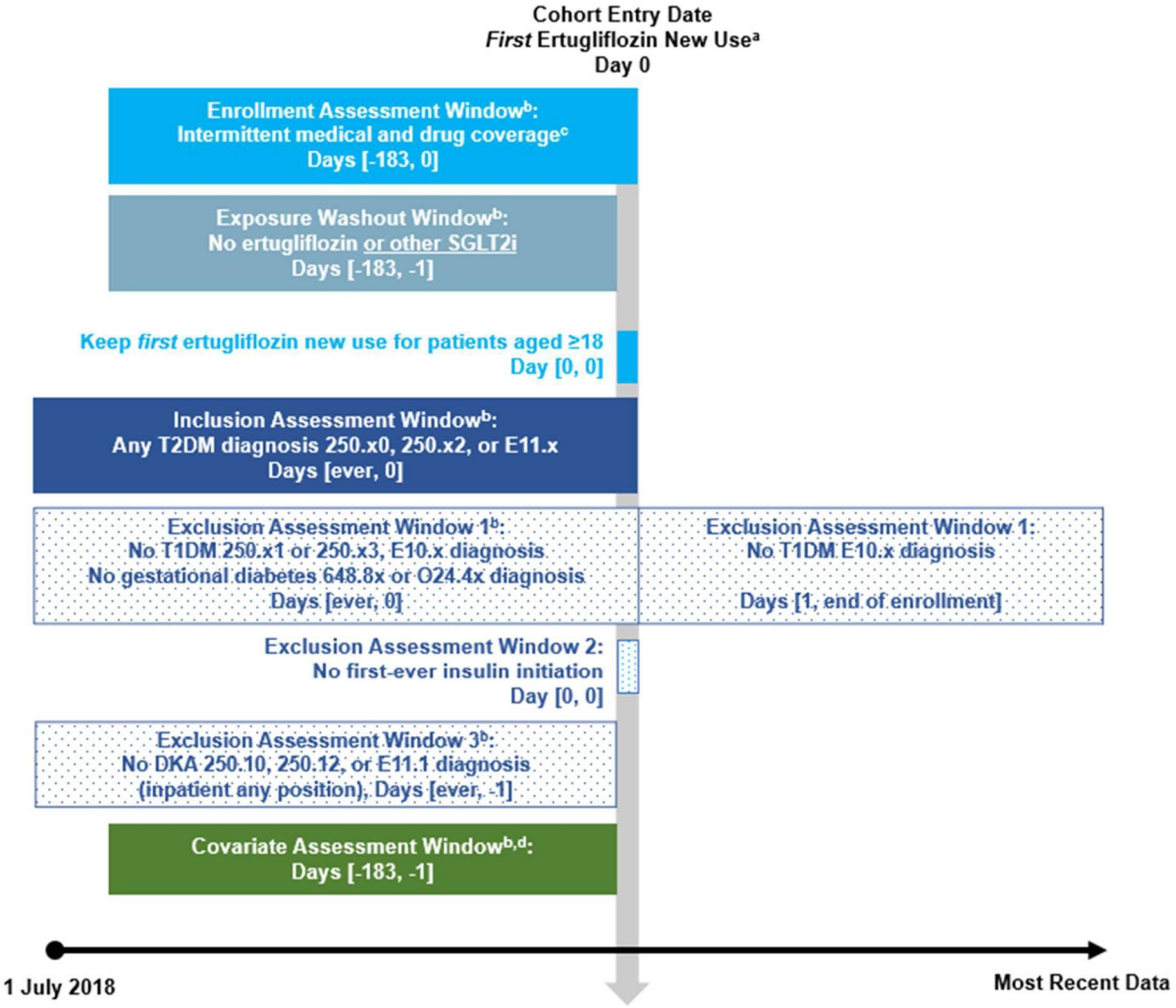
b. Assessment window may start before query start date 1 July 2018.

c. Up to 45-day gaps in medical or prescription drug plan enrollment will be allowed.

d. Except Day 0 assessment for: age, sex, calendar year, ongoing T2DM treatment, and number of ongoing unique drug use by class.

As a diagnosis of T1DM following T2DM may indicate that the T2DM diagnosis was incorrect, according to the protocol, we also assessed the number of new users of ertugliflozin and comparators 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) [Figure 2].

Figure 2. Design Schematic for Primary New Users, Narrow T2DM Definition: Ertugliflozin Example



Index date (Day 0) is defined by date of the first valid new use of ertugliflozin. Members are only allowed to enter the same exposure cohort once.

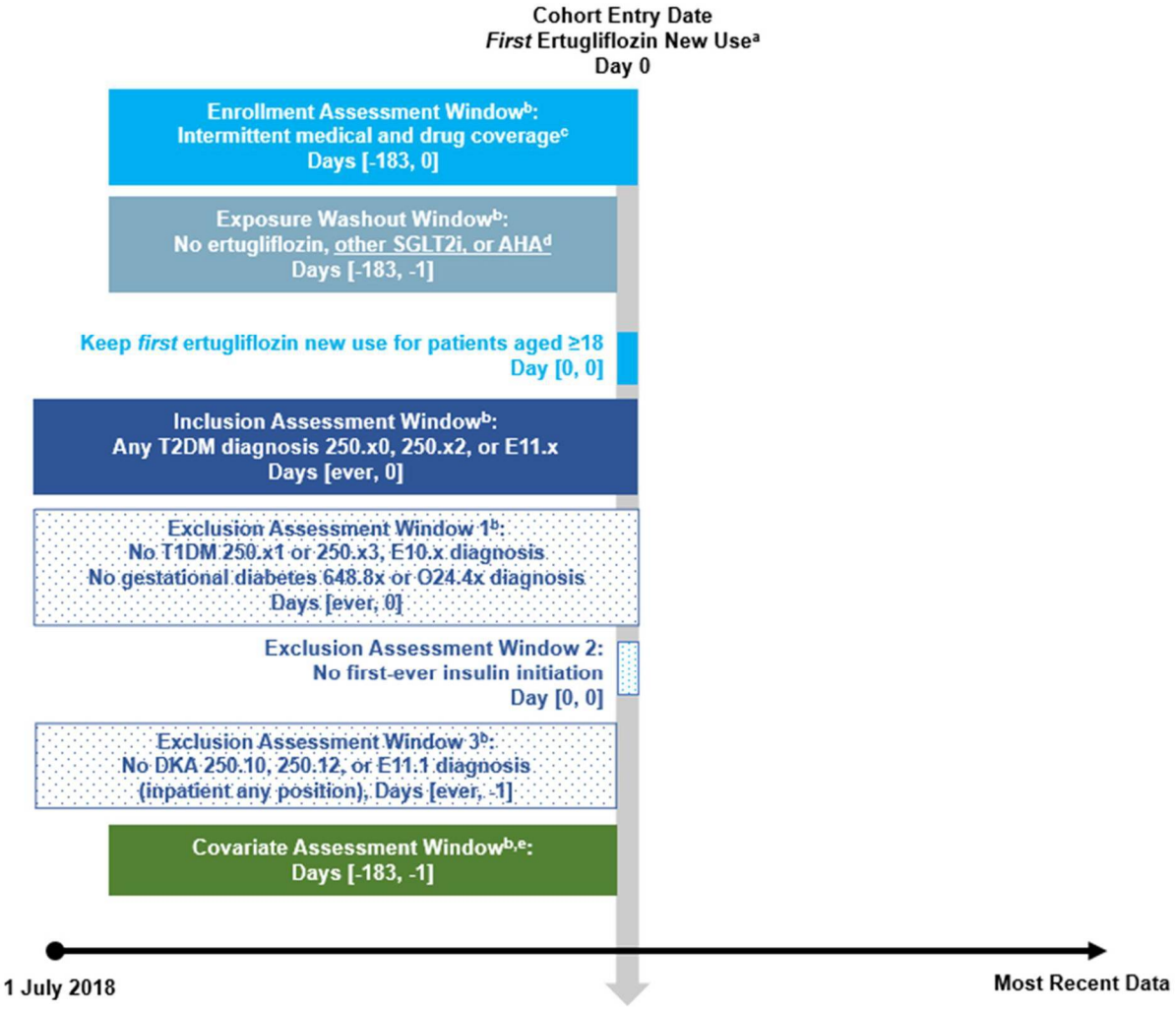
a. Assessment window may start before query start date 1 July 2018.

b. Up to 45-day gaps in medical or prescription drug plan enrollment will be allowed.

c. Except for Day 0 assessment for: age, sex, calendar year, ongoing T2DM treatment, and number of ongoing unique drug use by class.

Furthermore, we assessed the number of new users of ertugliflozin and comparators based on the "incident new user" definition, which required 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 when comparing ertugliflozin with incretin-based drugs) in the 6 months before the index date [Figure 3].

Figure 3. Design Schematic for Incident New Users: Ertugliflozin Example



a. Index date (Day 0) is defined by date of the first valid new use of ertugliflozin. Members are only allowed to enter the same exposure cohort once.

b. Assessment window may start before query start date 1 July 2018.

c. Up to 45-day gaps in medical or prescription drug plan enrollment will be allowed.

d. AHA alternates between SU/TZD and incretins (DPP-4i/GLP-1) in separate scenarios.

e. Except for assessment for: age, sex, calendar year, ongoing T2DM treatment, and number of ongoing unique drug use by class.

9.4.2 Covariates

This study examined baseline characteristics including demographics, AHA utilization, use of medications associated with DKA, comorbidity burden, pre-existing comorbidities, diabetic complications, and health services utilization. Unless otherwise specified, all characteristics were evaluated within the 6 months prior to the exposure index date (defined in [Sec. 9.3] Subjects, and medical conditions were assessed using medical encounter claims from any care setting. The full list of the patient characteristics examined is summarized in [Table 1].

Table 1. List of Variables

Category	Covariates
Demographics	Age, sex, calendar year of cohort entry
Use of AHAs	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	Charlson-Elixhauser combined comorbidity index [Ref. 5.4: 052TS6]
Pre-existing comorbidities	Acute illness (i.e., serious infection, trauma, acute febrile illness, or sepsis), surgical procedures, acute renal failure, cerebrovascular disease, coronary heart disease, heart failure, hyperlipidemia, hypertension, hypoglycemia, hypovolemia, hypoxemia, myocardial infarction, obesity, pancreatitis, peripheral artery disease, stroke, 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, amputation
Health services utilization	Number of generic medications, unique pharmacological classes, dispensing, inpatient encounters, non-acute institutional encounters, emergency department encounters, ambulatory encounters, and other ambulatory* encounters

AHA: antihyperglycemic agent; SU: sulfonylurea; TZD: thiazolidinediones; DPP-4: dipeptidyl peptidase-4; GLP-1: glucagon-like peptide-1; SGLT2: sodium-glucose cotransporter-2.
** 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.*

9.5 Data sources and measurement

9.5.1 IMEDS-DD

This study was conducted using existing electronic health insurance claims data in the IMEDS- DD, a subset of the FDA Sentinel Distributed Database. The IMEDS-DD is expected to be largely representative of the commercially insured population in US. As of 2020, the IMEDS- DD included electronic health insurance claims data available for research for over 110 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.

This study included data from five national and regional health insurers of the IMEDS-DD in the US. All listed network partners have access to their respective claims data and provided input and feedback for the study. Brief descriptions of the network partners are provided below:

Aetna, a CVS Health company, 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 an 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. The Harvard Pilgrim Health Care Institute (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 (HIRD®) is a proprietary, fully integrated, longitudinal claims database that combines medical, pharmacy, and clinical data since 2006. As of June 2021, there are 78.2 million individuals with medical and pharmacy coverage who may be included for research using the HIRD®. In addition, The HealthCore Integrated Research Environment (HIRE) has the ability to link the claims data in the HIRD® 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. The Institute is the research unit of the HealthPartners health plan and linked to claims and integrated health system data on more than 3.35 million people (both current and former enrollees) across Minnesota, Wisconsin and Iowa – with the highest density in the twin cities metropolitan area and its adjacent Wisconsin suburb.

Humana Healthcare Research (HHR) is a subsidiary of Humana Inc., headquartered in Louisville, Kentucky. Humana is a leading health plan and well-being company focused on making it easy for people to achieve their best health with clinical excellence through coordinated care. HHR conducts health economics and outcomes research focused on treatment effectiveness, drug and patient safety, patient centered research, adherence, medical and pharmacy benefit design, disease management programs, and other healthcare services. The research team also helps conduct internal research for the company. Team expertise includes areas of distributed research networks, multisite research, adherence, clinical outcomes, overall health costs, pragmatic trials Medicare benefit designs and coverage gaps, Medication Therapy Management services, survey data linking to claims, impact of clinical programs and prescription formulary design. The team has been a core part of several Distributed Research Networks.

The IMEDS-DD and Sentinel Distributed Database use the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] for standardization of demographic and clinical data elements and have routine analytic 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. Network 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.

- *Outpatient 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 network partners with such data varies [Ref. 5.4: 052WSP].

- *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 seven IMEDS-DD network partners participating in this study, six have death data and four have cause of death data [Ref. 5.4: 052X67]. 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.

Data contributing to this preliminary analysis were converted into the Sentinel Common Data Model version 7.0 or 8.0, varying by network partners. The analytic tools used were compatible to both.

9.5.2 IBM® MarketScan® Commercial Claims and Encounters Database and Medicare Supplemental Beneficiaries

As described in Objective 3, this preliminary analysis aimed to explore data solutions, should findings from Objective 1 indicate that the number of new users of ertugliflozin accumulated in the IMEDS-DD would not be anticipated to reach the target number required to perform the final analysis. As part of the exploration, this study used existing administrative claims in the IBM® MarketScan® Commercial Claims and Encounters Database (CCAE) linked with the IBM® MarketScan® Medicare Supplemental Beneficiaries (MDCR) to demonstrate the feasibility of including an additional data source to the IMEDS-DD. The reasons for using the IBM® MarketScan® CCAE and MDCR as an example included:

1) this dataset is widely used in non-interventional evaluations regarding the association of T2DM treatment and adverse outcomes in both US-based [Ref. 5.4: 05LSHQ, 052V37, 07WYN8, 07WZGS, 07WZM0, 07WYML] and international studies [Ref. 5.4: 04SDMH, 07WXVY, 07WZK0];

2) this dataset serves as the test dataset for the FDA's Sentinel System [Ref. 5.4: 07WZPC] and therefore has well-established compatibility with the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] and Sentinel routine analytic tools [Ref. 5.4: 07X0GP].

The IBM® MarketScan® databases are health insurance claims databases and capture information on outpatient, inpatient, health expenditure, enrollment and prescription drug claims of more than 147 million individuals together. The CCAE is a medical and drug insurance claims database of unique patients that includes active employees, early retirees and their dependents insured by employer-sponsored plans. The MDCR is an administrative health claims database for Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans. Both datasets together are expected to be largely representative of the employment-based insured population in US.

Like the IMEDS-DD and Sentinel Distributed Database, the IBM® MarketScan® databases used in this study were converted [Ref. 5.4: 07X0GP] into the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] for standardization of demographic and clinical data elements. Refer to [Sec. 9.5.1] IMEDS-DD for specific data information in the Sentinel Common Data Model. As a strength of both the IMEDS-DD and Sentinel Distributed Database, once a dataset is converted into the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K], the dataset shares compatibility of the Sentinel routine analytic tools to permit rapid queries. The IBM® MarketScan® databases is no exception. The Sentinel Common Data Model, when in use together with the Sentinel routine analytic tools, eliminates heterogeneity in data format and analytic programming which is otherwise commonly seen in the common protocol approach [Ref. 5.4: 07WYRJ]. In this method, transparency preservation for study design and analytic execution is high, and reproducibility becomes a unique advantage of a distributed data

network [Ref. 5.4: 07WYRJ]. Data contributing to this preliminary analysis were converted into the Sentinel Common Data Model version 8.0. All data conversion and quality assurance procedures were conducted by the IMEDS Analytic Center at HPHCI.

9.6 Bias

Despite the strengths of the IMEDS-DD and IBM® MarketScan® databases, the potential for misclassification remains due to the use of diagnostic codes, drug claim codes, or procedure codes for identification of specific medical conditions. For example, a diagnosis code could be used to rule out a certain condition instead of indicating the presence of disease, or alternatively, a disease that is truly present might not be coded in the database. There is also a substantial underestimate of obesity, as this condition is often under-recorded in electronic health record (EHR) or missing from electronic health insurance claims data. Further, analyses were limited to information obtained in the claim database of individual data partners; therefore, information on diagnoses, procedure and prescriptions outside of the contracted health care systems within individual network partners may not be captured in the IMEDS-DD or IBM® MarketScan® databases. Lastly, we required all patients to have a minimum of 6 months of medical records available prior to their index date in the database but since we examined all patient data available in the database prior to the index date for their diabetes diagnosis during the study period, the duration of medical history available varied across patients. The detailed discussion is included in [Sec. 11.2] Limitations.

9.7 Study size required to conduct MK8835-062

The study size required to conducted full study of MK8835-062 [Annex 1] was provided in Section 7.5 of the study protocol. Sample size estimates assuming different combinations of hazard ratio (HR), power, and DKA incidence rate in the comparator AHA new users are provided in [Table 2]. 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 were estimates for the matched sample after 1:1 propensity score matching. These results assumed 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 would be 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 an DKA incidence rate of 2.5 per 1,000 person-years among T2DM patients treated with comparator AHAs [Ref. 5.4: 052Y70, 052V37, 0576MY]. The 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 2. 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

		Total	DKA Incidence Rate (per 1,000 Person-Years)				
Hazard Ratio	Power	DKA Events	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.

The above preliminary study size estimate remains valid, considering similar range of DKA incidence rates in Type 2 diabetes mellitus patients treated with comparator AHAs reported by more recent literature [Ref. 5.4: 05LSHQ, 05LSH3].

The number of new users of ertugliflozin accumulated in the database so far has been assessed in this interim report and compared with target sample size to determine whether additional database(s) or data partner(s) would be necessary to help reach the target number required to perform the final analyses.

9.8 Data transformation

9.8.1 Data management

9.8.1.1 IMEDS-DD

As described in [Sec. 9.5] Data sources and measurement, 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 [Ref. 5.4: 052WPT, 052WWP, 052XV9]. 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 [Ref. 5.4: 052TNG, 052WY2] 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) [Ref. 5.4: 052X99, 052VKQ] administered by HPHCI. In this approach, data remain behind each data partner's local firewall, and network 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 employs the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] to allow data standardization across network partners. 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 was used. Data management and conversion of the Sentinel Common Data Model to analysis variables were performed using SAS software version 9.4 and above (SAS Institute, Inc., Cary, North Carolina).

For quality assurance of IMEDS-DD data, refer to [Sec. 9.10] Quality control.

As described in more detail in [Sec. 10.1.1] Protection of Human Subjects, this study was 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 was determined as exempt from IRB/IEC review. Data used in this study were anonymized and no personal identifiers were available to maintain patient confidentiality.

9.8.1.2 IBM® MarketScan® CCAE and MDCR

In order to leverage the Sentinel Common Data Model and Sentinel routine analytic tools, the IBM® MarketScan® databases used in this study were made compatible with the Sentinel infrastructure and shared the same data management, privacy protection methods, and all quality assurance standards with the IMEDS-DD, as described in [Sec. 9.8.1.1]. All IBM® MarketScan® database queries were conducted locally within the IMEDS Analytic Center at HPHCI.

9.9 Statistical methods

9.9.1 Main statistical methods

Patient characteristics, comorbidities, and health services utilization were summarized via descriptive analyses. Continuous variables were reported as means and standard deviations (SDs), and categorical variables were summarized as number and proportion of the total study population in each cohort.

9.9.2 Missing values

The study included three continuous variables in the general characteristic assessment: age, Charlson-Elixhauser combined comorbidity score [Ref. 5.4: 052TS6], AHA utilization, and health services utilization metrics. All were expected to be non-missing, given that cohort members were required to have age information available in order to meet eligibility requirement and that both comorbidity score and the count of health services have their respective numeric lower boundaries (for example, zero or no non-antihyperglycemic use).

The study dichotomized all categorical variables in the general characteristic assessment. The Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] allows the assignment to "unknown" value for these exception variables but no missing value. Absence of any diagnosis, procedure, or drug code required in the condition or drug utilization algorithms was considered that no condition or drug utilization was present.

9.9.3 Sensitivity analyses

None.

9.9.4 Amendments to the statistical analysis plan

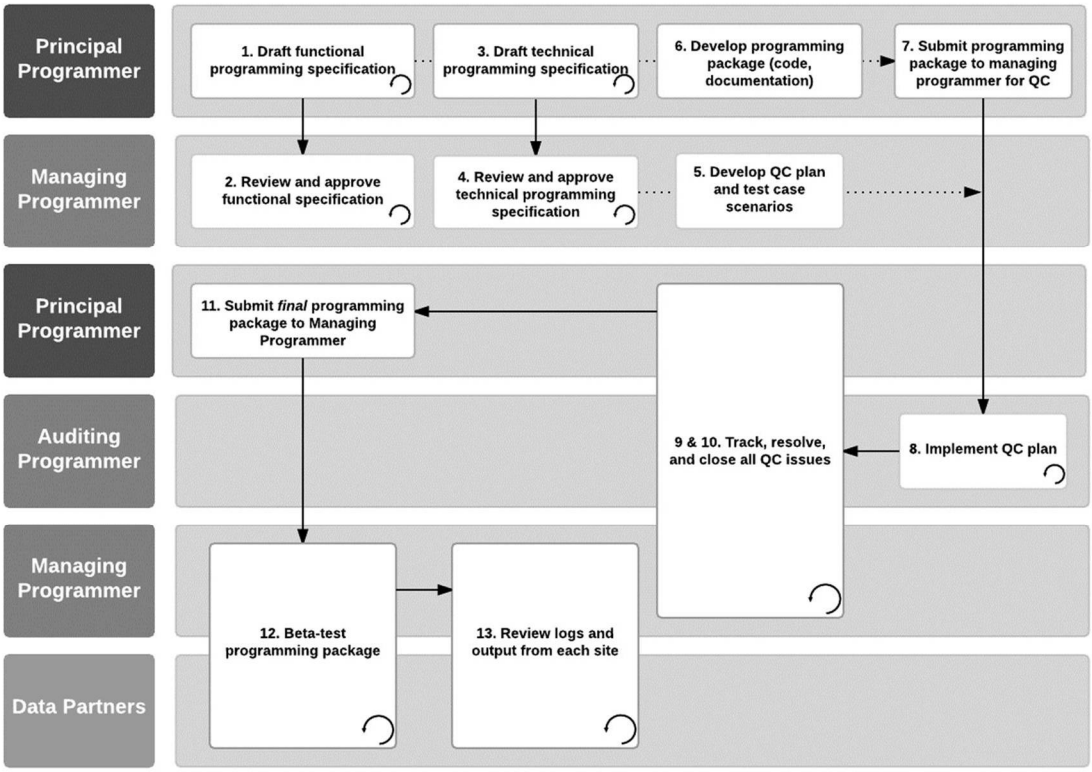
None.

9.10 Quality control

As described in [Sec. 9.5] Data sources and measurement, the IMEDS-DD is a subset of Sentinel Distributed Database and share 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 [Ref. 5.4: 052XV9, 052XSG]. 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 [Ref. 5.4: 052W62]. 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 custom SAS programs and deliverables. [Figure 4] illustrates the standard operating procedures (SOPs) for SAS programming quality assurance and quality control in the Sentinel System.

Figure 4 Standard Operating Procedure for SAS Programming Quality Assurance and Quality Control in the Sentinel System



10 RESULTS

10.1 Participants

10.1.1 Protection of Human Subjects

The preliminary analysis used existing databases IMEDS-DD and IBM® MarketScan® databases. Data were anonymized and no personal identifiers were available to maintain patient confidentiality. This work was determined to be exempt from Institution Review Board review (HPHC IRB Review #1077644 and FDA Foundation's IRB Review, WCG IRB, formerly NEIRB, # IRB2187).

The preliminary analysis was conducted in accordance with all legal and regulatory requirements. Additionally, the conduct of feasibility assessment was adhered 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 [Ref. 5.4: 052W62].

10.2 Main results

10.2.1 New users of ertugliflozin in the IMEDS-DD

A total of 647 patients who initiated ertugliflozin were identified in the IMEDS-DD between 1 July 2018 and 28 February 2021, based on the "primary new user" definition. Among those, 636 met the criteria for "narrow T2DM population" and did not have a T1DM diagnosis at any time during the study. There were 419 new users of ertugliflozin meeting the criteria for "incident new users" definition relevant to SU/TZD, defined as not having used any SGLT2 inhibitors (including ertugliflozin) nor SU or TZD during the six months prior to the index date. A total of 373 new users of ertugliflozin met the criteria for "incident new users" definition relevant to DPP-4 inhibitors/GLP-1 receptor agonists, defined as not having used any SGLT2 inhibitors (including ertugliflozin) nor any incretin-based drugs during the six months prior to the index date.

The baseline characteristics of primary new users of ertugliflozin are shown in [Table 3]. Of 647 ertugliflozin initiators, 43.0% were female and the mean age was 57.2 years (SD=11.0 years). One hundred and three (15.9%) patients initiated ertugliflozin as monotherapy. The most concomitantly utilized AHA class at the index date was metformin (61.8%), followed by DPP-4 inhibitors (27.0%) and SU (23.8%). The most common comorbidities included hyperlipidemia (72.8%), hypertension (69.7%), and obesity (36.2%). Individuals with a history of cardiovascular disease (CVD) represented 19.2% of new users of ertugliflozin, categorized based on ICD-10-CM diagnoses for myocardial infarction, coronary heart disease, heart failure, peripheral artery disease, cerebrovascular disease, or stroke.

Table 3. Baseline Characteristics of New Users of Ertugliflozin identified in the IMEDS Distributed Database between 1 July 2018 and 28 February 2021

	Ertugliflozin[1]	
	N/Mean	**%/Std Dev[2]**
Number of patients	647	100%
Demographics on the index date		
Age, in years	57.2	11.0
18-44	88	13.6%
45-64	415	64.1%
65-74	104	16.1%
≥75	40	6.2%
Sex, female	278	43.0%
Calendar year of initiation		
2018	102	15.8%
2019	272	42.0%

(EU GUIDANCE: 23 JANUARY 2013 EMA/738724/2012)

	Ertugliflozin[1]	
	N/Mean	%/Std Dev[2]
2020	273	42.2%
2021	0	0.0%
Antihyperglycemic management on the index date		
Monotherapy	103	15.9%
Dual therapy	232	35.9%
Triple therapy or more	312	48.2%
Concomitant antihyperglycemic agent use[1] on the index date		
Metformin	400	61.8%
DPP-4 inhibitor	175	27.0%
SU	154	23.8%
GLP-1 agonist	101	15.6%
Insulin	98	15.1%
TZD	34	5.3%
Others (alpha-glucosidase inhibitor, meglitinide)	6	0.9%
Use of medications[1] associated with DKA on the index date		
Clozapine/olanzapine	1	0.2%
Lithium	1	0.2%
Terbutaline	0	0.0%
Oral corticosteroid	13	2.0%
Thiazide	55	8.5%
Pentamidine	0	0.0%
Comorbidity burden		
Charlson-Elixhauser combined comorbidity score	0.8	1.6
Comorbidity/pre-existing conditions		
Acute illness (i.e., serious infection, trauma, acute febrile illness, or sepsis)	75	11.6%
Any surgical procedures	471	72.8%
Surgery, inpatient only	11	1.7%
Acute renal failure	8	1.2%
Cardiovascular disease	124	19.2%
Cerebrovascular disease	28	4.3%
Coronary heart disease	89	13.8%
Heart failure	16	2.5%
Myocardial infarction	17	2.6%
Peripheral artery disease	26	4.0%
Stroke	16	2.5%
Hyperlipidemia	471	72.8%
Hypertension	451	69.7%
Hypoglycemia	7	1.1%
Hypovolemia	0	0.0%

	Ertugliflozin[1]	
	N/Mean	%/Std Dev[2]
Hypoxemia	10	1.5%
Obesity	234	36.2%
Pancreatitis	4	0.6%
Thyroid disorders	122	18.9%
Diabetic complications		
Moderate-to-severe renal insufficiency	36	5.6%
Nephropathy	61	9.4%
Neuropathy	111	17.2%
Retinopathy	49	7.6%
Amputation	0	0.0%
Health Services Utilization		
Number of unique drug classes	8.1	4.7
Number of unique generic medications	8.1	5
Number of dispensings	24.3	18.8
Number of inpatient encounters	0.1	0.3
Number of non-acute institutional encounters	0.0	0.1
Number of emergency department encounters	0.3	0.8
Number of ambulatory encounters	7.2	7.3
Number of other ambulatory encounters	1.2	3.2

[1] Index exposure and individual drug class utilization are identified based on 9-digit National Drug Codes (NDCs), whereas drug class utilization summary is identified based on 11-digit NDCs recorded in outpatient pharmacy dispensings.

[2] Value represents standard deviation (Std Dev or Std) where no % follows.

T2DM: Type 2 diabetes mellitus. T1DM: Type 1 diabetes mellitus. SGLT2: sodium-glucose cotransporter-2. SU: sulfonylurea. TZD: thiazolidinedione. DPP-4: dipeptidyl peptidase-4. GLP-1: glucagon-like peptide-1.

10.2.2 New users of SU/TZD in the IMEDS-DD

A total of 161,941 patients who initiated SU or TZD were identified in the IMEDS-DD between 1 July 2018 and 28 February 2021, based on "primary new user" definition. Among those, 158,831 met the criteria for "narrow T2DM population" and did not have a T1DM diagnosis at any time during the study. There were 161,881 new users of SU or TZD meeting the criteria for "incident new users", defined as not having used any SGLT2 inhibitors (including ertugliflozin) nor SU or TZD during the six months prior to the index date.

The baseline characteristics of primary new users of SU or TZD are summarized in [Table 4]. Among 161,941 new users of SU or TZD, mean age was 66.8 years (SD=11.1 years), 48.0% were women, and 30.2% initiated SU or TZD as monotherapy. The most concomitantly utilized AHA class at the index date was metformin (61.2%), followed by DPP-4 inhibitors (9.6%) and insulin (8.9%). The most common comorbidities included hypertension (76.3%), hyperlipidemia (68.4%), and obesity (32.5%), and 31.6% had a history of CVD.

Table 4. Baseline Characteristics of New Users of Sulfonylurea or Thiazolidinedione (SU/TZD) in the IMEDS Distributed Database between 1 July 2018 and 28 February 2021

	SU/TZD[1]	
	N/Mean	%/Std Dev[2]
Number of patients	161,941	100%
Demographics on the index date		
Age, in years	66.8	11.1
18-44	7,681	4.7%
45-64	52,073	32.2%
65-74	65,005	40.1%
≥75	37,182	23.0%
Sex, female	77,654	48.0%
Calendar year of initiation		
2018	32,977	20.4%
2019	68,379	42.2%
2020	60,369	37.3%
2021	216	0.1%
Antihyperglycemic management on the index date		
Monotherapy	48,913	30.2%
Dual therapy	90,834	56.1%
Triple therapy or more	22,194	13.7%
Concomitant antihyperglycemic agent use[1] on the index date		
Metformin	99,098	61.2%
DPP-4 inhibitor	15,610	9.6%
Insulin	14,448	8.9%
GLP-1 agonist	6,204	3.8%
SGLT2 inhibitors	1,061	0.7%
Others (alpha-glucosidase inhibitor, meglitinide)	905	0.6%
Use of medications[1] associated with DKA on the index date		
Clozapine/olanzapine	589	0.4%
Lithium	278	0.2%
Terbutaline	2	0.0%
Oral corticosteroid	5,784	3.6%
Thiazide	18,417	11.4%
Pentamidine	0	0.0%
Comorbidity burden		
Charlson-Elixhauser combined comorbidity score	1.8	2.5
Comorbidity/pre-existing conditions		
Acute illness (i.e., serious infection, trauma, acute febrile illness, or sepsis)	29,213	18.0%
Any surgical procedures	121,945	75.3%
Surgery, inpatient only	8,065	5.0%

	SU/TZD[1]	
	N/Mean	**%/Std Dev[2]**
Acute renal failure	8,829	5.5%
Cardiovascular disease	51,247	31.6%
Cerebrovascular disease	13,106	8.1%
Coronary heart disease	33,290	20.6%
Heart failure	15,166	9.4%
Myocardial infarction	8,568	5.3%
Peripheral artery disease	15,064	9.3%
Stroke	11,622	7.2%
Hyperlipidemia	110,743	68.4%
Hypertension	123,521	76.3%
Hypoglycemia	2,140	1.3%
Hypovolemia	539	0.3%
Hypoxemia	7,803	4.8%
Obesity	52,564	32.5%
Pancreatitis	1,123	0.7%
Thyroid disorders	29,020	17.9%
Diabetic complications		
Moderate-to-severe renal insufficiency	30,723	19.0%
Nephropathy	40,142	24.8%
Neuropathy	34,730	21.4%
Retinopathy	12,614	7.8%
Amputation	406	0.3%
Health Services Utilization		
Number of unique drug classes	7.9	4.8
Number of unique generic medications	7.8	5
Number of dispensings	20.1	16.6
Number of inpatient encounters	0.1	0.5
Number of non-acute institutional encounters	0	0.3
Number of emergency department encounters	0.5	1.5
Number of ambulatory encounters	8.6	10
Number of other ambulatory encounters	1.6	5.1

[1] Index exposure and individual drug class utilization are identified based on 9-digit National Drug Codes (NDCs), whereas drug class utilization summary is identified based on 11-digit NDCs recorded in outpatient pharmacy dispensings.

[2]Value represents standard deviation (Std Dev or Std) where no % follows.

T2DM: Type 2 diabetes mellitus. T1DM: Type 1 diabetes mellitus. SGLT2: sodium-glucose cotransporter-2. SU: sulfonylurea. TZD: thiazolidinedione. DPP-4: dipeptidyl peptidase-4. GLP-1: glucagon-like peptide-1.

10.2.3 New users of incretin-based drugs in the IMEDS-DD

A total of 157,203 patients who initiated incretin-based drug were identified in the IMEDS- DD between 1 July 2018 and 28 February 2021, based on the "primary new user" definition. Among those, 153,490 met the criteria for "narrow T2DM population" and did not have T1DM diagnosis any time during the study. There were 157,106 incident new users of incretin-based drug meeting the criteria for "incident new users", defined as not having used any SGLT2 inhibitors (including ertugliflozin) nor DPP-4 inhibitors or GLP-1 receptor agonists during the six months prior to the index date.

The baseline characteristics of primary new users of incretin-based drugs are presented in [Table 5]. Of 157,203 new users of incretin-based drugs, 52.3% were female and the mean age was 65.3 years (SD=11.4 years). About one-fifth patients initiated either DPP-4 inhibitors or GLP-1 receptor agonists as monotherapy. The most concomitantly utilized AHA class at the index date was metformin (60.3%), followed by SU (31.4%) and insulin (19.9%). The most common comorbidities included hypertension (79.6%), hyperlipidemia (72.4%), and obesity (40.2%), and 32.6% had a history of CVD.

Table 5. Baseline Characteristics of New Users of Incretin-Based Drugs in the IMEDS Distributed Database between 1 July 2018 and 28 February 2021

	Incretin-Based Drugs[1]	
	N/Mean	**%/Std Dev[2]**
Number of patients	157,203	100%
Demographics on the index date		
Age, in years	65.3	11.4
18-44	9,622	6.1%
45-64	57,317	36.5%
65-74	59,430	37.8%
≥75	30,834	19.6%
Sex, female	82,274	52.3%
Calendar year of initiation		
2018	29,081	18.5%
2019	66,688	42.4%
2020	61,159	38.9%
2021	275	0.2%
Antihyperglycemic management on the index date		
Monotherapy	30,621	19.5%
Dual therapy	74,253	47.2%
Triple therapy or more	52,329	33.3%
Concomitant antihyperglycemic agent use[1] on the index date		
Metformin	94,807	60.3%
SU	49,425	31.4%
Insulin	31,221	19.9%

	Incretin-Based Drugs[1]	
	N/Mean	%/Std Dev[2]
TZD	8,293	5.3%
SGLT2 inhibitors	3,726	2.4%
Others (alpha-glucosidase inhibitor, meglitinide)	1,339	0.9%
Use of medications[1] associated with DKA on the index date		
Clozapine/olanzapine	654	0.4%
Lithium	353	0.2%
Terbutaline	2	0.0%
Oral corticosteroid	4,097	2.6%
Thiazide	18,981	12.1%
Pentamidine	0	0.0%
Comorbidity burden		
Charlson-Elixhauser combined comorbidity score	1.9	2.5
Comorbidity/pre-existing conditions		
Acute illness (i.e., serious infection, trauma, acute febrile illness, or sepsis)	29,456	18.7%
Any surgical procedures	125,200	79.6%
Surgery, inpatient only	7,821	5.0%
Acute renal failure	9,312	5.9%
Cardiovascular disease	51,223	32.6%
Cerebrovascular disease	12,850	8.2%
Coronary heart disease	33,845	21.5%
Heart failure	15,924	10.1%
Myocardial infarction	8,291	5.3%
Peripheral artery disease	14,902	9.5%
Stroke	11,325	7.2%
Hyperlipidemia	113,791	72.4%
Hypertension	125,071	79.6%
Hypoglycemia	3,728	2.4%
Hypovolemia	597	0.4%
Hypoxemia	8,031	5.1%
Obesity	63,160	40.2%
Pancreatitis	697	0.4%
Thyroid disorders	32,153	20.5%
Diabetic complications		
Moderate-to-severe renal insufficiency	32,303	20.5%
Nephropathy	42,062	26.8%
Neuropathy	38,347	24.4%
Retinopathy	15,315	9.7%
Amputation	422	0.3%

	Incretin-Based Drugs[1]	
	N/Mean	%/Std Dev[2]
Health Services Utilization		
Number of unique drug classes	9.2	4.9
Number of unique generic medications	9.1	5.1
Number of dispensings	24.8	18.1
Number of inpatient encounters	0.1	0.5
Number of non-acute institutional encounters	0	0.2
Number of emergency department encounters	0.5	1.4
Number of ambulatory encounters	9.8	10.4
Number of other ambulatory encounters	1.7	5.3

[1] Index exposure and individual drug class utilization are identified based on 9-digit National Drug Codes (NDCs), whereas drug class utilization summary is identified based on 11-digit NDCs recorded in outpatient pharmacy dispensings.

[2] Value represents standard deviation (Std Dev or Std) where no % follows.

T2DM: Type 2 diabetes mellitus. T1DM: Type 1 diabetes mellitus. SGLT2: sodium-glucose cotransporter-2. SU: sulfonylurea. TZD: thiazolidinedione. DPP-4: dipeptidyl peptidase-4. GLP-1: glucagon-like peptide-1.

10.2.4 Results from considering additional data sources

As presented in [Sec. 10.2.1] above, it is not anticipated that the required study sample size will be reached by the milestone date for the final report. Recognizing this issue, the first preliminary analysis explored data solutions to increase the sample size, as described in Objective 3.

Using the IBM® MarketScan® databases as an example, the exploratory work demonstrated the advantageous ability of a distributed data network such as the IMEDS-DD for sample size expansion and deployment of analytics. Replicating the same study design described in [Sec. 9] and using the same Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K] and analytic programming, a total of 2,004 patients who initiated ertugliflozin were identified in the IBM® MarketScan® databases between 1 July 2018 and 30 June 2020, based on the "primary new user" definition [Table 6]. Among these, 43.8% were female and the mean age was 53.3 years (SD=9.2). The most concomitantly utilized AHA class at the index date was metformin (68.6%), followed by DPP-4 inhibitors (30.0%) and SU (21.1%). The most common comorbidities included hyperlipidemia (67.7%), hypertension (65.0%), and obesity (33.7%). Individuals with a history of CVD represented 9.6% of new users of ertugliflozin. In contrast, 95,375 primary SU/TZD new users and 126,343 primary incretin new users were identified in the IBM® MarketScan® databases.

Table 6. Baseline Characteristics of New Users of Ertugliflozin in the IMEDS Distributed Database versus in the IBM® MarketScan® Commercial Claims and Encounters Database (CCAE) linked with the IBM® MarketScan® Medicare Supplemental Beneficiaries (MDCR)

	Ertugliflozin[1]			
	IMEDS-DD		**CCAE-MDCR**	
	N/Mean	**%/Std Dev[2]**	**N/Mean**	**%/Std Dev[2]**
Number of patients	647	100%	2,004	100.0%
Demographics on the index date				
Age, in years	57.2	11.0	53.3	9.2
18-44	88	13.6%	363	18.1%
45-64	415	64.1%	1,559	77.8%
65-74	104	16.1%	68	3.4%
≥75	40	6.2%	14	0.7%
Female sex	278	43.0%	877	43.8%
Calendar year of initiation				
2018	102	15.8%	437	21.8%
2019	272	42.0%	1,084	54.1%
2020	273	42.2%	483	24.1%
Antihyperglycemic management on the index date				
Monotherapy	103	15.9%	258	12.9%
Dual therapy	232	35.9%	738	36.8%
Triple therapy or more	312	48.2%	1,008	50.3%
Concomitant antihyperglycemic agent use[1] on the index date				
Metformin	400	61.8%	1,374	68.6%
DPP-4 inhibitor	175	27.0%	601	30.0%
SU	154	23.8%	423	21.1%
GLP-1 agonist	101	15.6%	342	17.1%
Insulin	98	15.1%	297	14.8%
TZD	34	5.3%	130	6.5%
Others (alpha-glucosidase inhibitor, meglitinide)	6	0.9%	7	0.3%
Use of medications[1] associated with DKA on the index date				
Clozapine/olanzapine	1	0.2%	1	0.0%
Lithium	1	0.2%	4	0.2%
Terbutaline	0	0.0%	0	0.0%
Oral corticosteroid	13	2.0%	20	1.0%
Thiazide	55	8.5%	132	6.6%
Pentamidine	0	0.0%	0	0.0%
Comorbidity burden				
Charlson-Elixhauser combined comorbidity score	0.8	1.6	0.5	1.2
Comorbidity/pre-existing conditions				
Acute illness (i.e., serious infection, trauma, acute	75	11.6%	191	9.5%

	Ertugliflozin[1]			
	IMEDS-DD		CCAE-MDCR	
febrile illness, or sepsis)				
Any surgical procedures	471	72.8%	1,476	73.7%
Surgery, inpatient only	11	1.7%	22	1.1%
Acute renal failure	8	1.2%	11	0.5%
Cardiovascular disease	124	19.2%	193	9.6%
Cerebrovascular disease	28	4.3%	34	1.7%
Coronary heart disease	89	13.8%	123	6.1%
Heart failure	16	2.5%	24	1.2%
Myocardial infarction	17	2.6%	21	1.0%
Peripheral artery disease	26	4.0%	50	2.5%
Stroke	16	2.5%	30	1.5%
Hyperlipidemia	471	72.8%	1,357	67.7%
Hypertension	451	69.7%	1,303	65.0%
Hypoglycemia	7	1.1%	13	0.6%
Hypovolemia	0	0.0%	0	0.0%
Hypoxemia	10	1.5%	24	1.2%
Obesity	234	36.2%	675	33.7%
Pancreatitis	4	0.6%	6	0.3%
Thyroid disorders	122	18.9%	300	15.0%
Diabetic complications				
Moderate-to-severe renal insufficiency	36	5.6%	46	2.3%
Nephropathy	61	9.4%	134	6.7%
Neuropathy	111	17.2%	213	10.6%
Retinopathy	49	7.6%	89	4.4%
Amputation	0	0.0%	1	0.0%
Health Services Utilization				
Number of unique drug classes	8.1	4.7	7.8	4.2
Number of unique generic medications	8.1	5	7.7	4.4
Number of dispensings	24.3	18.8	21	14.9
Number of inpatient encounters	0.1	0.3	0.0	0.2
Number of non-acute institutional encounters	0.0	0.1	0.0	0.0
Number of emergency department encounters	0.3	0.8	0.2	0.6
Number of ambulatory encounters	7.2	7.3	6.7	6.9
Number of other ambulatory encounters	1.2	3.2	1.5	2.7

[1] Index exposure and individual drug class utilization are identified based on 9-digit National Drug Codes (NDCs), whereas drug class utilization summary is identified based on 11-digit NDCs recorded in outpatient pharmacy dispensings.

[2]Value represents standard deviation (Std Dev or Std) where no % follows.

T2DM: Type 2 diabetes mellitus. T1DM: Type 1 diabetes mellitus. SGLT2: sodium-glucose cotransporter-2. SU: sulfonylurea. TZD: thiazolidinedione. DPP-4: dipeptidyl peptidase-4. GLP-1: glucagon-like peptide-1.

11 DISCUSSION

11.1 Key results

A total of 647 primary new users of ertugliflozin were identified in the IMEDS-DD between 1 July 2018 and 28 February 2021, compared to 161,941 primary new users of SU or TZD and 157,203 primary new users of incretin-based drug identified during the same period. The demographic and clinical characteristics of these ertugliflozin new users were comparable with those reported in observational studies of other SGLT2 inhibitors. The reference studies include EMPagliflozin compaRative effectIveness and SafEty (EMPRISE; EUPAS20677) [Ref. 5.4: 05LSHM, 07WZGS]; Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases (OBSERVE-4D) [Ref. 5.4: 07WZM0]; Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors (CVD-REAL; CVD- REAL Nordic) [Ref. 5.4: 04SDMH, 07WXVY, 07WZK0]; Diabetic Ketoacidosis in Patients with Type 2 Diabetes Treated with Sodium Glucose Co-Transporter 2 Inhibitors versus Other Antihyperglycemic Agents: An Observational Study of Four US Administrative Claims Databases (EUPAS23705) [Ref. 5.4: 05LSHQ]; and two Canadian Network for Observational Drug Effect Studies (CNODES) studies assessing the association of SGLT2 inhibitor exposure and various outcomes [Ref. 5.4: 07WYNH, 05LSH3]. Of note, most of these large-scale observational studies adopted a multi-database approach in the analysis which resembles the IMEDS framework used here.

The preliminary analyses in the current phase indicate that the required study sample size will not be reached [Sec. 9.7] by the milestone date for the final report if data are limited to those from five regional and national health insurers. Recognizing this issue, this first preliminary analysis explored data solutions that will be utilized in the next preliminary analysis to increase the sample size. The unique advantage of a distributed data network such as the IMEDS-DD is its capacity of sample size expansion and reproducibility of the same analytic programming among different databases. The IMEDS-DD adopts the Sentinel Common Data Model [Ref. 5.4: 052TNG, 052Y0K]. The Sentinel Common Data Model, when used with the Sentinel routine analytic tools, permits different network partners to replicate the same study design and use the same analytic programming, which in turn eliminates heterogeneity in data format and analytic programming otherwise typically seen in the common protocol approach [Ref. 5.4: 07WYRJ]. Using the Sentinel Common Data Model and the same analytic programming, a total of 2,004 patients who initiated ertugliflozin were identified in the IBM® MarketScan® databases between 1 July 2018 and 30 June 2020.

Using the same methods for replication in the IBM® MarketScan® databases, we also consider adding the following data sources to the data network in the next preliminary analysis: 1) Optum© Database; 2) the Centers for Medicare & Medicaid Services (CMS) Medicare Fee- for-Service Research Identifiable Files; and/or 3) multi-state Medicaid database. The results will be summarized in the Interim Report 2.

Optum is an adjudicated administrative health claims database for members with private health insurance, who are fully insured in commercial plans, Legacy Medicare Choice Lives (prior to January 2006), and Medicare Advantage (Medicare Advantage Prescription Drug coverage starting January 2006). The population is primarily representative of US commercially insured (0-65 years old) with some Medicare-insured (65+ years old) members.

The CMS Medicare fee-for-service is a nationwide federal health insurer in the United States for eligible individuals primarily aged 65 years and older that provides coverage for inpatient and outpatient services and prescription medications. Based on utilization records publicly available on the CMS website [Ref. 5.4: 07XC64], there were over 7,000 ertugliflozin prevalent users as of the end of 2019 in the Medicare Fee-for-Service population before applying the inclusion and exclusion criteria pre-specified in the protocol MK8835-062 [Annex 1]. Both Optum and CMS Medicare fee-for-service are existing contributing datasets to the Sentinel Distributed Database [Ref. 5.4: 07X0GQ] and are actively used by the Sentinel System.

Multi-state Medicaid database represents adjudicated health insurance claims for multiple statewide Medicaid, public health insurance program for low-income residents, in United States, and includes hospital discharge diagnoses, outpatient diagnoses and procedures, and outpatient pharmacy claims. Although the multi-state Medicaid database has not yet been used by the Sentinel System, we will explore the possibilities of converting the data into the Sentinel Common Data Model in the next preliminary analysis. Once the dataset is converted into the Sentinel Common Data Model, the dataset will share compatibility of the Sentinel routine analytic tools to permit reproducibility of the same analytic programming and integrability into the IMEDS-DD.

Of note, all these candidate data sources have been widely used in other recent observational assessments for risk of DKA with other SGLT2 inhibitors [Ref. 5.4: 052Y70, 05LSHQ, 05LSH3, 052V37, 05LSHP].

11.2 Limitations

Several limitations should be considered in the context of the preliminary analysis.

First, despite the strengths of the IMEDS-DD or IBM® MarketScan® CCAE and MDCR, there is the potential for misclassification due to the use of diagnostic, drug, or procedure codes for identification of specific medical conditions. For example, a diagnosis code could be used to rule out a certain condition instead of indicating the presence of disease, or alternatively, a disease that is truly present might not be coded in the database. Data input errors could also be present in the databases.

Second, the preliminary analysis was limited to information captured in the IMEDS-DD or IBM® MarketScan® CCAE and MDCR. As with any other non-interventional database studies using health insurance administrative claims, patient medical history and

treatment exposure in this study were captured by health services utilization. These types of data are 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). Race/ethnicity, clinical details (e.g., Hemoglobin A_{1c} [HbA_{1c}] laboratory results), and death are often missing or incomplete. There is also substantial underestimate of obesity or lifestyle measures, such as alcohol use, as they are often under-recorded in EHR or missing from health insurance administrative and claims data.

Third, drug exposure was inferred from outpatient pharmacy claims in this study, and days of supply information on these claims was used to determine study drug exposure and baseline or concomitant utilization. These outpatient pharmacy claims do not necessarily mean that patient consumed the drug and are a surrogate measure of actual drug exposure. This could result in drug exposure misclassification.

Fourth, as a distributed data network, the IMEDS-DD does not guarantee data uniqueness at the patient 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 study results are generalizable to the commercial health insurance population from which the study population was derived as well as others with similar characteristics but may not be representative of the uninsured or elderly people of US population with fee-for-service Medicare insurance.

11.3 Interpretation

The preliminary analysis was first conducted within five regional and national health insurer network partners contributing to the IMEDS-DD. The distributed data network in the IMEDS- DD resembles the multi-database design commonly adopted in recently published observational safety studies of SGLT2 inhibitors such EMRPISE, OBSERVE-4D, and CVD- REAL. The IMEDS framework employs the Sentinel Common Data Model for standardization of demographic and clinical data elements from various network partners [Ref. 5.4: 052TNG, 052Y0K]. By doing so, the IMEDS-DD enables rapid queries across the included databases thanks to compatibility and use of the analytic tools established and actively maintained by the Sentinel System. Successful replication of the planned analysis in the IBM® MarketScan® databases, as described in [Sec. 10.2.4], demonstrated feasibility of sample size expansion via this approach. In this way the sample size will be increased significantly for the next interim report and the final analysis.

Using the same methods for replication in the IBM® MarketScan® databases, we also propose to add the following data sources to the data network in the next preliminary analysis: 1) Optum© Database, 2) CMS Medicare Fee-for-Service Research Identifiable

Files, and/or 3) multi-state Medicaid database, and continue to monitor the numbers of new users of ertugliflozin and AHA comparators in the next preliminary analysis due for submission to EMA in 2022. The number of person-years of exposure to ertugliflozin that has been accumulated in the database will also be provided in the next preliminary analysis (Interim Report 2), when one more year of follow up data becomes available.

Despite the insufficient cohort size of new users of ertugliflozin projected for the final analysis by this preliminary analysis, the baseline characteristics of these identified patients remain largely similar to the real-world SGLT2 inhibitor user profiles reported by recent publications [Ref. 5.4: 05LSHM, 05LSHQ, 07WZGS, 07WZM0, 04SDMH, 07WXVY, 07WZK0, 07WYNH, 05LSH3].

11.4 Generalisability

The study results are generalizable to the commercial health insurance population from which the study population was derived as well as other populations with similar characteristics. Introducing additional data sources, such as multi-state Medicaid database and/or the CMS Medicare Fee-for-Service Research Identifiable Files, will improve the precision of the effect estimates. It will also improve the generalizability with real-world T2DM patients who receive treatment, including the publicly insured population with Medicare or Medicaid coverage in the United States, in addition to the commercially insured population.

12 CONCLUSION

After the implementation of inclusion and exclusion criteria pre-specified in the study protocol, we identified 2,651 patients who initiated ertugliflozin (647 in IMEDS-DD and 2,004 in the IBM® MarketScan® databases) between 1 July 2018 and the most recent data available in each database (IMEDS-DD: 21 February 2021; IBM® MarketScan® databases: 30 June 2020). The demographic and clinical characteristics of these ertugliflozin new users were comparable with those reported in observational studies of other SGLT2 inhibitors, such as EMPRISE [Ref. 5.4: 05LSHM, 07WZGS], OBSERVE-4D [Ref. 5.4: 07WZM0], and CVD-REAL [Ref. 5.4: 04SDMH, 07WXVY, 07WZK0]. We also identified 257,316 new users of SU/TZD (161,941 in IMEDS-DD and 95,375 in IBM® MarketScan® databases) and 283,546 new users of incretin-based drugs (157,203 in IMEDS-DD and 126,343 in IBM® MarketScan® databases) during the same study period.

The analyses in the current phase indicate that study sample size of new users of ertugliflozin accrued in the IMEDS-DD is unlikely to reach the target number by the milestone date for the final report, if data are limited to those from five regional and national health insurers included in this first preliminary analysis. However, the successful replication of the planned preliminary analysis in the IBM® MarketScan® databases demonstrated the feasibility of sample size expansion within the IMEDS framework. To reach the target number by the milestone date for the final report, we will consider adding data from additional network partners or other sources, including Optum© Database, multi-state Medicaid database, and/or the CMS Medicare Fee-for-Service Research Identifiable Files.

Upon approval of approach to increase sample size as described in the [Sec. 11.1] Key results of the Interim Report 1 by the EMA, an amendment of the protocol will be adopted to reflect the recommendation made based on preliminary analyses presented in the current interim report.

REFERENCES

[Ref. 5.4: 04SDMH] Kosiborod M, Cavender MA, Fu AZ, Wilding JP, Khunti K, Holl RW, et al. Lower risk of heart failure and death in patients initiated on sodium-glucose cotransporter-2 inhibitors versus other glucose-lowering drugs: The CVD-REAL study (comparative effectiveness of cardiovascular outcomes in new users of sodium-glucose cotransporter-2 inhibitors). Circulation. 2017 Jul 18;136(3):249-259.

[Ref. 5.4: 04YCLB] 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.

[Ref. 5.4: 052TMC] 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.

[Ref. 5.4: 052TNG] 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.

[Ref. 5.4: 052TS6] 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.

[Ref. 5.4: 052V37] 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.

[Ref. 5.4: 052VKQ] 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.

[Ref. 5.4: 052W62] 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.

[Ref. 5.4: 052W8Y] 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.

[Ref. 5.4: 052WPT] 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.

[Ref. 5.4: 052WPW] 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.

[Ref. 5.4: 052WSP] 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.

[Ref. 5.4: 052WWP] 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.

[Ref. 5.4: 052WY2] 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.

[Ref. 5.4: 052X67] 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.

[Ref. 5.4: 052X99] PopMedNet [Internet]. Boston (MA): Department of Population Medicine; PopMedNet; [cited 2018 Oct 29]. Available from: https://www.popmednet.org/.

[Ref. 5.4: 052XSG] 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.

[Ref. 5.4: 052XV9] 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.

[Ref. 5.4: 052Y0K] 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.

[Ref. 5.4: 052Y70] 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.

[Ref. 5.4: 0576MY] 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.

[Ref. 5.4: 05LSH3] Douros A, Lix LM, Fralick M, Dell'Aniello S, Shah BR, Ronksley PE, et al. Sodium-glucose cotransporter-2 inhibitors and the risk for diabetic ketoacidosis: a multicenter cohort study. Ann Intern Med. 2020 Sep 15;173(6):417-25.

[Ref. 5.4: 05LSHM] Patorno E, Pawar A, Franklin JM, Najafzadeh M, Deruaz-Luyet A, Brodovicz KG, et al. Empagliflozin and the risk of heart failure hospitalization in routine clinical care: a first analysis from the EMPRISE study. Circulation. 2019 Jun 18;139(25):2822-30.

[Ref. 5.4: 05LSHP] Ueda P, Svanstrom H, Melbye M, Eliasson B, Svensson AM, Franzen S, et al. Sodium glucose cotransporter 2 inhibitors and risk of serious adverse events: nationwide register based cohort study. BMJ. 2018;363:k4365.

[Ref. 5.4: 05LSHQ] Wang L, Voss EA, Weaver J, Hester L, Yuan Z, DeFalco F, et al. Diabetic ketoacidosis in patients with type 2 diabetes treated with sodium glucose co-transporter 2 inhibitors versus other antihyperglycemic agents: an observational study of four US administrative claims databases. Pharmacoepidemiol Drug Saf. 2019;28:1620-8.

[Ref. 5.4: 07WXVY] Birkeland KI, Jorgensen ME, Carstensen B, Persson F, Gulseth HL, Thuresson M, et al. Cardiovascular mortality and morbidity in patients with type 2 diabetes following initiation of sodium-glucose co-transporter-2 inhibitors versus other glucose- lowering drugs (CVD-REAL Nordic): a multinational observational analysis. Lancet Diabetes Endocrinol. 2017 Sep;5:709-17.

[Ref. 5.4: 07WYML] Chen YW, Voelker J, Tunceli O, Pericone CD, Bookhart B, Durkin M. Real-world comparison of hospitalization costs for heart failure in type 2 diabetes mellitus patients with established cardiovascular disease treated with canagliflozin versus other antihyperglycemic agents. J Med Econ. 2020;23(4):401-6.

[Ref. 5.4: 07WYN8] Dawwas GK, Smith SM, Park H. Cardiovascular outcomes of sodium glucose cotransporter-2 inhibitors in patients with type 2 diabetes. Diabetes Obes Metab. 2019;21:28-36.

[Ref. 5.4: 07WYNH] Filion KB, Lix LM, Yu OHY, Dell'Aniello S, Douros A, Shah BR, et al. Sodium glucose cotransporter 2 inhibitors and risk of major adverse cardiovascular events: multi-database retrospective cohort study. BMJ. 2020;370:m3342.

[Ref. 5.4: 07WYRJ] Huang TY, Welch EC, Shinde MU, Platt RW, Filion KB, Azoulay L, et al. Reproducing protocol-based studies using parameterizable tools-comparison of analytic approaches used by two medical product surveillance networks. Clin Pharmacol Ther. 2020 Apr;107(4):966-77.

[Ref. 5.4: 07WZGS] Patorno E, Najafzadeh M, Pawar A, Franklin JM, Deruaz-Luyet A, Brodovicz KG, et al. The EMPagliflozin compaRative effectIveness and SafEty (EMPRISE) study programme: design and exposure accrual for an evaluation of empagliflozin in routine clinical care. Endocrinol Diabetes Metab. 2020;3:e00103.

[Ref. 5.4: 07WZK0] Persson F, Nystrom T, Jorgensen ME, Carstensen B, Gulseth HL, Thuresson M, et al. Dapagliflozin is associated with lower risk of cardiovascular events and all-cause mortality in people with type 2 diabetes (CVD-REAL Nordic) when compared with dipeptidyl peptidase-4 inhibitor therapy: a multinational observational study. Diabetes Obes Metab. 2018;20:344-51.

[Ref. 5.4: 07WZM0] Ryan PB, Buse JB, Schuemie MJ, DeFalco F, Yuan Z, Stang PE, et al. Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: a real-world meta-analysis of 4 observational databases (OBSERVE-4D). Diabetes Obes Metab. 2018;20:2585-97.

[Ref. 5.4: 07WZPC] Welch EC, Woodworth TS, Menzin TJ, Jane Huang TY. Evaluation of test data in distributed research networks: a Sentinel system example. Presented at: 33rd International Conference on Pharmacoepidemiology (ICPE) and Therapeutic Risk Management; 2017 Aug 26-30; Montreal (Canada).

[Ref. 5.4: 07X0GP] Sentinel Operations Center [Internet]. SAS code for transforming the IBM MarketScan research databases (MarketScan) into the Sentinel Common Data Model. Boston (MA): Sentinel Initiative; 2020 Dec 31 [cited 2021 Nov 3]; [about 2 screens]. Available from: https://www.sentinelinitiative.org/methods-data-tools/sentinel- common-data-model/sas-code-transforming-ibm-marketscan- research.

[Ref. 5.4: 07X0GQ] Sentinel Operations Center [Internet]. Who is involved. Boston (MA): Sentinel Initiative; [cited 2021 Nov 3]; [about 10 screens]. Available from: https://www.sentinelinitiative.org/about/who-involved.

[Ref. 5.4: 07XC64] Centers for Medicare and Medicaid Services [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services. Medicare Part D Spending by Drug; 2019 [cited 2021 Nov 19]; [about 2 screens]. Available from: https://data.cms.gov/summary-statistics-on-use-and- payments/medicare-medicaid-spending-by-drug/medicare-part- d-spending-by-drug/data/2019.

ANNEX LIST

Annex 1 Study protocol

(05BK7R)