Janssen Research & Development *

Statistical Analysis Plan

Comparison of Canagliflozin vs. Alternative Antihyperglycemic Treatments on Risk of Heart Failure Hospitalization and Amputation for Patients with Type 2 Diabetes Mellitus and the Subpopulation with Established Cardiovascular Disease

Protocol RRA-20250 AMENDMENT INT-4

JNJ-28431754 (canagliflozin)

Status:	Approved
Date:	13 February 2018
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-ERI-149847048

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

ABBREVIATIONS **3**

1. INTRODUCTION **4**
1.1. Study Objectives 4
1.2. Study Design 4
1.3. Statistical Hypotheses for Study Objectives 7
1.4. Sample Size 7

2. GENERAL ANALYSIS DEFINITIONS **8**
2.1. Analysis Sets 8
2.2. Study Time to Health Outcome 9

3. SUBJECT INFORMATION **9**
3.1. Demographics and Baseline Characteristics 9
3.2. Extent of Exposure 10

4. OUTCOME OF INTEREST **11**
4.1. Analysis Specifications 11
4.1.1. Level of Significance 11
4.1.2. Data Handling Rules 11
4.2. Propensity Scores 12
4.3. Primary Health Outcome(s) 13
4.3.1. Definition 13
4.3.2. Analysis Methods 13
4.4. Negative Control Outcome 14

5. REFERENCES **16**

ATTACHMENT 1: DATA SOURCE **17**

ATTACHMENT 2: COHORT FOR SENSITIVITY ANALYSIS **22**

ATTACHMENT 3: ADDITIONAL COMPARISON FOR SENSITIVITY ANALYSIS **24**

ATTACHMENT 4: PROJECT NUMBER OF PATIENTS IN THE COHORTS SERVED FOR SENSITIVITY ANALYSIS **27**

ABBREVIATIONS

AHA	antihyperglycemic agents
AUC	area under the curve
CCAE	Truven Health MarketScan™ Commercial Claims and Encounters Database
CDM	Common Data Model
CVD	cardiovascular disease
DPP-4	dipeptidyl peptidase-4
ETL	extract, transform, and load
GLP-1	glucagon-like peptide-1
HHF	hospitalization for heart failure
ITT	Intent-to-Treat
LASSO	Least Absolute Shrinkage and Selection Operator
MDCD	Truven Health MarketScan™ Multi-state Medicaid Database
MDCR	Truven Health MarketScan™ Medicare Supplemental and Coordination of Benefits Database
NDC	National Drug Code
OMOP	Observational Medical Outcomes Partnership
Optum	OptumInsight's de-identified Clinformatics™ Datamart, Extended – Date of Death
ROC	Receiver Operating Characteristic
SGLT2i	sodium-glucose co-transporter 2 inhibitors
SU	sulfonylureas
T2DM	type 2 diabetes mellitus
TZD	thiazolidinediones

1. INTRODUCTION

Patients with type 2 diabetes mellitus (T2DM) have many treatment options to manage their disease, including metformin, sodium-glucose co-transporter 2 inhibitors (SGLT2i), dipeptidyl peptidase-4 (DPP-4) inhibitors, glucagon-like peptide-1 (GLP-1) agonist, thiazolidinediones (TZD), sulfonylureas (SU), insulin or other antihyperglycemic agents (AHA). Canagliflozin is one of the drugs in the SGLT2i class. Clinical trials have suggested canagliflozin may reduce the risk of major adverse cardiovascular events, relative to placebo. These findings are largely consistent with results from a placebo-controlled clinical trial for empagliflozin, another drug in the SGLT2i class [1]. However, there have been no studies that have provided a direct comparison among the drugs in the SGLT2i class to determine if the safety and effectiveness profile of each drug follow a common class effect. An observational database study was conducted to compare all SGLT2i drugs with other glucose-lowering drugs for hospitalization of heart failure and all-cause mortality, and showed lower risk of both outcomes associated with SGLT2i [2]. The study did not provide molecule-specific effect estimates, nor did it provide real-world evidence about any of the other adverse outcomes that have been observed in SGLT2i trials, such as amputation.

This study aims to examine the comparative effects of canagliflozin vs. other SGLT2 inhibitors and other alternative treatments for T2DM, overall and within the subpopulation of patients with established cardiovascular diseases, across 4 US administrative claims databases.

1.1. Study Objectives

The objective of the study is to estimate the incidence of health outcomes of

- hospitalization for heart failure (HHF),
- below knee lower extremity amputation,

in patients diagnosed with T2DM who are newly exposed to SGLT2i, including canagliflozin, or other AHA's, and to estimate the hazard ratio of the outcomes across the treatment groups.

In addition, the estimation and the comparisons will be conducted in the subgroup of the patients with established cardiovascular disease (CVD).

1.2. Study Design

This observational study is based on four administrative claims databases in the US:

1. Truven Health MarketScan™ Commercial Claims and Encounters Database (CCAE)
2. Truven Health MarketScan™ Medicare Supplemental and Coordination of Benefits Database (MDCR)
3. Truven Health MarketScan™ Multi-state Medicaid Database (MDCD)

4. OptumInsight's de-identified Clinformatics™ Datamart, Extended – Date of Death (Optum)

A description of these databases can be referred in Attachment 1.

Subsets of patients in these databases who satisfied one or more inclusion criteria for a duration of time are used to construct 12 exposure cohorts for clinical characterization, and population-level effect estimation. The 12 cohorts are defined by patient population as follows.

T2DM Patients		T2DM Patients with established CV Disease	
Cohort 1	New user of canagliflozin	Cohort 2	New user of canagliflozin
Cohort 3	New user of empagliflozin	Cohort 4	New user of empagliflozin
Cohort 5	New user of dapagliflozin	Cohort 6	New user of dapagliflozin
Cohort 7	New user of empagliflozin or dapagliflozin	Cohort 8	New user of empagliflozin or dapagliflozin
Cohort 9	New users of any DPP-4 inhibitor, GLP-1 agonist, or other AHA	Cohort 10	New users of any DPP-4 inhibitor, GLP-1 agonist, or other AHA
Cohort 11	New users of any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA	Cohort 12	New users of any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA

The specific ingredients included in the above drug classes are listed below. Note, metformin is not included in any cohort definition, as it is often prescribed as first line of therapy, and therefore not expected to be meaningfully comparable with canagliflozin exposure. Metformin-containing combination products are included in all cohort definitions.

- DPP-4 inhibitors: alogliptin, linagliptin, saxagliptin, sitagliptin, vildagliptin
- GLP-1 agonists: albiglutide, dulaglutide, exenatide, liraglutide, lixisenatide
- TZDs: pioglitazone, rosiglitazone, troglitazone
- Sulfonylureas: glipizide, glyburide, glimepiride, chlorpropamide, tolazamide, tolbutamide, acetohexamide
- Other select AHAs: acarbose, bromocriptine, miglitol, nateglinide, repaglinide

To address the potential impact from time-dependent complexity of non-metformin AHA switching during the 365 days prior to the index date of cohort defining drug, the cohorts above are further stratified by the AHA switching history. They are served for the purpose of sensitivity analysis. A list of these additional cohorts is presented in Attachment 2.

Generally, each cohort has an index cohort entry event of an initial drug exposure and satisfies the following inclusion criteria:

- First exposure to the drug(s) specified in the cohort definition (Index Day)
- Exposure start is between 1 April 2013 and 15 May 2017.
- At least 365 days of continuous observation time prior to Index Day
- At least 1 condition occurrence of 'Type II diabetes' any time in the prior continuous observation time (which is at least 365 days long) before or on the Index Day
- Exactly 0 condition occurrences of 'Type I diabetes' any time in the prior continuous observation time (which is at least 365 days long) before or on the Index Day
- Exactly 0 condition occurrences of 'Secondary diabetes' any time in the prior continuous observation time (which is at least 365 days long) before or on the Index Day
- Note, there is no patient age restriction, and there is no restriction on the healthcare setting(s) of the study population for the exposure cohorts.

Each cohort with 'established cardiovascular disease' additionally satisfies the following inclusion criteria:

- at least 1 occurrence of 'conditions indicating established cardiovascular disease' on or any time in the prior continuous observation time (which is at least 365 days long) prior to the Index Day (defined as diagnosis codes classified as 'arteriosclerotic vascular disease', 'cerebrovascular disease', 'late effects of cerebrovascular disease', 'myocardial infarction', 'angina pectoris', 'precerebral arterial occlusion', 'chronic ischemic heart disease', 'acute ischemic heart disease', 'intracranial hemorrhage', 'subdural hemorrhage', 'cerebral hemorrhage', subarachnoid hemorrhage', 'coronary occlusion', 'cerebral artery occlusion', 'peripheral vascular disease', or 'peripheral circulatory disorder associated with diabetes mellitus'), or
- at least 1 occurrence of 'procedures indicated established cardiovascular disease' on or any time in the prior continuous observation time (which is at least 365 days long) prior to the Index Day (defined by all procedure codes classified as 'operative procedure on coronary artery' or 'operative procedure on artery of extremity', and mainly comprised of 'percutaneous transluminal coronary angioplasty', 'coronary artery bypass graft', or 'revascularization').

Each cohort has cohort exit criteria based on the end of the period of persistent exposure to the cohort-defining drug, using a 30-day persistence gap window. Additionally, cohort exit is censored if a drug era started for a non-metformin AHA other than the cohort-defining drug(s) is observed after cohort entry.

1.3. Statistical Hypotheses for Study Objectives

The objective of the study is to estimate the risk of developing the health outcomes of HHF, and amputation in the 12 exposure cohorts (as defined in Section 1.2). For comparison purpose, hazard ratio with 95% confidence interval will be estimated. The following 14 comparisons by patient population will be made.

T2DM Patients	T2DM Patients with established CVD	Associated Treatment
Cohort 1 vs. Cohort 3	Cohort 2 vs. Cohort 4	canagliflozin vs. empagliflozin
Cohort 1 vs. Cohort 5	Cohort 2 vs. Cohort 6	canagliflozin vs. dapagliflozin
Cohort 1 vs. Cohort 7	Cohort 2 vs. Cohort 8	canagliflozin vs. (empagliflozin or dapagliflozin)
Cohort 1 vs. Cohort 9	Cohort 2 vs. Cohort 10	canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)
Cohort 1 vs. Cohort 11	Cohort 2 vs. Cohort 12	canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA)
Cohort 7 vs. Cohort 9	Cohort 8 vs. Cohort 10	(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)
Cohort 7 vs. Cohort 11	Cohort 8 vs. Cohort 12	(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin or other AHA)

To support the robustness the above comparisons, sensitivity analysis will be performed where additional comparisons stratified by the exposure to the background AHA will be conducted. A list of the additional comparisons is presented in Attachment 3.

1.4. Sample Size

The sample size of the 12 cohorts across the four databases is listed below. These patient counts represent the initial population, prior to statistical adjustment.

			Database			
New User of	Patient	Cohort	CCAE	MDCD	MDCR	Optum
canagliflozin	T2DM	1	78,541	7,497	11,678	45,321
	T2DM w CVD	2	19,440	2,554	6,319	14,802
empagliflozin	T2DM	3	35,586	549	2,956	14,372
	T2DM w CVD	4	9,415	187	1,649	5,138
dapagliflozin	T2DM	5	48,226	1,104	2,958	10,098
	T2DM w CVD	6	11,193	401	1,513	2,964
empagliflozin or dapagliflozin	T2DM	7	80,409	1,630	5,717	23,297
	T2DM w CVD	8	19,696	578	3,055	7,727

any DPP4i, GLP-1 agonist, or other AHA	T2DM	9	173,694	23,613	38,679	112,161
	T2DM w CVD	10	40,238	8,283	22,835	44,997
any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	T2DM	11	223,202	43,407	45,333	152,011
	T2DM w CVD	12	40,238	8,283	22,835	44,997

The projected number of patients in the cohorts served for sensitivity analysis purpose are presented in Attachment 4.

For the descriptive analyses as part of clinical characterization, these cohorts should provide sufficient sample to adequately summarize and compare baseline characteristics within the cohorts. For population-level effect estimation, where our aim is to produce an unbiased estimate of the average treatment effect, the precision we will achieve will vary by the incidence rate of each outcome.

2. GENERAL ANALYSIS DEFINITIONS

All analyses will be performed independently within each of the four databases to produce four source-specific results for each analysis. No patient-level data will be pooled across the databases for any analysis, in part to avoid the potential risk of 'double-counting' cases for duplicate patients, i.e., the databases will be kept separate to produce database-specific effect estimates (although we will potentially combine database-specific results meta-analytically). The extent of overlap between patients in these databases is unknown, and because all databases have been de-identified, it is not possible or allowable to attempt to link patients across databases. CCAE represents a privately-insured population of employees and their dependents, making it unlikely they would also qualify for a state Medicaid program, as represented in MDCD. A small proportion of patients do transition from CCAE to MDCR when they retire from their employee and purchase supplemental Medicare insurance, but in these cases, the patients would not be duplicated at the same period of time, but rather their person-time would be split between the two databases. It would be similarly unlikely that a person opting to purchase supplemental Medicare insurance qualifies for state Medicaid benefits. Persons with data in any of the Truven datasets may potentially be captured in Optum if the employer or state programs contained in Truven are administered by plans within the Optum catchment, but employers and plans are de-identified so the extent of overlap cannot be determined.

2.1. Analysis Sets

Analysis sets consist of 2 components: (1) analysis population (refer Section 1.2 for the 12 cohorts), which specifies the subjects included in an analysis; (2) data period, defining the time window during which data will be included in the analysis. The four time periods that will be applied are specified as follows.

1. On-Treatment period: defined as the time from Index Day (refer Section 1.2) to 0 days from cohort end date, where exposure cohort end date was defined as the persistent period of exposure, allowing for 30 day gap between successive exposures until the final exposure record, representing the date the subject expected to finish the supply of the

last drug dispensing per the prescription information. The exposure cohort end date may be censored by the start of a drug era for a non-metformin AHA other than the cohort defining drug(s).

2. Intent-to-Treat (ITT) period: defined as the time from Index Day to the end of the patient's observation period for which the cohort start date occurred.

3. Lagged On-Treatment period: defined as the time from 60 days after the cohort start date to 60 days after the cohort end date, defined as the time from 60 days after the exposure cohort start date to 60 days after the exposure cohort end date, where exposure cohort end date was similarly defined as in 1) above.

4. Modified Intent-to-Treat (MITT) period: defined as the time from 1 day after exposure cohort start date (based on the date of first exposure to the cohort-defining drug(s)) until the end of the patient's observation period for which the exposure cohort start date occurred for those persons who only qualify for the target or comparator cohort, and from 1 day after exposure start date of the first cohort entry until the date of the second cohort entry for those persons who qualify for both the target and comparator cohorts.

2.2. Study Time to Health Outcome

Index Day refers to the start of the use of the agent(s) associated the cohorts. The study time will be in terms of the day relative to Index Day.

3. SUBJECT INFORMATION

3.1. Demographics and Baseline Characteristics

Descriptive analysis of the baseline covariates will be generated to provide a characterization of the 64 exposure cohorts. The following baseline variables will be included.

- Demographics
  - Gender
  - Age group (5-year bands)
  - Initial drug exposure year
  - Initial drug exposure month
- Condition occurrence record for the concept or any its descendants observed during 365d on or prior to cohort index
- Drug exposure record for the concept or any its descendants observed during 365d on or prior to cohort index
- Procedure occurrence record for the concept or any its descendants observed during 365d on or prior to cohort index
- Measurement record for the verbatim concept observed during 365d on or prior to cohort index
- Charlson Index - Romano adaptation, using conditions all time on or prior to cohort index

- Number of distinct conditions observed in 365d on or prior to cohort index (defined as unique SNOMED condition concepts)
- Number of distinct drugs observed in 365d on or prior to cohort index (defined as unique RxNorm ingredient concepts)
- Number of distinct procedures observed in 365d on or prior to cohort index (defined as unique CPT4/HCPCS/ICD9P/ICD10P concepts)
- Number of distinct observations observed in 365d on or prior to cohort index
- Number of distinct measurements observed in 365d on or prior to cohort index (defined as unique LOINC concepts)
- Number of visits observed in 365d on or prior to cohort index
- Number of inpatient visits observed in 365d on or prior to cohort index
- Number of ER visits observed in 365d on or prior to cohort index
- Prior outcome history, observed during all time prior to or on the cohort index, for all study outcomes

An explicit head-to-head comparison between two cohorts of baseline covariates, using standardized difference as a metric to compare individual factors, will be conducted. Covariates with standardized difference > 10% will be highlighted as potential imbalanced confounding factors. This descriptive analysis will be performed for all the 56 comparisons, including the 14 primary comparisons as listed in section 1.3.

3.2. Extent of Exposure

For all exposure cohorts, we derive 'drug eras' as periods of persistent exposure to a particular active ingredient or set of ingredients. 'Drug eras' are derived from verbatim drug exposure records, including outpatient pharmacy dispensing claims records.

From an outpatient pharmacy dispensing claims record, we define the drug exposure start date = dispensing date, and define the exposure end date = dispensing date + days supply. 'Drug eras' consolidate drug exposure records of a particular ingredient for a given patient by combining together records where the drug exposure end date of one record is within a 'persistence window' of the drug exposure start date of a subsequent record. A 'persistence window' of 30 days means that a person will be assumed to be persistently exposed to a drug so long as two successive dispensings have no more than a 30-day gap between the end date of the first and the start date of the second.

A 'surveillance window' may be appended to the end date of each period of persistent exposure. A 'surveillance window' of 30 days means that a person who completed a drug regimen will continue to be followed for 30 days after the last exposure end date.

As an example, if a person had an initial prescription dispensing for canagliflozin on 1Jan2015, which came with 30 days supply, and a second prescription for canagliflozin on 15Feb2015, also with 30 days supply, the data would be processed as follows: the first exposure record would be assigned start date = 1Jan2015 and end date = 31Jan2015 (1Jan2015 + 30d). the second exposure record would be assigned start date = 15Feb2015, end date = 17Mar2015 (15Feb2015+30d). Because the two exposures were less than 30 days apart (15Feb2015 –

31Jan2015 = 16 days), these two exposure records would be consolidated into one drug era record, with start date = 1Jan2015 and end date = 17Mar2015. With an additional 'surveillance window' of 30 days, this drug era record would be expanded to have start date = 1Jan2015 and end date = 17Apr2015 (17Mar2015 + 30d). The drug era logic used here does not account for stockpiling if subsequent prescriptions overlap in time, nor does it account for dose tailoring, since the period of exposure is defined at the ingredient level.

4. OUTCOME OF INTEREST

For each of the following health outcome, the time to the first event from the new use of the cohort medication will be formulated to allow for clinical characterization and population-level effect estimation.

- Hospitalization of heart failure
- Below Knee Lower Extremity Amputation events

4.1. Analysis Specifications

4.1.1. Level of Significance

The objective of the study is to estimate the magnitude of the treatment effect based on the observational data. Nominal 95% confidence interval will be derived for the incidence and the hazard ratio. No multiplicity adjustment will be performed.

4.1.2. Data Handling Rules

Patients must have at least 1 day of continuous observation after Index Day.

For all analyses, a qualifying person from an exposure cohort will contribute time-to-outcome from time-at-risk start date until some censoring event occurs. For the 'on treatment' and 'intent-to-treat' analyses, time-at-risk start date is defined as the date of first exposure to the particular drug(s) in the database, where the exposure start is between 1April2013 and 15May2017; for the 'on treatment with delayed induction and delayed lag', the time-at-risk start date is defined as 60 days after the date of first exposure to the particular drug(s) in the database.

The time-to-event of outcome among patients in the cohorts is determined by calculating the number of days from the start of the time-at-risk window, 1 day from cohort start date, until the earliest event among 1) occurrence of the outcome post-index before 0 days from cohort end date, 2) the end of the time-at-risk window, 0 days from cohort end date, and 3) the end of the observation period that spans the time-at-risk start.

For the 'on treatment' period, the censoring event is the earliest of: 1) cohort end date, as defined by the end of the persistent period of exposure censored at the time of a drug era start for a non-metformin AHA other than the cohort-defining drug, and 2) observation period end date (which is defined by end of continuous insurance eligibility or death).

For the 'intent-to-treat' period, the censoring event is the observation period end date.

For the 'on treatment with delayed induction and delayed lag' period, the censoring event is the earliest of: 1) 60 days after the cohort end date, and 2) observation period end date.

For the 'modified intent-to-treat' period, the censoring event is the earliest of 1) the observation period end date, and 2) the start of the cohort entry for the second of the target and comparator cohorts for those patients qualifying for both cohorts.

In all circumstances, a patient is not censored by experiencing hospitalization for other reasons.

For each of the 14 pairwise comparisons listed in section 1.3 and all sensitivity analyses, patients in the two comparative cohorts must have at least 1 day of continuous observation after Index Day. The study time is restricted to the calendar time when both treatments were observed in the database. Patients are excluded from consideration if they qualified for both cohorts at any time in their record.

4.2. Propensity Scores

Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between the cohorts at comparison in baseline covariates. The propensity score is the probability of a patient being classified in the cohorts, given a set of observed covariates. The baseline covariates used to fit the propensity score model are listed in section 3.1.

The propensity score is estimated for each patient, using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation, using a starting variance of 0.01 and a tolerance of 2e-7. Covariates to be used in the propensity score model are listed in section 3.1, as was applied in prior studies [3].

All covariates that occur in fewer than 10 persons between the two cohorts at comparison combined will be excluded prior to model fitting.

Covariate balance will be summarized in tabular form by showing the mean value for all baseline covariates in the cohorts, with the associated standardized mean difference computed for each covariate.

Once the propensity score model is fit, we will plot the propensity score distribution of the target and comparator cohorts to evaluate the comparability of the two cohorts. The plot will be scaled to the preference score, normalizing for any imbalance in cohort size. The area under the Receiver Operating Characteristic (ROC) curve (AUC) will be reported. The covariates selected within the propensity score model, with associated coefficients will also be reported.

A plot showing the propensity score distributions for both cohorts before and after matching will be provided. Covariate balance will be evaluated by plotting the standardized mean difference of each covariate before propensity score stratification against the standardized mean difference for each covariate after propensity score stratification.

An attrition diagram will be provided to detail the loss of patients from the original target cohort and comparator cohort to the subpopulations that remain after all design considerations have been applied.

The cohorts at comparison will be matched using variable ratio matching on the propensity, with a maximum ratio of 100 and using a caliper of 0.25 times the standard deviation of the propensity score distribution. The final outcome model will apply a Cox proportional hazard model conditioned on the matched sets.

4.3. Primary Health Outcome(s)

4.3.1. Definition

For each of the following health outcome, the time to the first event from the new use of the cohort medication will be formulated to allow for clinical characterization and population-level effect estimation.

- Hospitalization of heart failure
- Below Knee Lower Extremity Amputation events

4.3.2. Analysis Methods

Each analysis will be performed against all 4 databases.

Incidence will be computed for the outcomes across all 12 target cohorts. The number of persons and person-years-at-risk for each exposure-outcome-database combination will be provided, along with the number of events during the time-at-risk periods. Incidence proportions and incidence rates will be computed. Incidence summary will be stratified by gender.

All clinical characterization analyses involve direct observation of the experience of patients, which can provide context about the real-world patterns of utilization in different populations, but cannot be used for causal inference or to draw comparative conclusions about the effects of any treatment.

For each of the 14 pairwise comparisons listed in section 1.3, a final outcome model will be fit for each outcome. The model will be a Cox proportional hazards model conditional on the matched set from the propensity score analysis. The hazards ratio and associated 95% confidence interval will be derived from the model. The number of persons, amount of time-at-risk, and number of outcomes in each cohort will also be reported.

Heterogeneity of the hazards ratios will be estimated across the 4 databases, using I2 as a metric [4]. If there is sufficient homogeneity across sources (I2<40%), [5], source-specific estimates will be pooled through a DerSimonian-Laird random effect meta-analysis. Source-specific and composite estimates will be presented in a forest plot.

Subgroup analysis will also be conducted for patients with no history of the outcome,

A collection of negative control outcomes, listed in section 4.4, as a diagnostic tool to quantify residual bias and empirically calibrate statistics generated during the analysis.

4.4. Negative Control Outcome

Negative control outcomes, outcomes known not to be causally associated with any of the exposure cohorts [6], and are to be used for empirical calibration [7]. The same analysis performed for each health outcome in each cohort comparison will also be performed for each negative control outcome. Because of the a priori assertion of no effect, we assume the true relative risk for each negative control outcome is 1, and the distance between RR=1 and the observed effect estimate will be classified as error. The sample of negative controls will therefore be used to construct an empirical distribution of the error distribution, which will be used to calibrate the p-values from the unknown outcomes of interest. Analyses performed for negative control outcomes should not be considered formal hypothesis tests, and should not be included in any multiplicity adjustment. The negative control outcomes used in this study in the table below.

These negative control outcomes were selected on the basis of being unlisted as potential effects on the product labels of any cohort-defining drugs, not having an observed statistical association from disproportionality analysis of FAERS, and having no identified published literature linking the outcome to any of the exposures of interest. The *a priori* assumption that these outcomes have no causal relation with the exposure is inherently untestable. The use of a large sample of negative controls also helps to mitigate the risk that any particular negative control has a non-null effect that would bias the error distribution estimate. Negative control outcomes may have a myriad of sources of systematic error, including measurement error due to outcome misclassification for coding imprecision, and unobserved confounding with the exposures of interest; the role of empirical calibration is to use the estimated systematic error from the negative control sample as a proxy for the potential residual bias that may exist for the unknown outcome of interest.

concept_id	concept_name
376707	Acute conjunctivitis
433753	Alcohol abuse
257007	Allergic rhinitis
442077	Anxiety disorder
436665	Bipolar disorder
380094	Carpal tunnel syndrome
255573	Chronic obstructive lung disease
257012	Chronic sinusitis
443617	Conduct disorder
134438	Contact dermatitis
78619	Contusion of knee
378752	Corneal opacity
137063	Corns and callus
133228	Dental caries
134681	Diffuse spasm of esophagus

432251	Disease caused by parasite
378161	Disorder of ear
139057	Disorder of oral soft tissues
31057	Disorder of pharynx
138225	Disorder of sebaceous gland
440329	Herpes zoster without complication
441788	Human papilloma virus infection
140673	Hypothyroidism
374375	Impacted cerumen
139099	Ingrowing nail
436962	Insomnia
201322	Internal hemorrhoids without complication
132466	Lumbar sprain
255891	Lupus erythematosus
444100	Mood disorder
440374	Obsessive-compulsive disorder
380733	Otalgia
372328	Otitis media
4002650	Plantar fasciitis
373478	Presbyopia
436073	Psychotic disorder
438688	Sarcoidosis
432597	Schizoaffective schizophrenia
435783	Schizophrenia
372409	Sciatica
73562	Solitary sacroiliitis
133141	Tinea pedis
436070	Vitamin D deficiency
434008	White blood cell disorder

5. REFERENCES

1. Zinman, B., et al., Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med, 2015. 373(22): p. 2117-28.
2. Kosiborod, M., et al., Lower Risk of Heart Failure and Death in Patients Initiated on SGLT-2 Inhibitors Versus Other Glucose-Lowering Drugs: The CVD-REAL Study. Circulation, 2017.
3. Wang, Y., 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: p. 83-90.
4. Higgins, J.P., et al., Measuring inconsistency in meta-analyses. BMJ, 2003. 327(7414): p. 557-60.
5. Cochrane reviews. 9.5.2 Identifying and measuring heterogeneity. Cochrane Handbook for Systematic Reviews of Interventions March 2011; Available from: http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm.
6. Ryan, P.B., et al., Defining a reference set to support methodological research in drug safety. Drug Saf, 2013. 36 Suppl 1: p. S33-47.
7. Schuemie, M.J., et al., Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med, 2014. 33(2): p. 209-18.
8. Committee for Medicinal Products for Human Use. Guideline on the Investigation of Bioequivalence. January 2010.
9. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations, March 2003.
10. Shankar G., Arkin S., Cocea L., Devanarayan V., Kirshner S., Kromminga A., quarmby V., Richardws,S, Schnieder C.K. Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides – Harmonized Terminology and Tactical Recommendations. The AAPS Journal. e-ISSn 1550-7416. AAPS J DOI 10.1208/s12248-014-9599-2

ATTACHMENT 1: DATA SOURCE

This study will be conducted in four observational health databases:

1. Truven Health MarketScan™ Commercial Claims and Encounters Database (CCAE)
2. Truven Health MarketScan™ Medicare Supplemental and Coordination of Benefits Database (MDCR)
3. Truven Health MarketScan™ Multi-state Medicaid Database (MDCD)
4. OptumInsight's de-identified Clinformatics™ Datamart, Extended – Date of Death (Optum)

Each database is described below:

1. Truven Health MarketScan™ Commercial Claims and Encounters Database (CCAE)

   CCAE is a medical and drug insurance claims database of unique de-identified patients that include active employees, early retirees, COBRA continuers, and their dependents insured by employer-sponsored plans. The database contains inpatient admission records, outpatient services, prescription drugs, populations, eligibility status, and costs of services.

   As of 3 January 2018 (data lock on version CDM_Truven_CCAE_v656),, the latest available version of CCAE contains more than 125 million patients with observations from January 2000 through July 2017. The major data elements contained within this database that will be used within this analysis include: outpatient pharmacy dispensing claims – which provide dispensing date, National Drug Code (NDC) to identify the prescription product, days supply, and quantity; inpatient and outpatient medical claims – which provide service date, procedure codes (coded in CPT-4, HCPCS, ICD-9-CM or ICD-10-PCS), and diagnosis codes (coded in ICD-9-CM or ICD-10-CM); and enrollment records – which provide the dates of insurance eligibility, as well as year of birth and sex. The database does contain selected laboratory test results (those sent to a contracted third-party laboratory service provider) for a non-random sample of the population, but this data will not be used within these analyses. As an administrative claims database, records may exist for each reimbursement transaction, such that it is possible for the database to document multiple drugs, diagnoses, and procedures during inpatient and outpatient medical services for a given person over time, during the enrollment period. Claims submitted using the CMS-1500 form are limited to 12 diagnoses per claim. A person may experience discontinuation in their eligibility within the same database and the same database may contain observations during multiple discontinuous periods of time for a given individual.

   The following limitations of Truven CCAE should be noted:

- The commercially insured patients represent a higher socioeconomic status than the overall US population.
- Data are based on financial claims filed for reimbursement; disease coding may reflect financial incentives for reimbursement rather than clinically and systematically verified definitions.
- Prescriptions are those filled at outpatient pharmacies, not those prescribed or administered within inpatient services. The extent to which prescribed records went unfulfilled is not known. It is also not known whether medications were actually taken as directed, although repeated dispensing of the same drug would suggest that this is the case.
- There is a data lag; Truven only sends records that are 100% paid, which can take about 6 months after year end.

2. Truven Health MarketScan™ Medicare Supplemental and Coordination of Benefits Database (MDCR)

   MDCR is an administrative health claims database for Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans (predominantly fee-for-service plans). Only plans where both the Medicare-paid amounts and the employer-paid amounts were available and evident on the claims were selected for this database. MDCR captures person-specific clinical utilization, expenditures, and enrollment across inpatient, outpatient, and prescription drug. It also includes results for outpatient lab tests processed by large national lab vendors.

   As of 3 January 2018 (data lock on version CDM_Truven_MDCR_v657), the latest available version of MDCR contains more than 9.8 million patients with observations from January 2000 through July 2017. The major data elements contained within this database that will be used within this analysis include: outpatient pharmacy dispensing claims – which provide dispensing date, National Drug Code (NDC) to identify the prescription product, days supply, and quantity; inpatient and outpatient medical claims – which provide service date, procedure codes (coded in CPT-4, HCPCS, ICD-9-CM or ICD-10-PCS), and diagnosis codes (coded in ICD-9-CM or ICD-10-CM); and enrollment records – which provide the dates of insurance eligibility, as well as year of birth and sex. The database does contain selected laboratory test results (those sent to a contracted third-party laboratory service provider) for a non-random sample of the population, but this data will not be used within these analyses. As an administrative claims database, records may exist for each reimbursement transaction, such that it is possible for the database to document multiple drugs, diagnoses, and procedures during inpatient and outpatient medical services for a given person over time, during the enrollment period. Claims submitted using the CMS-1500 form are limited to 12 diagnoses per claim. A person may experience discontinuation in their eligibility and the database may contain observations during multiple discontinuous periods of time for a given individual.

The following limitations of MDCR should be noted:

- The commercially insured patients represent a higher socioeconomic status than the overall Medicare population.
- Exact birth date is not available, only year of birth.
- Data based on financial claims filed for reimbursement, disease coding may reflect financial incentives for reimbursement rather than clinically and systematically verified definitions.
- Prescriptions are those filled, not those prescribed. We do not know the universe of prescribed records that went unfulfilled.
- There is data lag, MarketScan only sends records that are 100% paid, which can take about 6 months after year end.

3. Truven Health MarketScan™ Multi-state Medicaid Database (MDCD)

MDCD contains the pooled healthcare experience of Medicaid enrollees from multiple states. MDCD includes medical claims from inpatient and outpatient services and prescription drug claims, as well as information on enrollment, long-term care, and other medical care.

As of 3 January 2018 (data lock on version CDM_Truven_MDCD_v635),, the latest available version of MDCD contains more than 25.5 million patients with observations from February 2006 through December 2016. The major data elements contained within this database that will be used within this analysis include: outpatient pharmacy dispensing claims – which provide dispensing date, National Drug Code (NDC) to identify the prescription product, days supply, and quantity; inpatient and outpatient medical claims – which provide service date, procedure codes (coded in CPT-4, HCPCS, ICD-9-CM or ICD-10-PCS), and diagnosis codes (coded in ICD-9-CM or ICD-10-CM); and enrollment records – which provide the dates of insurance eligibility, as well as year of birth and sex. The database does not contain any laboratory test results. As an administrative claims database, records may exist for each reimbursement transaction, such that it is possible for the database to document multiple drugs, diagnoses, and procedures during inpatient and outpatient medical services for a given person over time, during the enrollment period. Claims submitted using the CMS-1500 form are limited to 12 diagnoses per claim. A person may experience discontinuation in their eligibility and the database may contain observations during multiple discontinuous periods of time for a given individual.

The following limitations of MDCD should be noted:

- No state information is available.
- Exact birth date is not available, only year of birth.
- Lab tests processed by large national lab vendors are not available for MDCD patients.
- Members eligible for Medicare may have incomplete data.
- Data based on financial claims filed for reimbursement, disease coding may reflect financial incentives for reimbursement rather than clinically and systematically verified definitions.
- Prescriptions are those filled, not those prescribed. We do not know the universe of prescribed records that went unfulfilled.

4. OptumInsight's de-identified Clinformatics™ Datamart, Extended – Date of Death (Optum)

   Optum is an administrative health claims database for members of United Healthcare, who are fully insured in commercial plans or in administrative services only (ASOs) and commercial Medicare. For each group, only members with both medical and prescription drug coverage are included. Optum captures person-specific clinical utilization, expenditures, and enrollment across inpatient, outpatient, and prescription drug, and includes results for outpatient lab tests processed by large national lab vendors (about 30% of the population).

   As of 3 January 2018 (data lock on version CDM_Optum_Extended_DoD_v654),, the latest available version of Optum contains more than 80million patients with observations from June 2000 through June 2017. The major data elements contained within this database that will be used within this analysis include: outpatient pharmacy dispensing claims – which provide dispensing date, National Drug Code (NDC) to identify the prescription product, days supply, and quantity; inpatient and outpatient medical claims – which provide service date, procedure codes (coded in CPT-4, HCPCS, ICD-9-CM or ICD-10-PCS), and diagnosis codes (coded in ICD-9-CM or ICD-10-CM); and enrollment records – which provide the dates of insurance eligibility, as well as year of birth and sex. The database does contain selected laboratory test results (those sent to a contracted third-party laboratory service provider) for a non-random sample of the population, but this data will not be used within these analyses. As an administrative claims database, records may exist for each reimbursement transaction, such that it is possible for the database to document multiple drugs, diagnoses, and procedures during inpatient and outpatient medical services for a given person over time, during the enrollment period. Claims submitted using the CMS-1500 form are limited to 12 diagnoses per claim. A person may experience discontinuation in their eligibility and the database may contain observations during multiple discontinuous periods of time for a given individual.

The following limitations of Optum should be noted:

- Family enrollment, capitated plan information and exact birth date are not available
- All claims are limited to first 5 diagnostic codes.
- Data based on financial claims filed for reimbursement, disease coding may reflect financial incentives for reimbursement rather than clinically and systematically verified definitions.
- Prescriptions are those filled, not those prescribed. We do not know the universe of prescribed records that went unfulfilled.

All four databases have been standardized into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM),[a] which includes a standard representation of health care experiences (such as information related to drug utilization and condition occurrence), as well as common vocabularies for coding clinical concepts and enables consistent application of analyses across multiple disparate data sources [5, 6]. Complete specifications for the extract, transform, and load (ETL) process for each database is available at: https://github.com/OHDSI/ETL-CDMBuilder.

The standardized vocabularies within OMOP CDM enable disparate source codes to be mapped into a common referent standard for each clinical domain, such as conditions, drugs, and procedures. As an example of particular relevance for this analysis, the standard vocabulary for conditions is SNOMED-CT and source vocabularies that have been mapped into SNOMED-CT include both ICD-9-CM and ICD-10-CM. In the US, with the transition of diagnosis recording from ICD-9-CM to ICD-10-CM occurring in October2015, the use of OMOP standardized vocabularies is quite valuable because it means a common definition for exposures, outcomes, and covariates can be applied to all data, both pre- and post-October 2015 using SNOMED concepts instead of ICD-9-CM or ICD-10-CM codes. The source codes that map to all standard concepts are provided in Annex 1, and the full OMOP standard vocabulary is available for download at: http://athena.ohdsi.org/.

[a] https://github.com/OHDSI/CommonDataModel

ATTACHMENT 2: COHORT FOR SENSITIVITY ANALYSIS

The subjects in the following cohorts are T2DM patients or T2DM patients with established CV disease (CVD) who newly used the cohort-defining drug. The cohorts are stratified by exposure to the specific background AHA during the 365 days prior to the index date of the cohort-defining drug.

Cohort-Defining Drug	Patient	Exposed to Background AHA	Cohort
canagliflozin	T2DM	At least one exposure to non-Metformin AHA	13
		No exposure to non-Metformin AHA	14
		No exposure to empagliflozin	15
		No exposure to dapagliflozin	16
		No exposure to empagliflozin or dapagliflozin	17
		No exposure to any DPP4i, GLP-1 agonist, or other AHA	18
		No exposure to any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	19
	T2DM w CVD	At least one exposure to non-Metformin AHA	20
		No exposure to non-Metformin AHA	21
		No exposure to empagliflozin	22
		No exposure to dapagliflozin	23
		No exposure to empagliflozin or dapagliflozin	24
		No exposure to any DPP4i, GLP-1 agonist, or other AHA	25
		No exposure to any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	26
empagliflozin	T2DM	At least one exposure to non-Metformin AHA	27
		No exposure to non-Metformin AHA	28
		No exposure to canagliflozin	29
	T2DM w CVD	At least one exposure to non-Metformin AHA	30
		No exposure to non-Metformin AHA	31
		No exposure to canagliflozin	32
dapagliflozin	T2DM	At least one exposure to non-Metformin AHA	33
		No exposure to non-Metformin AHA	34
		No exposure to canagliflozin	35
	T2DM w CVD	At least one exposure to non-Metformin AHA	36
		No exposure to non-Metformin AHA	37
		No exposure to canagliflozin	38
empagliflozin or dapagliflozin	T2DM	At least one exposure to non-Metformin AHA	39
		No exposure to non-Metformin AHA	40
		No exposure to canagliflozin	41
		No exposure to any DPP4i, GLP-1 agonist, or other AHA	42
		No exposure to any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	43
	T2DM w CVD	At least one exposure to non-Metformin AHA	44
		No exposure to non-Metformin AHA	45
		No exposure to canagliflozin	46
		No exposure to any DPP4i, GLP-1 agonist, or other AHA	47

		No exposure to any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	48
any DPP4i, GLP-1 agonist, or other AHA	T2DM	At least one exposure to non-Metformin AHA	49
		No exposure to non-Metformin AHA	50
		No exposure to canagliflozin	51
		No exposure to empagliflozin or dapagliflozin	52
	T2DM w CVD	At least one exposure to non-Metformin AHA	53
		No exposure to non-Metformin AHA	54
		No exposure to canagliflozin	55
		No exposure to empagliflozin or dapagliflozin	56
any DPP4i, GLP-1 agonist, TZD, SU, insulin, or other AHA	T2DM	At least one exposure to non-Metformin AHA	57
		No exposure to non-Metformin AHA	58
		No exposure to canagliflozin	59
		No exposure to empagliflozin or dapagliflozin	60
	T2DM w CVD	At least one exposure to non-Metformin AHA	61
		No exposure to non-Metformin AHA	62
		No exposure to canagliflozin	63
		No exposure to empagliflozin or dapagliflozin	64

ATTACHMENT 3: ADDITIONAL COMPARISON FOR SENSITIVITY ANALYSIS

The subjects in the following cohorts are T2DM patients who newly used the cohort-defining drug, their exposure to the background AHA were assessed in 365 days prior to the index date of their cohort-defining drug. The comparisons listed below are served for sensitivity analysis purpose.

Associated Cohort-Defining Drug	T2DM Patients	Exposure to background AHA in 365-day period
canagliflozin vs. empagliflozin	Cohorts 13 vs. 27	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 14 vs. 28	no prior non-metformin AHA exposure in both cohorts
	Cohorts 15 vs. 29	no prior empagliflozin exposure in Cohort 15 and no prior canagliflzoin exposure in Cohort 29
canagliflozin vs. dapagliflozin	Cohorts 13 vs. 33	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 14 vs. 34	no prior non-metformin AHA exposure in both cohorts
	Cohorts 16 vs. 35	no prior dapagliflozin exposure in Cohort 16 and no prior canagliflzoin exposure in Cohort 35
canagliflozin vs. (empagliflozin or dapagliflozin)	Cohorts 13 vs. 39	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 14 vs. 40	no prior non-metformin AHA exposure in both cohorts
	Cohorts 17 vs. 41	no prior empagliflozin or dapagliflozin exposure in Cohort 17 and no prior canagliflzoin exposure in Cohort 41
canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)	Cohorts 13 vs. 49	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 14 vs. 50	no prior non-metformin AHA exposure in both cohorts
	Cohorts 18 vs. 51	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure in Cohort 18 and no prior canagliflzoin exposure in Cohort 51
canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin or other AHA)	Cohorts 13 vs. 57	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 14 vs. 58	no prior non-metformin AHA exposure in both cohorts
	Cohorts 18 vs. 59	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure in Cohort 18 and no prior canagliflzoin exposure in Cohort 59
(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)	Cohorts 39 vs. 49	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 40 vs. 50	no prior non-metformin AHA exposure in both cohorts
	Cohorts 42 vs. 52	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure in Cohort 42 and no prior empagliflozin or dapagliflozin exposure in Cohort 52

Associated Cohort-Defining Drug	T2DM Patients	Exposure to background AHA in 365-day period
(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA)	Cohorts 39 vs. 57	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 40 vs. 58	no prior non-metformin AHA exposure in both cohorts
	Cohorts 43 vs. 60	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure in Cohort 43 and no prior empagliflozin or dapagliflozin exposure in Cohort 60

Similar comparisons are made for the T2DM subjects with established CV disease (CVD) and are listed as follows.

Associated Cohort-Defining Drug	T2DM Patients with established CVD	Exposure to background AHA in 365-day period
canagliflozin vs. empagliflozin	Cohorts 20 vs. 30	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 21 vs. 31	no prior non-metformin AHA exposure in both cohorts
	Cohorts 22 vs. 32	no prior empagliflozin exposure in Cohort 22 and no prior canagliflzoin exposure in Cohort 32
canagliflozin vs. dapagliflozin	Cohorts 20 vs. 36	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 21 vs. 37	no prior non-metformin AHA exposure in both cohorts
	Cohorts 23 vs. 38	no prior dapagliflozin exposure in Cohort 23 and no prior canagliflzoin exposure in Cohort 38
canagliflozin vs. (empagliflozin or dapagliflozin)	Cohorts 20 vs. 44	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 21 vs. 45	no prior non-metformin AHA exposure in both cohorts
	Cohorts 24 vs. 46	no prior empagliflozin or dapagliflozin exposure in Cohort 24 and no prior canagliflzoin exposure in Cohort 46
canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)	Cohorts 20 vs. 53	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 21 vs. 54	no prior non-metformin AHA exposure in both cohorts
	Cohorts 25 vs. 55	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure in Cohort 25 and no prior canagliflzoin exposure in Cohort 55
canagliflozin vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin or other AHA)	Cohorts 20 vs. 61	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 21 vs. 62	no prior non-metformin AHA exposure in both cohorts

Associated Cohort-Defining Drug	T2DM Patients with established CVD	Exposure to background AHA in 365-day period
	Cohorts 26 vs. 63	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure in Cohort 26 and no prior canagliflzoin exposure in Cohort 63
(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, or other AHA)	Cohorts 44 vs. 53	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 45 vs. 54	no prior non-metformin AHA exposure in both cohorts
	Cohorts 47 vs. 56	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure in Cohort 47 and no prior empagliflozin or dapagliflozin exposure in Cohort 56
(empagliflozin or dapagliflozin) vs. (any DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA)	Cohorts 44 vs. 61	at least 1 prior non-metformin AHA exposure in both cohorts
	Cohorts 45 vs. 62	no prior non-metformin AHA exposure in both cohorts
	Cohorts 48 vs. 64	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure in Cohort 48 and no prior empagliflozin or dapagliflozin exposure in Cohort 64

ATTACHMENT 4: PROJECT NUMBER OF PATIENTS IN THE COHORTS SERVED FOR SENSITIVITY ANALYSIS

Cohort-Defining Drug: canagliflozin			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	13	60851	6121	9935	35390
	no prior non-metformin AHA exposure	14	17690	1376	1743	9931
	no prior empagliflozin exposure	15	77869	7465	11617	44930
	no prior dapagliflozin exposure	16	75520	7394	11411	43276
	no prior empagliflozin or dapagliflozin exposure	17	74912	7364	11352	42920
	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure	18	31043	3625	3959	19814
	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure	19	14599	1178	1358	8243
T2DM w CVD	at least 1 prior non-metformin AHA exposure	20	15684	2183	5429	12136
	no prior non-metformin AHA exposure	21	3756	371	890	2666
	no prior empagliflozin exposure	22	19266	2546	6286	14653
	no prior dapagliflozin exposure	23	18745	2508	6172	14203
	no prior empagliflozin or dapagliflozin	24	18585	2500	6141	14071
	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure	25	6982	1140	2073	5971
	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure	26	2902	283	654	2048

Cohort-Defining Drug: empagliflozin			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	27	27747	448	2490	11212
	no prior non-metformin AHA exposure	28	7839	101	466	3160
	no prior canagliflozin exposure	29	27578	461	2444	12757
T2DM w CVD	at least 1 prior non-metformin AHA exposure	30	7648	156	1402	4204
	no prior non-metformin AHA exposure	31	1767	31	247	934
	no prior canagliflozin exposure	32	7257	149	1362	4552

Cohort-Defining Drug: dapagliflozin			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	33	36273	928	2481	7872
	no prior non-metformin AHA exposure	34	11953	176	477	2226
	no prior canagliflozin exposure	35	39024	979	2568	9311
T2DM w CVD	at least 1 prior non-metformin AHA exposure	36	8834	353	1277	2438
	no prior non-metformin AHA exposure	37	2359	48	236	526
	no prior canagliflozin exposure	38	8943	362	1328	2716

Cohort-Defining Drug: empagliflozin or dapagliflozin			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	39	60688	1354	4775	17931
	no prior non-metformin AHA exposure	40	19721	276	942	5366
	no prior canagliflozin exposure	41	64430	1426	4877	21097
	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure	42	34395	704	2031	10179
	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure	43	18006	251	824	4595
T2DM w CVD	at least 1 prior non-metformin AHA exposure	44	15586	499	2573	6271
	no prior non-metformin AHA exposure	45	4110	79	482	1456
	no prior canagliflozin exposure	46	15629	505	2612	6959
	no prior DPP-4 inhibitor, GLP-1 agonist, or other AHA exposure	47	7523	229	1020	3101
	no prior DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure	48	3499	70	394	1177

Cohort-Defining Drug: DPP-4 inhibitor, GLP-1 agonist, or other AHA			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	49	79176	13334	21033	58955
	no prior non-metformin AHA exposure	50	94518	10279	17646	53206
	no prior canagliflozin exposure	51	165243	22789	37738	107577
	no prior empagliflozin or dapagliflozin exposure	52	167004	23459	38334	110192
T2DM w CVD	at least 1 prior non-metformin AHA exposure	53	19748	4906	12625	25772
	no prior non-metformin AHA exposure	54	20490	3377	10210	19225
	no prior canagliflozin exposure	55	38202	8035	22337	43525
	no prior empagliflozin or dapagliflozin exposure	56	38679	8228	22670	44383

Cohort-Defining Drug: DPP-4 inhibitor, GLP-1 agonist, TZD, SU, insulin, or other AHA exposure			Database			
Patient	**Exposure to Background AHA in 365-day period**	**Cohort**	**CCAE**	**MDCD**	**MDCR**	**Optum**
T2DM	at least 1 prior non-metformin AHA exposure	57	8347	450	612	3577
	no prior non-metformin AHA exposure	58	214855	42957	44721	148434
	no prior canagliflozin exposure	59	217985	42994	44848	149231
	no prior empagliflozin or dapagliflozin exposure	60	218808	43335	45119	150820
T2DM w CVD	at least 1 prior non-metformin AHA exposure	61	19748	4906	12625	25772
	no prior non-metformin AHA exposure	62	20490	3377	10210	19225
	no prior canagliflozin exposure	63	38202	8035	22337	43525
	no prior empagliflozin or dapagliflozin exposure	64	38679	8228	22670	44383