Protocol for non-interventional studies based on existing data

Document Number:	[redacted]
BI Study Number:	1245.221
BI Investigational Product(s):	Jardiance® (Empagliflozin)
Title:	**Empagliflozin vs. DPP-4 inhibitors and GLP-1 Receptor Agonists Cost of Care ("EDGE-COS") Study: a German claims data analysis**
Protocol version identifier:	1.0
Date of the last version of the protocol:	20 January 2020 [redacted]
PASS:	No
EU PAS register number:	31950
Active substance:	A10BK03 Empagliflozin ("EMPA") A10BH Dipeptidyl peptidase 4 inhibitors ("DPP-4") • A10BH01 Sitagliptin ("SITA") A10BJ Glucagon-like peptide-1 receptor agonists ("GLP-1-RA")
Medicinal product:	Jardiance
Product reference:	Europe: EMEA/H/C/002677
Procedure number:	N/A
Joint PASS:	No
Research question and objectives:	The main objective of this study is to compare healthcare cost associated with different antidiabetic drug treatments among the incident users of Empagliflozin, any DPP-4i (specifically Sitagliptin) or any GLP-1-RA, based on claims data covering up to four years of follow-up.
Countries of study:	Germany
Country-specific analyses:	This protocol is the locally amended German protocol based on global master protocol template
Author:	[redacted]

	[redacted]
Marketing authorisation holder:	[redacted]
MAH contact person:	
In case of PASS, add: **<EU-QPPV:>**	N/A
In case of PASS, add: <**Signature of EU-QPPV:>**	N/A
Date:	20 January 2020 [redacted]

1. TABLE OF CONTENTS

2. LIST OF ABBREVIATIONS

aDCSI	Adapted Diabetes Complications Severity Index
ATC	Anatomical Therapeutic Chemical classification
BI	Boehringer Ingelheim International GmbH
BMI	Body Mass Index
CABG	Coronary artery bypass grafting
CHF	Congestive heart failure
CI	Confidence interval
CCI	Charlson-Comorbidity-Score
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COPD	Chronic obstructive pulmonary disease
CV	Cardiovascular
DM	Diabetes mellitus
DPP-4i	Dipeptidyl peptidase-4 inhibitor
EMPA	Empagliflozin
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESRD	End-stage renal disease
FGS	"Facharztgruppenschlüssel"
GLM	Generalized linear model
GLP-1-RA	Glucagon-Like Peptide-1 Receptor Agonist
GP	General physician
HbA1c	Glycated hemoglobin A1c
HCRU	Healthcare resource utilization
ICD-10	International Classification of Diseases, 10th revision
KDIGO	Kidney Disease Improving Global Outcomes
MA	Marketing authorization
MET	Metformin
MI	Myocardial infarction
OPS	"Operationen- und Prozedurenschlüssel"
PS	Propensity score
PSM	Propensity score matching
PTCA	Percutaneous transluminal coronary angioplasty
SAP	Statistical analysis plan
SD	Standard deviation
SGLT-2i	Sodium-glucose cotransporter-2 inhibitor
SITA	Sitagliptin
STD.	Standardized difference
SU	Sulfonylureas
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus

3. RESPONSIBLE PARTIES

Function	Name	Affiliation
Project [redacted]	[redacted]	*Boehringer Ingelheim*
Medical Advisor	*tbd*	*Boehringer Ingelheim*
Scientific Advisor	[redacted]	[redacted]
Scientific Advisor		
Scientific [redacted]		
Main Project Management		
Claims Data Management		

4. ABSTRACT

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®			
Name of active ingredient: Empagliflozin (A10BK03)			
Protocol date: 21 May 2019	**Study number:** 1245.221	**Version/Revision:** 1.0	**Version/Revision date:** 14 September 2022
Title of study:	Empagliflozin vs. DPP-4 inhibitors and GLP-1 Receptor Agonists Cost of Care ("EDGE-COS") Study: a German claims data analysis		
Rationale and background:	In recent years, for most of the available SGLT-2 inhibitors (SGLT-2i), DPP-4 inhibitors (DPP-4i) and GLP-1 receptor agonists (GLP-1-RA), cardiovascular outcome studies have been conducted. However, much less is known about cost (both direct and indirect cost) and healthcare resource use (HCRU) associated with therapy of type 2 diabetes mellitus (T2DM) patients with SGLT-2i, DPP-4i and GLP-1-RA. Accordingly, the objective of this study is to collect these data.		
Research question and objectives:	The overall objective of this study is to examine direct and indirect healthcare cost associated with use of Empagliflozin, any DPP-4i (specifically Sitagliptin) or any GLP-1-RA among newly treated T2DM patients in Germany. Costs will be compared between Empagliflozin and DPP-4 inhibitor (in particular Sitagliptin) or GLP-1-RA use. Additional subgroup analyses will be conducted according to observed/unobserved concomitant antidiabetic therapies (e.g. insulin or metformin use).		
Study design:	This will be a non-interventional, retrospective study using claims data from (at least) one major German sick fund / sick-fund service organization (here: [redacted]).		
Population:	This study will include all continuously insured patients with at least one inpatient or outpatient diagnosis of T2DM (ICD-10 code: E11.-) in the [redacted] database between 2014 and 2018[1]. The study population will consist of patients initiating use of either Empagliflozin, any DPP-4i or any GLP-1-RA.		
Variables:	Variables describing all-cause and T2DM-associated healthcare cost will be medication costs (based on list prices), costs for prescribed medical aids and remedies, costs related to outpatient visits, inpatient costs, rehabilitation costs and indirect costs caused by days absent from		

[1] extended to 30/06/2019 in case this data can already be provided by [redacted] at time of data validation (end of Q1 2020)

	work. These costs can be distinguished and directly T2DM-related and T2DM-unrelated costs based on documented primary diagnoses for each visit (identified via ICD-10 E11; E16), respectively ATCs/PZNs for medication and aid costs. Variables related to healthcare resource use (HCRU) will be number of outpatient visits and number of (first and recurrent) hospitalizations, rehabilitations, length of stay and days absent from work. The composite outcomes for overall direct and indirect costs will be based on these variables. Other variables collected from the data sources, according to data availability, will include covariates at baseline on patient characteristics like: age, sex, comorbidity level (using Charlson Comorbidity Index and Top 10 comorbidities that are not covered by CCI), level of care, prior use of antidiabetic drugs, previously observed (T2DM- and non-T2DM related) healthcare costs and days absent from work.
Data sources:	This study will utilize a Germany-wide claims database with a probable size of two million insured people, provided by [redacted]
Study size:	The anticipated number of T2DM patients in the [redacted] database is expected to be approximately 200,000 – 250,000.
Data analysis:	The propensity score matching (PSM) will be used to reduce confounding in the comparative cost analyses across study cohorts. Individual patient characteristics that are relevant to treatment assignment are combined into one item, the propensity score (PS), which measures the conditional probability that the patient is assigned to a treatment rather than the alternative one. Pairwise PS models between i) EMPA and DPP-4i and ii) EMPA and SITA as well as iii) EMPA and GLP-1-RA will be developed using logistic regression including appropriate covariates. For all included patients, descriptive statistics will be generated separately in each cohort before and after matching. Continuous variables will be described by mean, standard deviation, median, 25th, and 75th percentiles, minimum and maximum. Categorical variables and continuous variables that are also categorized will be described by proportion and frequency in each category. All-cause and T2DM-related healthcare costs, as well as HCRU outcomes, will be analysed and compared across cohorts by cost/numbers per patient year (PY). Additionally, subgroup analyses according to observed concomitant antidiabetic treatment will be conducted for 1) insulin-naïve patients, 2) patients receiving only a single antidiabetic drug (AD), 3) patients receiving a dual therapy with metformin, and 4) patients receiving insulin in addition to the respective study drug. Patients with a) discontinuation of drug of interest (DDD-based grace period of 90 days), b) switch to or c) treatment intensification with any drug of the two other study arms[2] will be censored accordingly.
Milestones:	A first steering board meeting for discussion of the study design as outlined in this protocol draft is planned for Q1 2020. Completion of data validation and statistical analysis is planned for Q2/Q3 2020

[2] (or any other SGLT-2i in the EMPA group or any other DPP-4i in the SITA group)

	depending on the data delivery process. The final report will become available end of Q3 2020.

5. AMENDMENTS AND UPDATES

Not applicable.

6. MILESTONES

The following table estimates timelines for the milestones of this German study.

Milestone	Planned Date
Study protocol final (1st Steering board)	Q1 2020
Start of data validation/analysis	Q2 2020
Results presentation (2nd Steering board) and final report available	Q3 2020

7. RATIONALE AND BACKGROUND

The number of patients with type 2 diabetes mellitus (T2DM) is increasing all over the world [1]. In Europe, the number of patients with diabetes is estimated to be 55 million, and it is estimated to increase to 67 million by 2030. T2DM constitutes 90% of these numbers. The prevalence of T2DM in Germany ranges between 7.2% (population aged 18 to 79 years) based on health examination surveys of the [redacted] and 9.9% (among all age groups) based on statutory health insurance data [2].

According to the German National Disease Management Guideline on the Treatment of Type 2 Diabetes, there are four stages in the treatment of T2DM [3]. When individualized HbA1c target is not achieved after three to six months treatment, the first line drug, metformin as monotherapy, is added to the first step lifestyle therapy (education, diet, physical activity). If metformin is not tolerated, DPP-4 inhibitors (DPP-4i), insulin, sulfonylurea or SGLT-2 inhibitors (SGLT-2i) are used. The next step is a two-drug combination or insulin mono therapy.

SGLT-2i represent one of the most recent pharmacological treatment options of T2DM [7]. These drugs have been available since 2012. The globally available SGLT-2i include canagliflozin, dapagliflozin, empagliflozin, and ertugliflozin. Additional SGLT-2i (e.g. ipragliflozin, tofogliflozin) are available also in Asia [7]. SGLT-2i are often used as a second-line treatment of T2DM [8]. In addition to the glucose-lowering effects, these drugs also have beneficial effects on weight, blood pressure, and potentially also lipids [7].

In recent years, for most of the available SGLT-2 inhibitors (SGLT-2i), DPP-4 inhibitors (DPP-4i) and GLP-1 receptor agonists (GLP-1-RA), cardiovascular outcome studies have been conducted. CV events contribute to up to 50% of healthcare costs in the treatment of T2DM. The main cost drivers were hospitalization, drug, and outpatient care costs. However, much less is known about the drug-specific cost and healthcare resource use (HCRU) associated directly with therapy of type 2 diabetes mellitus (T2DM) patients with SGLT-2i, DPP-4i and GLP-1-RA. Accordingly, the objective of this study is to collect these data. The study results shall be used to enter discussions with stakeholders in the German healthcare system (payers, physicians´ associations, etc.) about the real-world use and relevance of Empagliflozin in the treatment of T2DM patients.

8. RESEARCH QUESTION AND OBJECTIVES

The overall objective of this study is to examine healthcare cost associated with the treatment with Empagliflozin versus any DPP-4i, Sitagliptin and any GLP-1-RA therapies in Germany.

8.1 PRIMARY OBJECTIVES

Primary objective: Compare

a) all-cause b) T2DM-associated c) CV-event-associated	healthcare costs (excluding indirect costs of days absent from work)
a) all-cause b) T2DM-associated c) CV-event-associated	healthcare resource utilization (HCRU)

among patients newly treated with:

i. Empagliflozin vs. any DPP-4i
ii. Empagliflozin vs. Sitagliptin
iii. Empagliflozin vs. any GLP-1-RA

8.2 SECONDARY OBJECTIVES

Secondary objective 1: Compare all-cause, T2DM-associated and CV-event-associated healthcare costs (direct costs and indirect costs of days absent from work[3]) between above defined patients (i – iii) for the following subgroups[4]:

a. Only patients without concomitant insulin treatment
b. Only patients without any concomitant antidiabetic treatment
c. Only patients with concomitant metformin treatment (only dual therapy)
d. Only patients with concomitant insulin treatment

Secondary objective 2: Compare all-cause as well as T2DM-associated healthcare resource utilization (HCRU) between above defined patients (i – iii) for the following subgroups:

a. Only patients without concomitant insulin treatment
b. Only patients without any concomitant antidiabetic treatment
c. Only patients with concomitant metformin treatment (only dual therapy)
d. Only patients with concomitant insulin treatment

[3] Costs displayed for patients with confirmed status of employment only
[4] Healthcare costs including indirect costs of days absent from work will also be compared for the overall population

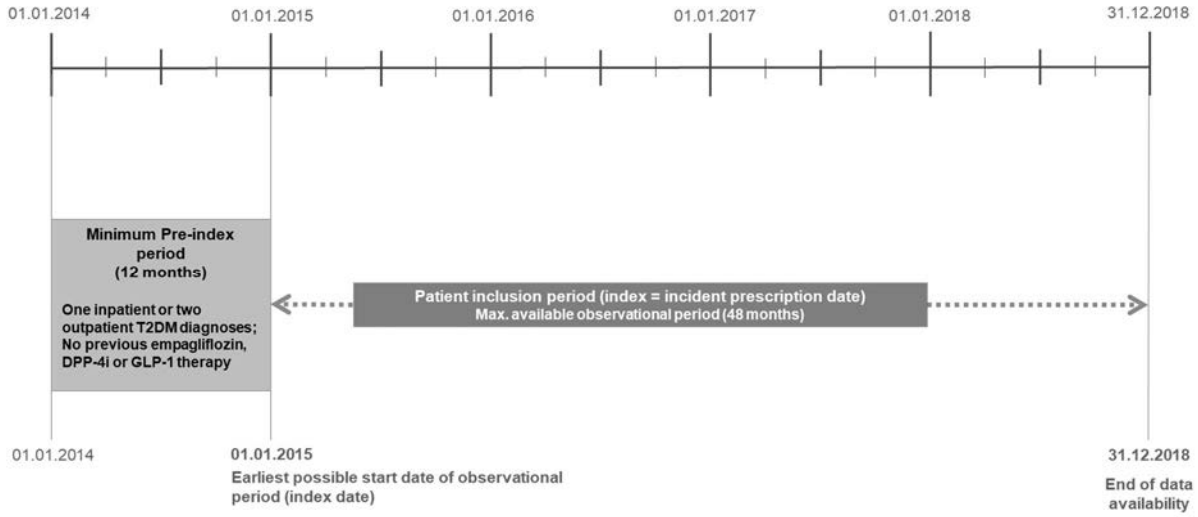
9. RESEARCH METHODS

9.1 STUDY DESIGN

This will be a non-interventional study using existing data and including patients with a diagnosis of T2DM in German claims data. German claims data are a valid and often used source for health-economic analyses [2–4]. So far, the analysis is planned to be performed in collaboration with [redacted], a company providing services to several sickness funds. The strengths of a claims data analysis are the use of a large database and the unselected nature of the data and representativeness of the underlying population. Moreover, HCRU and cost are well covered in such a database. Generally, a claims dataset seems to be an appropriate data source to answer the study questions mentioned above. Therefore, claims data are extremely valuable for the effective examination of health care costs, and HCRU.

The dataset will cover the period 01/01/2014 – 31/12/2018[5] and include all T2DM patients who initiated a treatment with either Empagliflozin, DPP-4i or GLP-1-RA (=index date); specific inclusion criteria are outlined below. **Figure 1** below describes the main design.

Figure 1. Study inclusion and follow-up periods



[5] extended to 30/06/2019 in case this data can already be provided by [redacted] at time of data validation (end of Q1 2020)

9.2 SETTING

9.2.1 Overview

Claims data from several German public sickness funds (one uniform dataset provided by [redacted]) with a probable size of two million insured people will be requested through a university-affiliated partner, [redacted] (coverage: 01/01/2014 – 31/12/2018[6]). The database contains anonymized patient-based data regarding prescriptions, outpatient and inpatient visits (including procedures and diagnoses), and outpatient prescriptions and selected dispensed medications during inpatient treatment (only in case of drug-specific reimbursement) of the patients. A general overview of information that is available in German claims data is provided in **Figure 2**; a detailed description of variables that will be used in this study is provided in section 9.3.

Figure 2: Available variables in German claims databases

Sociodemo-graphic Characteristics	Inpatient Care	Outpatient Care	Outpatient Medication (prescriptions only)	Other Data
Age	Data about initial diagnoses including day of admission	All physicians' visits including type of physician (esp. GPs and specialists in different specialist groups)	Type of medication (medication number – in Germany: PZN), ATC code, number of packs, dates of prescription and of dispensing pharmacy	Costs for outpatient devices/other supportive measures Outpatient surgeries
Gender	All documented diagnoses/ procedures	Documentation of diagnoses/ measures (EBM, GONr, OPS)	Prescribing physician	Other services paid by the insurance, e.g. salary co-payments etc. and Costs of other diagnostic/ therapeutic measures
Type of insurance	Length of stay in days	Description of "safety" of diagnoses	Medication-specific data (DDD, other information)	Outpatient/in-patient long-term care data
Partly: Socioeconomic status	Costs including specific DRG	Dates (physician visits; all diagnostic/ therapeutic measures)	Costs; indirectly by calculation: patients' co-payments	Days absent from work
Mortality	Data about inpatient/ outpatient rehabilitation clinic stays/other follow-up measures	Costs of outpatient care based on activity points documented by doctors		Partly DMP data: height, weight, HbA1c, blood pressure, hypoglycemic events

[6] extended to 30/06/2019 in case this data can already be provided by [redacted] at time of data validation (end of Q1 2020)

9.2.2 Study population

Generally, only prevalent T2DM patients continuously insured by the respective sickness funds will be considered in this study and thus, the following basic selection criteria need to be met:

Inclusion criteria:

- Continuous insurance by the sickness funds for the entire period (01/01/2014 – 31/12/2018; death of a patient is the only accepted exception from this rule)
- At least two outpatient T2DM diagnoses (ICD E11.-) in two different quarters and/or at least one inpatient T2DM diagnosis (ICD E11.-) in the period: 12 months before index date – till 3 months after index date
- At least one prescription of Empagliflozin, DPP-4i or GLP-1-RA in the inclusion period; first prescription date = index date (between 01/01/2015 – 31.12.2018)

Exclusion criteria:

- Any prescription of SGLT-2i, DPP-4i or GLP-1-RA in the baseline ("washout") period (01/01/2014 – 31/12/2014); depending on identified patient numbers in the empagliflozin vs. GLP-1-RA comparison, prior use of DPP-4i will be allowed within the period before the index date
- Patients with two outpatient diagnoses or one inpatient – OR – any diagnosis of
    - T1DM (E10),
    - Secondary diabetes
        - Malnutrition-related diabetes mellitus (E12)
        - Other specified diabetes mellitus (E13) or
    - Gestational diabetes (O24),

  within the 1-year "lookback period" before index date
- Patients with pre-existing diagnosis of neoplasms (C00-C97, D00-D48) or treated with cancer drugs, within the lookback period (ATC main group: L)
- Patients with another pre-existing diagnosis of one of the top-5 most cost-intensive comorbidities (according to the 80 most cost-intensive diseases as defined by the German Federal Insurance Office: "Hierarchisierte Morbiditätsgruppen"; HMGs)

These criteria ensure that only patients with T2DM, and not any other type of diabetes, are included in the study. In addition, the exclusion of patients with conditions that will likely skew the costs (e.g., cancer) ensures unbiased cost estimates.

Based on all patients fulfilling the basic criteria mentioned above, different analysis samples will be selected to answer specific research questions. **Table 1** provides an overview of the analysis samples used in this study as well as the respective selection criteria.

Table 1. Overview on analysis (sub-)samples

Sample	Abbreviation	Selection criteria	Usage
Newly treated Empagliflozin patients	EMPA	All patients with at least one prescription of Empagliflozin during the inclusion period (index date = date of 1st prescription of Empagliflozin) No prescription of DPP-4i, GLP-1-RA or any other SGLT-2i before index date, except prior use of DPP-4i being allowed in the EMPA vs. GLP-1 comparison (in case of small sample size)	PRIMARY OBJECTIVE - Basis for PSM
Newly treated DPP-4i patients	DPP-4	All patients with at least one prescription of any DPP-4i during the inclusion period (index date = date of 1st prescription of any DPP-4i) No prescription of GLP-1-RA or SGLT-2i before index date	PRIMARY OBJECTIVE - Basis for PSM
Newly treated Sitagliptin patients	SITA	All patients with at least one prescription of Sitagliptin during the inclusion period (index date = date of 1st prescription of Sitagliptin) No prescription of SGLT-2i, GLP-1-RA or any other DPP-4i before index date	PRIMARY OBJECTIVE - Basis for PSM
Newly treated GLP-1-RA patients	GLP-1	All patients with at least one prescription of GLP-1-RA during the inclusion period (index date = date of 1st prescription of GLP-1-RA) No prescription of any DPP-4i or SGLT-2i before index date, except prior use of DPP-4i being allowed in the EMPA vs. GLP-1 comparison (in case of small sample size)	PRIMARY OBJECTIVE - Basis for PSM
Excluding unmatched patient for pairwise PSM comparisons:			

Newly treated Empagliflozin patients matched to newly treated DPP-4i patients	EMPA matched to DPP-4	Only EMPA patients with similar propensity scores when compared to the DPP-4 cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Newly treated Empagliflozin patients matched to newly treated Sitagliptin patients	EMPA matched to SITA	Only EMPA patients with similar propensity scores when compared to the SITA cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Newly treated Empagliflozin patients matched to newly treated GLP-1-RA patients	EMPA matched to GLP-1	Only EMPA patients with similar propensity scores when compared to the GLP-1 cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Newly treated DPP-4i patients matched to newly treated Empagliflozin patients	DPP-4 matched to EMPA	Only DPP-4 patients with similar propensity scores when compared to the EMPA cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Newly treated Sitagliptin patients matched to newly treated Empagliflozin patients	SITA matched to EMPA	Only SITA patients with similar propensity scores when compared to the EMPA cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Newly treated Empagliflozin patients matched to newly treated Empagliflozin patients	GLP-1 matched to EMPA	Only GLP-1 patients with similar propensity scores when compared to the EMPA cohort	PRIMARY OBJECTIVE - Comparison for costs - Comparison for HCRU
Sub-cohorts for “EMPA” (for pairwise comparisons unmatched patients will be excluded)[7]:			

[7] Including PS-matched sub-cohorts for each comparison: i) matched to DPP-4; ii) matched to SITA; iii) matched to GLP-1

EMPA without concomitant insulin treatment	EMPA w/o insulin	No insulin prescription within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
EMPA without concomitant antidiabetic treatment	EMPA – mono	No antidiabetic prescription within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
EMPA with concomitant metformin treatment (only dual therapy)	EMPA – dual	At least one metformin prescription within 3 months before/after index date; No further prescription of other antidiabetics within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
EMPA with concomitant insulin treatment	EMPA – insulin	At least one insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
Sub-cohorts for “DPP-4” (for pairwise comparisons unmatched patients will be excluded):			
DPP-4 without concomitant insulin treatment	DPP-4 – w/o insulin	No insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
DPP-4 without concomitant antidiabetic treatment	DPP-4 – mono	No antidiabetic prescription within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU

DPP-4 with concomitant metformin treatment (only dual therapy)	DPP-4 – dual	At least one metformin prescription within 3 months before/after index date; No further prescription of other antidiabetics within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
DPP-4 with concomitant insulin treatment	DPP-4 – insulin	At least one insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
Sub-cohorts for "SITA" (for pairwise comparisons unmatched patients will be excluded):			
SITA without concomitant insulin treatment	SITA w/o insulin	No insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
SITA without concomitant antidiabetic treatment	SITA - mono	No antidiabetic prescription within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
SITA with concomitant metformin treatment (only dual therapy)	SITA - dual	At least one metformin prescription within 3 months before/after index date; No further prescription of other antidiabetics within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
SITA with concomitant insulin treatment	SITA – insulin	At least one insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
Sub-cohorts for "GLP-1" (for pairwise comparisons unmatched patients will be excluded):			
GLP-1 without concomitant insulin treatment	GLP-1 w/o insulin	No insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES

			- Comparison for costs - Comparison for HCRU
GLP-1 without concomitant antidiabetic treatment	GLP-1 - mono	No antidiabetic prescription within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
GLP-1 with concomitant metformin treatment (only dual therapy)	GLP-1 - dual	At least one metformin prescription within 3 months before/after index date; No further prescription of other antidiabetics within 3 months before/after index date is observed	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU
GLP-1 with concomitant insulin treatment	GLP-1 - insulin	At least one insulin prescription within 3 months before/after index date	SECONDARY OBJECTIVES - Comparison for costs - Comparison for HCRU

9.2.3 Study period and follow-up

The inclusion period for all four treatment groups will begin on 01/01/2015[8]. The follow-up for the included patients will begin on the day when the first prescription of the above defined agents is observed (index date) and continue as the most recent data are available from the [redacted] database (max. follow-up of 4 years; until 31/12/2018). Sensitivity analyses for patients with a minimum follow-up time of 12 and 24 months will be conducted additionally. During follow-up, patients will be censored if they die before 31/12/2018, discontinue their initiated treatment with Empagliflozin, DPP-4i or GLP-1-RA or initiate a concomitant use of any SGLT-2i, DPP-4i or GLP-1-RA (including free or fixed-dose combinations).

In the main analyses, the follow-up will end on the date of the first occurrence of any of the following events:

- death,
- discontinuation of the initiated study drug (DDD + grace period of 180 days; see 9.3.1),
- prescription for any of the following antidiabetic therapies, if these were not initially used: any SGLT-2 inhibitor, any DPP-4 inhibitor (if Sitagliptin was initiated, any further DPP-4i will lead to censoring), any GLP-1-RA,
- or end of data availability.

In subgroup analysis additional events for censoring will be included, as described in **Table 2**

Table 2. Censoring criteria for subgroup analysis during the follow-up period

Samples	Censoring criteria (events after index date)
EMPA, DPP-4, SITA and GLP-1 without concomitant insulin treatment	none Sensitivity analysis: Any prescription of insulin
EMPA, DPP-4, SITA and GLP-1 without concomitant antidiabetic treatment	none
EMPA, DPP-4, SITA and GLP-1 with concomitant metformin treatment (only dual therapy)	none
EMPA, DPP-4, SITA and GLP-1 with concomitant insulin treatment	none Sensitivity analysis: Discontinuation of insulin (DDD + grace period of 90 days; see 9.3.1),

[8] 01/07/2015 in case of availability of data set

9.3 VARIABLES

9.3.1 Exposures: Study Drugs

Main exposures: Empagliflozin

The main exposures in this study are incident use of Empagliflozin (ATC codes: A10BK03, formerly A10BX12; A10BD20). The available SGLT-2 inhibitors in Germany to be excluded in any Empagliflozin sub-cohort are presented in **Table 3**. Fixed-dose combination drugs (e.g. Empagliflozin and metformin) will be included.

Main comparators: any DPP-4 inhibitor, Sitagliptin only and any GLP-1-RA,

Incident use of any DPP-4 inhibitor, Sitagliptin in particular and any GLP-1-RA are the main comparators in the analyses. The available DPP-4 inhibitors and GLP-1 receptor agonists in Germany are presented in **Table 3**. Fixed-dose combination drugs with the above defined comparators and another drug other than an SGLT-2 inhibitor (main exposure) will be included. GLP-1 receptor agonists, DPP-4 inhibitors as well as Sitagliptin specifically will be used as the active comparator sub-cohorts because they are used in the treatment of the same disease stage of T2DM as SGLT-2 inhibitors [17]. Some DPP-4 inhibitors might increase the risk of heart failure, which might cause increased indirect healthcare costs. Therefore, this study will include the utilization of Sitagliptin in a separate cohort, since it most likely does not affect HF risk and consequently not increase indirect, CV event-associated costs.

Concomitant use of any SGLT-2 inhibitor, DPP-4 inhibitor or GLP-1 receptor agonist excluded
As described in section 9.2.2, the inclusion and exclusion criteria require that patients have no previous use of any SGLT-2 inhibitor, any GLP-1-RA or any DPP-4 inhibitor (as listed in **Table 3**, in a free or fixed-dose combination). During follow-up, patients are censored if they switch to any SGLT-2 inhibitor, GLP-1-RA or DPP-4 inhibitor or initiate a concomitant use of any SGLT-2 inhibitors, GLP-1-RA and DPP-4 inhibitors (free or fixed-dose combinations), as described in section 9.2.3.

Definition of exposure periods
Drug exposure periods will be defined based on available data on prescription and dispensation data of drugs as provided by [redacted]. Drug exposure will be assumed to begin on the date of a dispensation. Patients are assumed to consume prescribed drugs according to the respective DDD without stock piling. Once no further drug prescription is observed, a grace period of 90 days is used after the latest date of dispensation. Discontinuation date is the date when this 90-day grace period ends.

Table 3. Study drugs form sub-cohorts.

Sub-cohorts and study drugs	ATC code
SGLT-2 inhibitor (including only Empagliflozin fixed-dose combinations with other drugs than DPP-4 inhibitor or GLP-1-RA)	
Empagliflozin	2014 – 2016: A10BX12 A10BK03
Empagliflozin and metformin	A10BD20
Dapagliflozin (listed for exclusion only)	2013 – 2016: A10BX09 A10BK01
Dapagliflozin and metformin (listed for exclusion only)	A10BD15
Any GLP-1 receptor agonist (including fixed-dose combinations with insulin)	
Exenatide	2013 – 2016: A10BX04 A10BJ01
Liraglutide	2013 – 2016: A10BX07 A10BJ02
Liraglutide and Insulin degludec (starting 01/05/2015, until 01/08/2016)	A10AE56
Lixisenatide (until 01/04/2014)	2014 – 2016: A10BX10 A10BJ03
Dulaglutide (starting 01/02/2015)	2015 – 2016: A10BX14 A10BJ05
Any DPP-4 inhibitor (including fixed-dose combinations of a DPP-4 inhibitor and another drug than SGLT-2 inhibitor or GLP-1-RA)	
Sitagliptin	A10BH01
Sitagliptin and metformin	A10BD07
Vildagliptin (until 01/07/2014; starting 01/12/2018)	A10BH02
Vildagliptin and metformin (until 01/07/2014; starting 01/12/2018)	A10BD08
Saxagliptin	A10BH03
Saxagliptin and metformin	A10BD10

ATC=Anatomical Therapeutic Chemical; DPP-4=dipeptidyl peptidase-4; SGLT-2=sodium-glucose cotransporter-2.

9.3.2 Covariates at baseline (Patient Characteristics)

The types of covariates at baseline (baseline patient characteristics) utilized in this study and their measurements periods are listed in **Table 4**. The complete list of covariates, representing each covariate type, and their operational definitions are available in **Annex 1**.

Table 4. Types of covariates at baseline

Covariate[1] at baseline	Measurement period
Sociodemographic and lifestyle characteristics	12 months preceding index date (or Index)
Cardiovascular complications	12 months preceding index date (inclusive)
Other comorbidities (incl. CCI)	12 months preceding index date (inclusive)
Prior use of other antidiabetic drugs	12 months preceding index date (inclusive)
Prior use of CV-related drugs	12 months preceding index date (inclusive)
Healthcare resource utilization covariates	12 months preceding index date (inclusive)
Inpatient and outpatient cost	12 months preceding index date (inclusive)
AD cost	12 months preceding index date (inclusive)
All other drug cost	12 months preceding index date (inclusive)

[1] The complete list of covariates, representing each covariate type, and their operational definitions are available in **Annex 1**.

9.3.3 Outcomes

The following **Table 5** provide an overview about available variables and/or related endpoints that will be analyzed within the study, and how they will be measured based on the claims dataset.

Table 5. Primary and secondary outcomes included in this study.

Outcomes corresponding to each objective	Variable(s)	Definition
Primary outcome 1: Direct healthcare costs (excluding indirect costs of days absent from work)		
Primary outcome 1a: All-cause direct healthcare costs (excluding indirect costs of days absent from work)	Total costs for hospitalizations	Total costs based on DRG reimbursement codes associated to all hospitalizations with any diagnosis ICD-10 code during the post-index period
	Total costs for outpatient physician visits	Total costs based on EBM outpatient settings[9] during the post-index period
	Total costs for outpatient prescriptions	Costs based on pharmacy sales price at prescription date during the post-index period
	Total costs for medical aids and remedies	Costs based on pharmacy sales price at prescription date during the post-index period
	Total costs for rehabilitation stay	Cost directly extracted from the database related to rehabilitation visits with any diagnosis ICD-10 code during the post-index period

[9] The majority of services are not invoiced directly by means of a monetary value but by a system of weighted points. To assess the monetary payment in the outpatient setting, weighted points are usually multiplied by a uniform orientation value, which is defined by the National Association of Statutory Health Insurance Physicians.

Outcomes corresponding to each objective	Variable(s)	Definition
Primary outcome 1b: T2DM-associated direct healthcare costs (excluding indirect costs of days absent from work)	Costs for DM-related hospitalizations	Associated to main or primary diagnosis ICD-10 codes: • T2DM: E11.-, • Hypoglycemia: E16.-
	Costs for DM-related outpatient physician visits	Total costs based on EBM outpatient settings for outpatient visits associated to primary diagnosis ICD-10 codes E11.-, E16.- documented by a physician with the following codes ("Facharztgruppenschlüssel"; FGS): • General physician (GP): FGS 1/2/3 • Diabetologist: FGS 25
	Costs for outpatient prescriptions of antidiabetic drugs	Costs for antidiabetic drugs (identified by ATC/PZN) based on pharmacy sales price at prescription date during the post-index period
	Costs for DM-related medical aids and remedies	Costs based on prescribed aids and remedies: needles, lancets and strips (identified by respective PZNs)
	Costs for DM-related rehabilitation stays	Associated to main or primary diagnosis ICD-10 codes: E11.-, E16.-
Primary outcome 1c: CV-associated direct healthcare costs (excluding indirect costs of days absent from work)	Costs for CV-related hospitalizations	Associated to main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
	Costs for CV-related outpatient physician visits	Total costs based on EBM outpatient settings for outpatient visits associated to primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
	Costs for outpatient prescriptions of CV drugs	Costs for CV drugs (identified by ATC/PZN) based on pharmacy sales price at prescription date during the post-index period
	Costs for CV-related medical aids and remedies	Costs based on prescribed aids and remedies
	Costs for CV-related rehabilitation stays	Associated to main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
Primary outcome 2: Healthcare resource utilization		
Primary outcome 2a:	Total count of hospitalizations	Total number of admissions per observed patient year during the post-index period

Outcomes corresponding to each objective	Variable(s)	Definition
All-cause healthcare resource utilization	Length of hospital stay (LOS)	I. Average length of stay per admission during the post-index period Total number of days in hospital per observed patient year during the post-index period
	Total count of quarters with outpatient GP visits	Outpatient visits documented by a GP (quarterly counted)
	Total count of quarters with outpatient diabetologist visits	Outpatient visits documented by a diabetologist (quarterly counted)
	Total count of quarters with outpatient visits of other specialist physicians	Outpatient visits documented by other specialist physicians (quarterly counted)
	Total count of days absent from work	Based on number of all days absent from work during the post-index period
	Total count of rehabilitation stays	Total number of rehabilitations stays per observed patient year during the post-index period
Primary outcome 2b: T2DM-associated healthcare resource utilization	Count of DM-related hospitalizations	Number of admissions with main or primary diagnosis ICD-10 codes: E11.- or E16.- per observed patient year during the post-index period
	Length of DM-related hospital stay (LOS)	I. Average length of stay per admission with main or primary diagnosis ICD-10 codes: E11.- or E16.- during the post-index period II. Total number of days in hospital with main or primary diagnosis ICD-10 codes: E11.- or E16.- per observed patient year during the post-index period
	Total count of quarters with/without DM-related outpatient GP visits	Outpatient visits with main or primary diagnosis ICD-10 codes: E11.- or E16. documented by a GP (quarterly counted)
	Total count of quarters with/without DM-related outpatient diabetologist visits	Outpatient visits with main or primary diagnosis ICD-10 codes: E11.- or E16. documented by a diabetologist (quarterly counted)
	Total count of quarters with/without DM-related outpatient visits of other specialist physicians	Outpatient visits with main or primary diagnosis ICD-10 codes: E11.- or E16. documented by other specialist physicians (quarterly counted)
	Count of DM-related days absent from work	Based on number of all days absent from work associated to main or primary diagnosis ICD-10 codes: E11.- or E16.- during the post-index period
	Count of DM-related rehabilitation stays	Total number of rehabilitations stays with main or primary diagnosis ICD-10 codes: E11.- or E16.- per observed patient year during the post-index period

Outcomes corresponding to each objective	Variable(s)	Definition
Primary outcome 2c: CV-associated healthcare resource utilization	Count of CV-related hospitalizations	Number of admissions with main or primary diagnosis ICD-10 codes: E11.- or E16.- per observed patient year during the post-index period
	Length of CV-related hospital stay (LOS)	I. Average length of stay per admission during the post-index period with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63 II. Total number of days in hospital per observed patient year during the post-index period with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
	Total count of quarters with/without CV-related outpatient GP visits	Outpatient visits with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63 documented by a GP (quarterly counted)
	Total count of quarters with/without CV-related outpatient diabetologist visits	Outpatient visits with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63 documented by a diabetologist (quarterly counted)
	Total count of quarters with/without CV-related outpatient visits of other specialist physicians	Outpatient visits with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63 documented by other specialist physicians (quarterly counted)

Outcomes corresponding to each objective	Variable(s)	Definition
	Count of CV-related days absent from work	Associated to main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
	Count of CV-related rehabilitation stays	Total number of rehabilitations stays per observed patient year during the post-index period with main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63
Secondary outcome 1: Direct and indirect healthcare costs (including indirect costs of days absent from work[10])		
Secondary outcome 1a: All-cause (direct and indirect) healthcare costs	Total costs for hospitalizations	See above (Primary outcome 1a)
	Total costs for outpatient physician visits	
	Total costs for outpatient prescriptions	
	Total costs for medical aids and remedies	
	Total costs for rehabilitation stay	
	Indirect costs of days absent from work	Based on number of days absent from work multiplied with average gross salary
Secondary outcome 1b: T2DM-associated (direct and indirect) healthcare costs	Costs for DM-related hospitalizations	See above (Primary outcome 1b)
	Costs for DM-related outpatient physician visits	
	Costs for outpatient prescriptions of antidiabetic drugs	
	Costs for DM-related medical aids and remedies	
	Costs for DM-related rehabilitation stays	

[10] Costs displayed for patients with confirmed status of employment only

Outcomes corresponding to each objective	Variable(s)	Definition
	DM-related indirect costs of days absent from work	Associated to main or primary diagnosis ICD-10 codes: E11.-, E16.-
Secondary outcome 1c: CV-event-associated (direct and indirect) healthcare costs	Costs for CV-related hospitalizations	See above (Primary outcome 1c)
	Costs for CV-related outpatient physician visits	
	Costs for outpatient prescriptions of CV drugs	
	Costs for CV-related medical aids and remedies	
	Costs for CV-related rehabilitation stays	
	CV-related indirect costs of days absent from work	Associated to main or primary diagnosis ICD-10 codes: • Heart failure: I11.0, I13.0, I13.2 or I50 • Coronary artery disease: I25 • Myocardial infarction: I21 • Ischemic stroke: I63

9.3.4 Other variables required for analyses

In addition, the variables listed below will be collected for performing the alternative and sensitivity analyses (see section 9.7):

- History of any CV disease in the 12 months prior to the index date (***Annex 1***)

9.4 DATA SOURCES

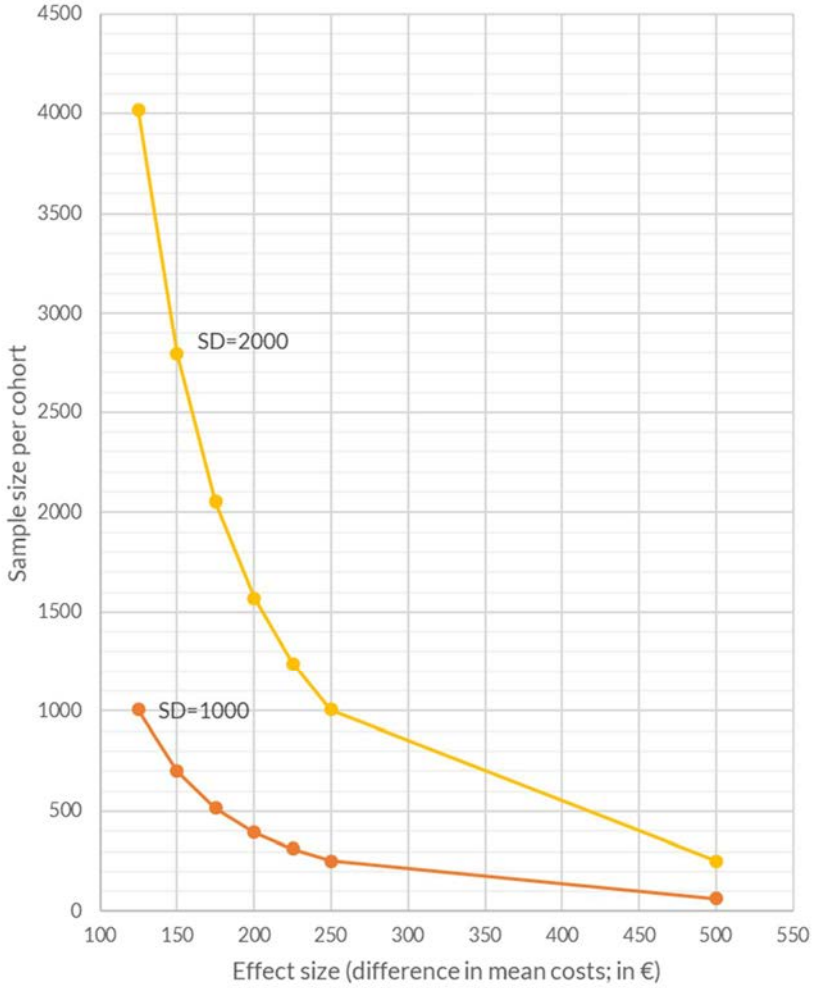
This study will be based on the [redacted] claims database. Health insurance is mandatory in Germany, with approximately 85% of the population (about 70 million persons) being covered by German Statutory Health Insurance (SHI). German health insurance companies provide nearly full coverage of all healthcare-related services (except for co-payments such as the costs for over-the-counter medication). Information on patient socio-demographics as well as diagnoses, procedures, and costs from primary care data, pharmaceuticals, hospitalizations, and inability to work are available.

[redacted] is a service provider for about [redacted] German-wide smaller sickness funds. The [redacted] dataset contains administrative data for approximately two million insured persons. All analyses will be conducted locally in Germany. [redacted] will involve the [redacted]), a university-affiliated partner with longstanding experience in conducting claims data analyses. [redacted] will close a contract with [redacted] (conditional on approval of study protocol), which will define inclusion/exclusion criteria for patient selection, all variables and formats to be used in the data extraction process, and the data protection measures needed to be implemented by [redacted]. [redacted] will participate in the steering board of the study and will co-author all future publications of study results.

9.5 STUDY SIZE

Figure 3 shows the sample size required to detect a 5% or 10% difference in the average cost with 80% power and 5% alpha level. Varying expected differences for mean costs (125-250 Euros in assumed cost average of 2,500 Euros per patient-year) are presented assuming two equally numbered groups for comparison.

Figure 3: Minimum sample sizes for a 0.05 significance level and 0.8 power level of a two-sample comparison of average costs per patient year



Differences in means (assumed cost average; per patient year in Euro)	125 (5%)	150 (6%)	175 (7%)	200 (8%)	225 (9%)	250 (10%)
Sample required, per cohort (after matching); assuming SD = 1,000	1006	699	514	394	312	253
Sample required, per cohort (after matching); assuming SD = 2,000	4020	2792	2052	1571	1242	1006

SD = Standard Deviation

For example, with assuming diabetes-associated costs of 2,500 € per patient year (SD = 1,000) for two equally numbered groups, the analyses are sufficiently powered (>80%) to detect a cost difference (effect size) of 250 € (10%) decrease in costs per patient year between these groups when 253 patients per cohort, i.e., 506 patients in total, can be included.

9.6 DATA MANAGEMENT

The study organization/research organization ([redacted].) will conduct the analysis according to all relevant legal requirements (EC data protection guidelines). All data will be checked for their validity using the median, arithmetic mean, max, and min for each variable as criteria. Patients with any implausible/counter-intuitive data will be excluded from the dataset.

9.7 DATA ANALYSIS

9.7.1 Baseline characteristics / Descriptive statistics

For all included patients, descriptive statistics will be generated separately in each sub-cohort. Continuous covariates will be described by the mean, standard deviation (SD), median, interquartile range (IQR), minimum and maximum. Categorical covariates and continuous covariates that were also categorized will be described by proportion and frequency in each category.

Patient characteristics at cohort entry for the main exposure cohort and the comparator cohort patients and the standardized difference before and after matching will be presented as in **Table 6** below.

Table 6. Illustration of baseline characteristics table. Among the main exposure cohorts (Empagliflozin cohort) and the comparator cohorts, baseline characteristics variables and their standardized difference (Std.) will be presented before and after matching.

	Before matching			**After matching**		
	Main exposure sub-cohort	**Comparator sub-cohort**	**Std.**	**Main exposure sub-cohort**	**Comparator sub-cohort**	**Std.**
Continuous variable 1 (e.g. age) that has been also categorized						
Category 1 (e.g. <39 years) N (%)	n (p.pp)	n (p.pp)	-	n (p.pp)	n (p.pp)	-
Category 2 (e.g. 40-60 years) N (%)	n (p.pp)	n (p.pp)	-	n (p.pp)	n (p.pp)	-
Category 3 (e.g. >60 years) N (%)	n (p.pp)	n (p.pp)	-	n (p.pp)	n (p.pp)	-
Mean (SD)	x.xx (x.xx)	x.xx (x.xx)	x.xx	x.xx (x.xx)	x.xx (x.xx)	x.xx
Median (IQR)	x.xx (x.xx)	x.xx (x.xx)	(based on	x.xx (x.xx)	x.xx (x.xx)	(based on mean)
Min, Max	x.xx, x.xx	x.xx, x.xx	mean)	x.xx, x.xx	x.xx, x.xx	
Categorical variable 1 (e.g. sex)						
Category 1 (e.g. male) N (%)	n (p.pp)	n (p.pp)	x.xx	n (p.pp)	n (p.pp)	x.xx
Category 2 (e.g. female) N (%)	n (p.pp)	n (p.pp)	-	n (p.pp)	n (p.pp)	-
Continuous variable 2 that has not been categorized						
Mean (SD)	x.xx (x.xx)	x.xx (x.xx)	x.xx	x.xx (x.xx)	x.xx (x.xx)	x.xx
Median (IQR)	x.xx (x.xx)	x.xx (x.xx)	(based on	x.xx (x.xx)	x.xx (x.xx)	(based on mean)
Min, Max	x.xx, x.xx	x.xx, x.xx	mean)	x.xx, x.xx	x.xx, x.xx	

Std.=standardized difference; SD=standard deviations; Q1=first quartile; Q3=third quartile.

9.7.2 Propensity score methodology

As mentioned above, a propensity-score matching (PSM) is planned to answer the primary analysis question. When comparing individuals receiving different treatments, the question of comparability due to systematic differences across treatment groups needs to be addressed. PSM is a common approach to adjust for patients' characteristics. The idea is to combine several individual characteristics that are relevant to treatment assignment into one item, the propensity score, which measures the conditional probability that the patient is assigned to a treatment rather than the alternative one. Patients with the same propensity score have theoretically the same probabilistic distribution over other covariates, independently of the treatment they received. Therefore, outcomes of patients with similar propensity scores can be compared to determine the treatment effect on each outcome. Based on the estimated propensity scores, patients initiating Empagliflozin will be matched to respective comparator cohorts [i) DPP-4; ii) SITA; iii) GLP-1 as well as sub-cohorts as described in **Table 1**] using a nearest-neighbor 1:1 matching algorithm with maximum caliper of 0.001. While a limited number of studies implemented a 1: n matching procedure [33], most studies in the literature use 1:1 matching [6, 16, 30, 34] which reduces the estimation bias. In practice, propensity scores will be estimated using logistic regression models, where the dependent variable is a binary indicator of group affiliation (Empagliflozin versus respective comparator). Covariates in the logistic regression should include any relevant patient characteristic or condition which is relevant to treatment assignment. This will include following variables, which refer to the index date or the related 12-month pre-index period:

- Age in years,
- Gender (female = 1; male = 0),
- Charlson Comorbidity Index (CCI),
- Top 10 comorbidities that are not covered by the CCI,
- Kidney disease (by stage)
    - Stage 1 (Y/N)
    - Stage 2 (Y/N)
    - Stage 3 (Y/N)
    - Unspecified (Y/N)
- Level of care,
- Antidiabetic treatment
    - insulin (Y/N)
    - metformin (Y/N)
    - sulfonylureas (Y/N)
    - other antidiabetic drugs except study drugs (e.g. thiazolidinediones; Y/N)
    - Number of antidiabetic drugs (prescribed in parallel)
- CV-related therapies:
    - Anticoagulant/ Platelet aggregation inhibition (Y/N)
    - RAAS inhibitors(Y/N)
    - Beta-blocking agents(Y/N)
    - Calcium-Channel blockers (Y/N)
    - Lipid modifying agents (Y/N)
- T2DM-related costs (cost per observed patient year)
    - Inpatient costs
    - Outpatient costs

  - Medication costs
  - Rehabilitation
  - Cost of remedies / aids
- Non-T2DM-related costs (all-cause costs minus T2DM-related costs)
  - Inpatient costs
  - Outpatient costs
  - Medication costs
  - Rehabilitation
  - Cost of remedies / aids
- Number of days absent from work
  - T2DM-related
  - not T2DM-related

Each covariate included in the PSM analysis is expected to affect both treatment assignment and the outcomes of interest. PSM quality will be assessed in two ways:

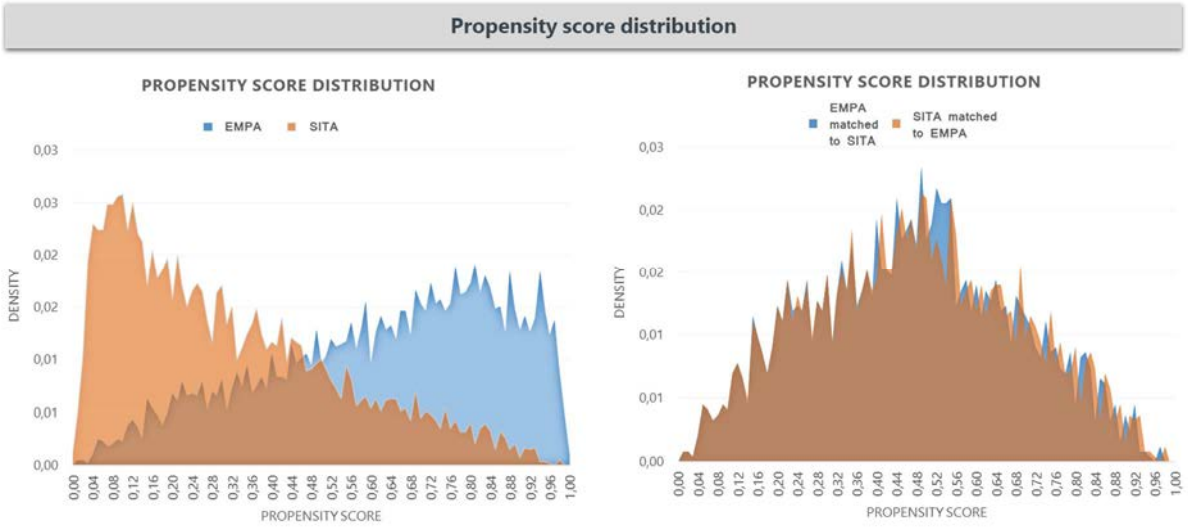
1) Standardized differences (STD) between comparison groups will be used to assess the balance of covariates after matching (number of variables with – still – significant differences (> 0.1) between Empagliflozin and comparators). **Table 7** lists the statistics that will be separately reported for cost and HCRU variables in each group comparison.
2) We will compare the propensity score distributions before and after matching as illustrated in **Figure 4**.

Table 7. Statistics reported for HCRU / cost variables

Variable name	Statistics
All-cause, T2DM-related and CV-related HCRU	
All-cause hospitalizations and hospitalizations related to T2DM	• Number (%) of patients with at least 1 hospitalization • Number of hospitalizations per PY: Mean, SD, Median • Number of hospitalization days per PY: Mean, SD, Median
Outpatient GP visits (any visit, related to T2DM and related CV diseases as described in Table 5)	• Number (%) of patients with at least 1 visit • Number of visits per PY: Mean, SD, Median
Outpatient diabetologist visits (any visit, related to T2DM and related CV diseases as described in Table 5)	• Number (%) of patients with at least 1 visit • Number of visits per PY: Mean, SD, Median
Rehabilitation stays (any visit, related to T2DM and related CV diseases as described in Table 5)	• Number (%) of patients with at least 1 rehabilitation stay • Number of rehabilitation days per PY: Mean, SD, Median
All-cause days absent from work, related to T2DM and related CV diseases as described in Table 5	• Number (%) of patients with at least 1 absence • Number of days per PY: Mean, SD, Median
Top-10 Aids and remedies	• Number (%) of patients with at least 1 aid or remedy • Number of aids or remedies per PY: Mean, SD, Median

Variable name	Statistics
All-cause, T2DM-related and CV-related costs	
Costs for hospitalizations	• Mean, SD, Median per PY
Costs for outpatient physician visits	
Costs for outpatient prescriptions	
Costs for rehabilitation stays	
Costs for medical aids and remedies	
Indirect costs for days absent from work	

Figure 4. Illustration of propensity score distribution among the main exposure cohort ("EMPA") and among the comparator cohorts (e.g. "SITA") before and after matching.



9.7.3 Primary outcomes

For the primary outcomes, average direct cost in Euros per observed patient year with corresponding standard deviations (SD) will be calculated, separately for each pairwise comparison of treatments and subgroups (see **Table 1** for subsample definitions). The mean difference of the cost parameters between the matched samples will be calculated and reported as outlined in **Table 8**.

Table 8. Reporting of average cost per observed patient year

Type of cost	EMPA matched to SITA	SITA matched to EMPA	Effect size
Cohort size	(n = XXX)	(n = XXX)	
Outpatient care	Mean (SD)	Mean (SD)	difference of means (p-value)
Inpatient care	Mean (SD)	Mean (SD)	difference of means (p-value)
Pharmaceuticals/Outpatient prescriptions	Mean (SD)	Mean (SD)	difference of means (p-value)
Rehabilitation	Mean (SD)	Mean (SD)	difference of means (p-value)
Medical aids and remedies	Mean (SD)	Mean (SD)	difference of means (p-value)
Total	Mean (SD)	Mean (SD)	difference of means (p-value)

9.7.4 Secondary outcomes

For secondary cost outcomes, indirect costs of days absent from work[11] will be added to direct costs analyzed in the respective primary outcomes (separately for all-cause and T2DM-related costs) (section 9.7.3). The analysis of secondary HCRU outcomes will be conducted in a similar way as the analysis of primary outcomes. The analysis of HCRU outcomes will include the description of the number of outcome events. The mean difference of the HCRU parameters between the matched samples will be calculated and reported as outlined in **Table 9**.

Table 9. Reporting of mean healthcare resource utilization per patient year

Type of resource utilization	EMPA matched to SITA	SITA matched to EMPA	Effect size
	(n = XXX)	(n = XXX)	
Number of outpatient visits	Mean (SD)	Mean (SD)	difference of means (p-value)
Number of hospitalizations	Mean (SD)	Mean (SD)	difference of means (p-value)
Number of days in hospital	Mean (SD)	Mean (SD)	difference of means (p-value)
Days absent from work (sick leave)	Mean (SD)	Mean (SD)	difference of means (p-value)
Days of rehabilitation	Mean (SD)	Mean (SD)	difference of means (p-value)

9.7.5 Sensitivity analyses

The core analyses for the primary outcome, as described in section 9.7.3, will be re-performed for patients with a minimum follow-up time of 12 and 24 months post-index, respectively. Furthermore, patients with history of any cardiovascular (CV) disease (see **Annex 1**) in the 12 months prior to the index date will be excluded in addition. Since applying these criteria will change the study population, the PS matching and balance analyses will also be re-performed.

In case that relevant differences in the medication costs and/or HCRU between Empagliflozin and GLP-1-RA would be identified, separate considerations of Dulaglutide and Liraglutide will be carried out in the context of a sensitivity analysis with respect to the primary outcomes.

[11] Costs displayed for patients with confirmed status of employment only

9.8 QUALITY CONTROL

The study will be conducted as specified in this protocol. The principal investigator, the co-investigators and the sponsors of the study must approve all revisions to the protocol. All changes must be documented as protocol amendments.

The study team must maintain adequate and accurate records to enable the conduct of the study to be fully documented, including but not limited to the protocol and protocol amendments, and documentation of IRB and Health Authority approval/notification (if applicable). The study team ensures that the datasets and statistical programs used for generating the data included in the final research report are kept in electronic format and are available for auditing and inspection. Records and documents pertaining to the conduct of this activity must be retained for at least 15 years after completion of the activity, or for the length of time required by relevant national or local health authorities, whichever is longer. After that period, the documents may be destroyed, subject to local regulations. No records may be disposed of without the written approval of the research initiator.

This study protocol has been written following the Code of Conduct by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) [23]. The protocol also follows the key elements of the Guideline for Good Pharmacoepidemiology Practices by International Society for Pharmacoepidemiology. TFS Trial Form Support, Quantify Research, the principal investigator, co-investigators, the Sponsors and individuals acting on their behalf commit to adhere to the rules of the ENCePP Code of Conduct in their entirety.

9.9 LIMITATIONS OF THE RESEARCH METHODS

Claims data are extremely valuable for the effective examination of health care outcomes, treatment patterns, health care resource utilization, and costs. Validity of recording and coding can be evaluated as high in these databases. This is especially the case for any inpatient and prescription data as this data is directly relevant for reimbursement of hospitals/pharmacies by sickness funds. The strengths of a claims data analysis are the use of a large database and the elimination of a patient/study site bias. However, all claims databases have certain inherent limitations because the claims are collected for the purpose of payment and not research:

9.9.1 Data sources and variables

Information available in claims data is limited to parameters which are generally relevant for reimbursement purposes and thus, the following variables will not be available: duration since first T2DM diagnosis, specific prescribed daily dosages, reasons for prescribing/not prescribing specific agents, laboratory values such as HbA1c

9.9.2 Identifying patients with T2DM

The presence of a diagnosis code on a medical claim is not positive presence of disease, as the diagnosis code may be incorrectly coded (i.e. wrong ICD-10 code) or included as rule-out criteria rather than actual disease (i.e. wrong extension of ICD-10 code: “a” for “ausgeschlossen” (excluded diagnosis) vs. “g” for “gesichert” (secured diagnosis). In addition, the beginning of T2DM symptoms cannot be determined in this study setting. Therefore, the duration of the disease cannot be determined.

9.9.3 Drug exposure

A claim for a filled prescription is not an indication the medication was consumed or taken as prescribed, and the available follow-up data will be limited to up to four years. Furthermore, physician-provided samples, samples taken as part of a clinical trial or over-the-counter medications will not be observed in claims data.

9.9.4 Bias and confounding

Missing data, e.g. only observing patient characteristics for a subset of the population, could introduce bias into the study if the data are systematically missing for selected samples. However, health care claims are generally required to contain the relevant pieces of information in order for the claim to be processed. All medical claims are required to have a diagnosis code, and all pharmacy claims will have a corresponding National Drug Code (German PZN) code. Thus, there will be residual confounding related to, for example, unobservable differences in patient characteristics, which may influence the findings of this study. However, these differences are expected to be reflective of the populations outside of the controlled setting of a clinical trial.

9.9.5 Analyses

This will be an observational study and therefore, the results will indicate a correlation between drug use and selected outcomes. Causality cannot be determined in observational studies.

9.10 OTHER ASPECTS

None specified.

9.11 SUBJECTS

Not applicable as this is a claims data study.

9.11.1 Cases

Not applicable as this is a claims data study.

9.11.2 Controls

Not applicable as this is a claims data study.

9.12 BIAS

This study will include new users of Empagliflozin, any GLP-1 receptor agonist, and any DPP-4 inhibitor. Immortal time bias [24] will be minimal as the follow-up will begin at the initiation of drug use. Adherence to drug use will be not specified. Potential bias from exposure misclassification will be addressed with grace periods.

Comparing the outcomes between new users of Empagliflozin and any DPP-4 inhibitor or GLP-1 receptor agonist will decrease confounding by indication, as patients in the Empagliflozin cohorts and patients in the comparison cohorts will have a similar indication to drug use [16]. Another approach of decreasing confounding by indication is a restriction of the study sample only to T2DM patients without additionally documented T1DM diagnosis.

The compared cohorts will be matched on the basis of propensity-scores (PS). The PS will be computed on the basis of a wide range of covariates related to sociodemographic characteristics, lifestyle, diabetes complications, other comorbidities and prior/concomitant use of other antidiabetic drugs. The PS matching will decrease the systematic differences between the cohorts and, thus, decrease confounding.

10. PROTECTION OF HUMAN SUBJECTS

Per design, this non-interventional study utilizes secondary data. Thus, the study does not affect the treatment or health outcomes of the study individuals. The study individuals will not be contacted in any phase of the study.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Not applicable. As stated in section VI.C.1.2.1. of BI GVP Module VI, adverse-reactions reporting is not required for non-interventional study designs which are based on secondary use of data.

There is no potential that any employee of BI or agent working on behalf of BI will access individual patient data in which the patient may be identified during data compilation, data reporting or data analysis.

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Manuscripts describing this work will be submitted for publication in peer-review journals. Results may also be submitted for presentation at scientific conferences.

13. REFERENCES

1 Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. *Nature Reviews Endocrinology* 2012;**8**:228–36. doi:10.1038/nrendo.2011.183

2 Heidemann C, Scheidt-Nave C. Prevalence, incidence and mortality of diabetes mellitus in adults in Germany – A review in the framework of the Diabetes Surveillance. *Journal of Health Monitoring* 2017;**2**:98–121.

3 Pfeiffer AFH, Klein HH. The Treatment of Type 2 Diabetes. *Dtsch Arztebl Int* 2014;**111**:69–82. doi:10.3238/arztebl.2014.0069

4 Einarson TR, Acs A, Ludwig C, *et al.* Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007-2017. *Cardiovasc Diabetol* 2018;**17**:83. doi:10.1186/s12933-018-0728-6

5 Kang YM, Kim Y-J, Park J-Y, *et al.* Mortality and causes of death in a national sample of type 2 diabetic patients in Korea from 2002 to 2013. *Cardiovascular Diabetology* 2016;**15**:131. doi:10.1186/s12933-016-0451-0

6 Einarson TR, Acs A, Ludwig C, *et al.* Economic Burden of Cardiovascular Disease in Type 2 Diabetes: A Systematic Review. *Value Health*;**21**:881–90. doi:10.1016/j.jval.2017.12.019

7 Hsia DS, Grove O, Cefalu WT. An update on sodium-glucose co-transporter-2 inhibitors for the treatment of diabetes mellitus. *Curr Opin Endocrinol Diabetes Obes* 2017;**24**:73–9. doi:10.1097/MED.0000000000000311

8 International Diabetes Federation. IDF Clinical Practice Recommendations for managing Type 2 Diabetes in Primary Care. 2017.https://www.idf.org/component/attachments/attachments.html?id=1270&task=download (accessed 25 Sep 2018).

9 Scheen AJ. Cardiovascular Effects of New Oral Glucose-Lowering Agents: DPP-4 and SGLT-2 Inhibitors. *Circ Res* 2018;**122**:1439–59. doi:10.1161/CIRCRESAHA.117.311588

10 Persson F, Nyström T, Jørgensen ME, *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. doi:10.1111/dom.13077

11 Kosiborod M, Cavender MA, Fu AZ, *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;**136**:249–59. doi:10.1161/CIRCULATIONAHA.117.029190

12 Kosiborod M, Lam CSP, Kohsaka S, *et al.* Cardiovascular Events Associated With SGLT-2 Inhibitors Versus Other Glucose-Lowering Drugs: The CVD-REAL 2 Study. *J Am Coll Cardiol* 2018;**71**:2628–39. doi:10.1016/j.jacc.2018.03.009

13 Birkeland KI, Jørgensen ME, Carstensen B, *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;**5**:709–17. doi:10.1016/S2213-8587(17)30258-9

14 Patorno E, Pawar A, Franklin J, *et al.* Empagliflozin And The Risk Of Heart Failure Hospitalization In Routine Clinical Care: A First Analysis From The Empagliflozin Comparative Effectiveness And Safety (EMPRISE) Study. Chicago, Illinois: : American Heart Association

15 Psaty, BM, Koepsell, TD, Lin, D, *et al.* Assessment and Control for Confounding by Indication in Observational Studies. *J Am Geriatr Soc*;**47**:749–54.

16 Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep* 2015;**2**:221–8. doi:10.1007/s40471-015-0053-5

17 Association AD. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2018. *Diabetes Care* 2018;**41**:S73–85. doi:10.2337/dc18-S008

18 Kidney International Supplements | KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease | ScienceDirect.com. https://www.sciencedirect.com/journal/kidney-international-supplements/vol/3/issue/1 (accessed 13 Mar 2019).

19 Zinman B, Wanner C, Lachin JM, *et al.* Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *New England Journal of Medicine* 2015;**373**:2117–28. doi:10.1056/NEJMoa1504720

20 Wang SV, Schneeweiss S, Rassen JA. Optimal Matching Ratios in Drug Safety Surveillance: *Epidemiology* 2014;**25**:772–3. doi:10.1097/EDE.0000000000000148

21 Austin PC. Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. *Pharmacoepidemiology and Drug Safety* 2008;**17**:1218–25. doi:10.1002/pds.1674

22 Dongsheng Yang, Jarrod E. Dalton. A unified approach to measuring the effect size between two groups using SAS. http://support.sas.com/resources/papers/proceedings12/335-2012.pdf

23 European Medicines Agency. The ENCePP Code of Conduct. http://www.encepp.eu/code_of_conduct/ (accessed 22 Nov 2018).

24 Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007;**16**:241–9. doi:10.1002/pds.1357

Annex 1. DEFINITIONS FOR COVARIATES AT BASELINE

Covariate	Definition (when applicable)	ICD-10-GM / OPS-2019
Sociodemographic characteristics (at index date)		
Age	By year	
Sex	female, male	
Covariates related to lifestyle (within 12 months before index date)		
Smoking	–	F17, Z72.0, T65.2
Alcohol abuse or dependence	–	F10, K70, G62.1, I42.6, Z71.4
Drug abuse or dependence	–	F11; F12, F13, F14, F15, F16, F18, F19, Z71.5
Comorbidities (12 months preceding index date)		
CCI	See Annex 2	See Annex 2
Major cardiovascular disease	Combined endpoint of diagnosis with myocardial infarction, heart failure, coronary artery disease or ischemic stroke	• I11.0 • I13.0 or I13.2 • I21 • I25 • I50 • I63
Kidney disease (acute/chronic)	Based on documented ICD-10 diagnoses, separated by stage 1-3	• N17 • N18 • N19
Sepsis	Based on the definition for the 80 most cost-intensive diseases (HMG) by the German Federal Insurance Office	• A02.1 • A20.7 • A22.7 • A26.7 • A32.7 • A39.1 • A39.2 • A39.3 • A39.4 • A40.0 • A40.1 • A40.2 • A40.3 • A40.8 • A40.9 • A41.0 • A41.1 • A41.2 • A41.3 • A41.4 • A41.5 • A41.51 • A41.52 • A41.58 • A41.8 • A41.9

		• A42.7 • B00.7 • B37.7 • O88.3 • P36.0 • P36.1 • P36.2 • P36.3 • P36.4 • P36.5 • P36.8 • P36.9 • R57.2 • R65.0 • R65.1 • R65.2 • R65.3 • R65.9
HIV / AIDS	Based on the definition for the 80 most cost-intensive diseases (HMG) by the German Federal Insurance Office	• B20 • B21 • B22 • B23.0 • B23.8 • B24 • O98.7 • U60.1 • U60.2 • U60.3 • U60.9 • U61.1 • U61.2 • U61.3 • U61.9 • U85 • Z21
Opportunistic infections (caused by pathogens)	Based on the definition for the 80 most cost-intensive diseases (HMG) by the German Federal Insurance Office	• A31.0 • A81.2 • B25.0 • B25.1 • B25.2 • B25.8 • B25.80 • B25.88 • B25.9 • B37.1 • B37.81 • B42.0 • B44.0 • B44.1 • B44.2 • B44.7 • B44.8 • B44.9 • B45.0

		• B45.1 • B45.3 • B45.7 • B45.8 • B45.9 • B46.0 • B46.1 • B46.2 • B46.4 • B46.5 • B46.8 • B59 • U83
Acquired hemolytic anemia	Based on the definition for the 80 most cost-intensive diseases (HMG) by the German Federal Insurance Office	• D59.0 • D59.1 • D59.2 • D59.3 • D59.4 • D59.5 • D59.6 • D59.8 • D59.9 • D63.0 • D63.8
Aplastic or sideroblastic anemia	Based on the definition for the 80 most cost-intensive diseases (HMG) by the German Federal Insurance Office	• D60.0 • D60.1 • D60.8 • D60.9 • D61.0 • D61.1 • D61.10 • D61.18 • D61.19 • D61.2 • D61.3 • D61.8 • D61.9 • D64.0 • D64.1 • D64.2 • D64.3
Prior/concomitant use of other antidiabetic drugs (12 months preceding index date and at index date)		
Covariate	**Definition (when applicable)**	**ATC**
Previous diabetes-related drug treatment (documented antidiabetic prescriptions including ATC/PZN codes in the 12-months pre-index period, including prescribed defined daily dosages (DDDs) per observed patient year),	Use of insulin, metformin, sulfonylureas and other antidiabetics (e.g. thiazolidinediones) as identified by respective ATC codes Number of antidiabetic drugs (prescribed in parallel)	• A10A-, insulins • A10BA-, MET • A10BB-, SU • A10B-, other AD; • A10X-, other drugs used in DM

Treatment cost (12 months preceding index date		
Covariate	**Definition (when applicable)**	**ATC/ ICD-10**
Inpatient cost	Associated to main or primary diagnosis ICD-10 codes: • any visit • related to T2DM • related CV	• Any ICD-10 code • E11 oder E16 • I11.0, I13.0, I13.2, I21, I25, I50 or I63
Outpatient cost	Total costs based on EBM outpatient settings for outpatient visits associated to primary diagnosis ICD-10 codes related to • any visit • T2DM • CV	• Any ICD-10 code • E11 oder E16 • I11.0, I13.0, I13.2, I21, I25, I50 or I63
Rehabilitation cost	Associated to main or primary diagnosis ICD-10 codes: • any visit • related to T2DM • related CV	• Any ICD-10 code • E11 oder E16 • I11.0, I13.0, I13.2, I21, I25, I50 or I63
Cost of remedies / aids	Costs based on prescribed aids and remedies: TD2M related: needles, lancets and strips (identified by respective PZNs)	-/-
Medication cost		
AD cost	Medication costs of prescribed antidiabetic drug based on list prices	A10
CV drug cost	Medication costs based on pharmaceutical list prices for the following therapies: • Antithrombotic agents • Beta blocking agents • Calcium channel blockers • RAAS inhibitors • lipid modifying agents	• B01 • C07 • C08 • C09 • C10
All other drug cost	Any other agent	Any ATC except the ones listed above

AD=Antidiabetics; ATC=Anatomical Therapeutic Chemical; CABG=coronary artery bypass grafting; CHF=congestive heart failure; CKD=chronic kidney disease; COPD=chronic obstructive pulmonary disease; CV=cardiovascular; DM=diabetes mellitus; ICD-10=International Classification of Diseases, 10th revision; MET=metformin; MI=myocardial infarction; PTCA=percutaneous transluminal coronary angioplasty; SU=sulfonylureas.

Annex 2. CHARLSON-COMORBIDITY-SCORE (CCI) AND ITS COMPONENTS (EXCL: AGE FACTOR)

COMORBIDITY	ICD-10 CODE	CHARLSON SCORE
Coronary artery disease	*I20-, I21-, I22-, I23-, I24-, I25-*	1
Congestive heart failure	*I11-, I50-*	1
Peripheral vascular disease	*I73-, I74-, I77-*	1
Cerebrovascular disease	*G45-, G46-, I6-*	1
Dementia	*F00-, F01-, F02-, F03-, G30-*	1
Chronic pulmonary disease	*J4-, J6- w/o J67-, J68-, J69-*	1
Connective tissue disorder	*M05-, M06-, M07-, M08-, M3-*	1
Peptic ulcer disease	*K25-, K26-, K27-, K28-*	1
Mild liver disease	*B18-, K70-, K73-, K75-*	1
Diabetes mellitus without complications	*E109-, E119-, E129-, E139-, E149-*	1
Hemiplegia	*G81-, G82-*	2
Moderate or severe renal disease	*N17-, N18-, N19-*	2
Diabetes mellitus with end-organ damage	*E10-, E11-, E12-, E13-, E14-*	2
Tumor without metastases, leukemia, lymphoma, multiple myeloma	*C-, w/o metastatic solid tumor [C77-, C78-, C79-, C80-]*	2
Moderate or severe liver disease	*K72-, K74-, I85-*	3
Metastatic solid tumor	*C77-, C78-, C79-, C80-*	6
AIDS	*B20-, B21-, B22-, B23-, B24-*	6