Observational Study Protocol MB102134

OBSERVATIONAL SINGLE-COHORT DATA BASE STUDY OF DAPAGLIFLOZIN UTILISATION IN EUROPE

BMS Study Director

[redacted]

AstraZeneca Study Director

[redacted]

External Investigator

[redacted]

Bristol-Myers Squibb Research and Development

[redacted]

&

AstraZeneca Pharmaceuticals LP

[redacted]

SYNOPSIS

Observational Study Protocol MB102134

Protocol Title: Observational Single-Cohort Data Base Study of Dapagliflozin Utilisation in Europe

Department: Bristol-Myers Squibb Epidemiology and AstraZeneca Global Epidemiology

Objective(s): ***Primary objective:*** To describe the characteristics of European patients newly prescribed dapagliflozin by age, sex, dapagliflozin dose, country, selected co-morbidities, and selected concomitant medications.

The proposed drug utilization study will specifically describe dapagliflozin use in:

- patients > 75 years-old,
- combination use with loop diuretics or pioglitazone,
- patients with a known history of moderate or severe renal impairment and in kidney failure,
- patients lacking a diagnostic code indicating type 2 diabetes.

Study Design: This is an observational single-cohort data base study with descriptive data analyses among patients receiving dapagliflozin within electronic medical records in Europe. The study will describe the utilization pattern of dapagliflozin during the first 3.5 years after marketing authorization and launch in Europe specifically in Belgium, France, Germany, Italy, Spain and United Kingdom.

Study Population: All patients identified in the database(s) who newly received at least one dapagliflozin prescription during the study period. Dapagliflozin use will be representative of the uptake of dapagliflozin in the respective countries.

Data Collection Methods: Cegedim Strategic Data (CSD) Longitudinal Patient Databases (LPDs) come directly from physicians' EMR. Participating physicians use CSD-provided software to record their daily patient interactions and data are transmitted regularly to the Cegedim coordinating center where they are cleaned and de-identified.

Data Sources: CSD's LPDs include data from 5.8 million active patients. Over 4,000 general practitioners (GPs) in six European countries (Belgium, France, Germany, Italy, Spain and United Kingdom) are part of the LPD panels.

Patient metrics in the LPDs include Demographics (Year of birth, gender, registration dates), Medical History (Event dates, diagnosis, symptoms, risk factors, co-morbidities, referrals), Prescription (Rx dates, indication, therapeutic class, molecule, brand, dosage, posology, duration), Clinical Data (Height, weight, BP, life habits) and laboratory data.

Exposures: A patient will be identified as newly exposed when he/she has at least one dapagliflozin prescription recorded in the study database(s).

Outcomes: Outcomes include but are not limited to:

- Patient demographics: age, sex, country,
- Baseline history of type 2 diabetes,
- Baseline history of moderate or severe renal impairment,
- Concomitant medications at baseline and during dapagliflozin use.

Follow-up: Follow-up will begin on the date a patient is first prescribed dapagliflozin, which will represent the beginning of follow-up (i.e., "index date"). Follow-up will continue until the end of study or discontinuation of dapagliflozin (i.e., the final day of the days' supply for the last prescription for dapagliflozin). Follow-up will only be used to determine combination use with loop diuretics or pioglitazone. The baseline period is defined as the 12 months in the database preceding the index date.

Data Analyses: Descriptive analyses of the data will be conducted. The demographic and clinical characteristics of patients identified to have received a dapagliflozin prescription will be described at the index date. The proportion of patients receiving the drug by age group, sex, selected co-morbidities, selected concomitant medications and frequency of laboratory testing will be calculated. We will also describe dapagliflozin use within and outside the

labeled indication of type 2 diabetes. Descriptive statistics will be calculated within each database by year, and trends over time will be described. Categorical variables will be summarized by frequencies and proportions, and continuous variables will be summarized by means and standard deviations or medians and interquartile ranges.

Sample Size/Power: This is a descriptive study that will quantify new dapagliflozin use according to selected patient characteristic. Estimates of exposure, based on commercial forecasting, were performed to describe the numbers of patients that may be included in the study databases.

Estimated Annual Number of Dapagliflozin Users in the Cegedim Databases

	Number of Type II Diabetics, 2010	"% of T2DM Patients Receiving OAD Treatment"	"Estimate Annual Number of OAD Users"	2012	2013	2014	2015	2016	2017	Total
France	65000	82.7%	53755	-	269	753	1344	2957	3924	9246
UK	85000	57.2%	48620	49	146	292	535	778	924	2723
Italy	45000	64.5%	29025	-	261	319	435	784	929	2728
Germany	70000	55.3%	38710	77	387	542	735	968	1161	3871
Belgium	15000	71.9%	10785	11	86	129	183	313	388	1111
Spain	18000	52.8%	9504	-	105	95	86	162	200	646
TOTAL				137	1254	2130	3318	5960	7526	20325

Limitations/Strengths: Difference in the availability of data variables across the databases; misclassification of exposure and outcome; and available sample sizes (due, in part, to dapagliflozin uptake) are the major potential limitations of this study.

TABLE OF CONTENTS

1 INTRODUCTION

Dapagliflozin (BMS-512148) is a highly potent, selective, and reversible inhibitor of the human renal sodium glucose co-transporter (SGLT2), the major transporter responsible for renal glucose reabsorption. Dapagliflozin lowers plasma glucose by inhibiting the renal reabsorption of glucose, and by promoting its urinary excretion making it a member of an emerging therapeutic class in the treatment of type 2 diabetes mellitus (T2DM).

Dapagliflozin is indicated in adults aged 18 years and older with type 2 diabetes mellitus (T2DM) to improve glycaemic control as:

- Monotherapy when diet and exercise alone do not provide adequate glycaemic control in patients for whom metformin is considered inappropriate due to intolerance.
- Add-on combination therapy in combination with other glucose lowering medicinal products including insulin, when these, together with diet and exercise, do not provide adequate glycaemic control.

According to the summary of product characteristics (SmPC), the recommended posology and method of administration are the following:

- The recommended dose is 10 mg dapagliflozin once daily for monotherapy and add-on combination therapy with other glucose lowering medications including insulin. When dapagliflozin is used in combination with insulin or an insulin secretagogue, such as a sulphonylurea, a lower dose of insulin or insulin secretagogue may be considered to reduce the risk of hypoglycaemia.
- Dapagliflozin can be taken orally once daily at any time of day with or without food. Tablets are to be swallowed whole.

The following precautions for use must be taken for the following special populations:

- Renal impairment: The efficacy of dapagliflozin is dependent on renal function, and efficacy is reduced in patients who have moderate renal impairment and likely absent in patients with severe renal impairment. Dapagliflozin is not recommended for use in patients with moderate to severe renal impairment (patients with creatinine clearance [CrCl] < 60 ml/min or estimated glomerular filtration rate [eGFR] < 60 ml/min/1.73 m^2). No dosage adjustment is indicated in patients with mild renal impairment.
- Hepatic impairment: No dosage adjustment is necessary for patients with mild or moderate hepatic impairment. In patients with severe hepatic impairment, a starting dose of 5 mg is recommended. If well tolerated, the dose may be increased to 10 mg.
- Elderly (≥ 65 years): In general, no dosage adjustment is recommended based on age. Renal function and risk of volume depletion should be taken into account. Due to the limited therapeutic experience in patients 75 years and older, initiation of dapagliflozin therapy is not recommended.
- Patients at risk for hypotension due to effect of dapagliflozin on diuresis and blood pressure: Caution should be exercised in patients for whom a dapagliflozin-induced drop in blood pressure could pose a risk, such as patients with known cardiovascular disease, patients on anti-hypertensive therapy with a history of hypotension or elderly patients. For patients receiving dapagliflozin, in case of intercurrent conditions that may lead to volume depletion,

careful monitoring of volume status (e.g. physical examination, blood pressure measurements, and laboratory tests including haematocrit) and electrolytes is recommended. Temporary interruption of treatment with dapagliflozin is recommended for patients who develop volume depletion until the depletion is corrected.

1.1 Study Rationale

This drug utilization study (DUS) is being conducted as part of the BMS/AZ Dapagliflozin Risk Management Plan. Per regulatory request, this study is being conducted to describe the patients using dapagliflozin in usual clinical practice in Europe, specifically the countries reporting to our data source: Belgium, France, Germany, Italy, Spain and United Kingdom. The selection of the Cegedim Strategic Data was made because it includes data across 6 different European countries. However, an analysis of dapagliflozin patients not only required a European approval but it also requires reimbursement at the country level. As of November 2013, reimbursement was achieved in the United Kingdom and was not achieved in Belgium. In Germany, dapagliflozin was only reimbursed for 2013. Reimbursement in Italy and Spain has yet to be determined and in France, dapagliflozin is not yet reimbursed but it is expected that, should we be reimbursed, it would be only for specialist use. Since the Cegedim database includes only general practitioners, we would not expect to see any dapagliflozin users from France. Therefore, at this time, sensitivity analyses around missing data will be limited to discussions of the UK data. As this study will be on-going through 2017, country reimbursement status may change potentially resulting in more useable data from the Cegedim database.

1.2 Research Question

Research question 1: What are the baseline characteristics of the patients prescribed dapagliflozin in Europe?

Research question 2: What proportion of patients prescribed dapagliflozin have baseline moderate to severe renal impairment?

Research question 3: What proportion of patients prescribed dapagliflozin are 75 years of age or older at the time of the index prescription?

Research question 4: What proportion of patients prescribed dapagliflozin are also users of loop diuretics or pioglitazone during the baseline period and the available follow-up period?

Research Question 5: What proportion of patients prescribed dapagliflozin does not have a diagnosis of Type II Diabetes during the baseline period or on the index date?

2 OBJECTIVES

2.1 Primary Objectives

To describe the characteristics of European patients newly prescribed dapagliflozin by age, sex, dapagliflozin dose, country, selected co-morbidities, and selected concomitant medications.

The proposed drug utilization study will specifically describe dapagliflozin use among all new dapagliflozin users including dapagliflozin users in the following subgroups (separate analysis for each subgroup):

- patients > 75 years-old,
- combination use with loop diuretics or pioglitazone,
- patients with a known history of moderate or severe renal impairment and in kidney failure,
- patients not reported to have T2DM.

2.2 Secondary Objectives

Not applicable.

2.3 Exploratory Objectives

Not applicable.

3 STUDY DESIGN

3.1 Overview of Study Design

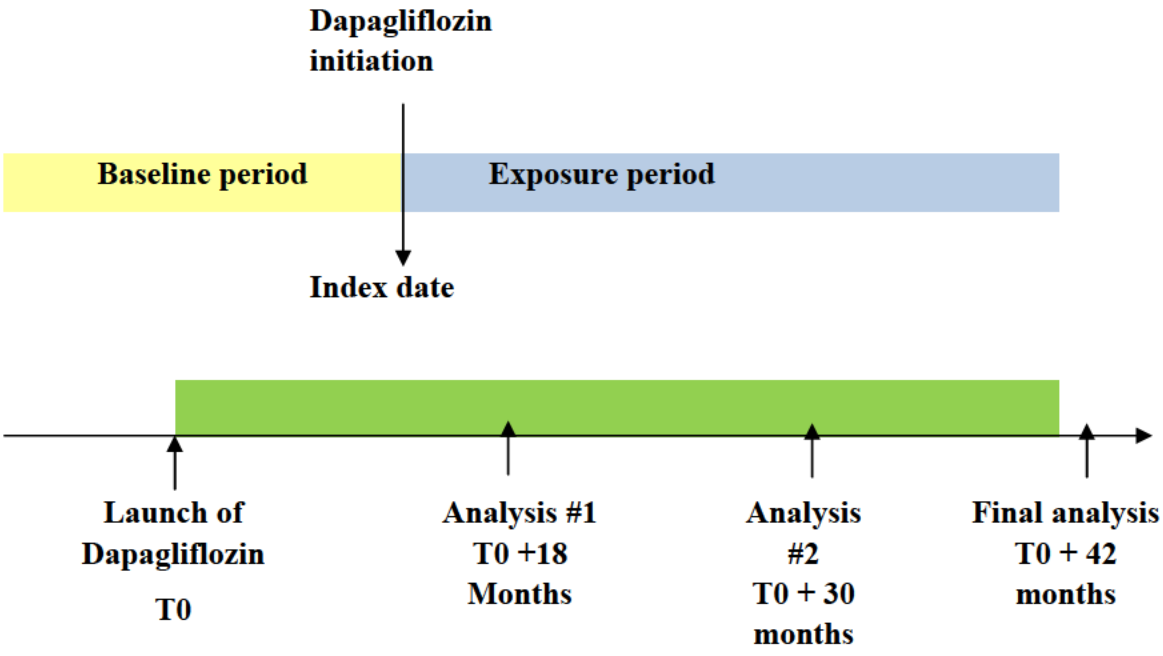
This is a cross-sectional observational single-arm cohort study with descriptive data analyses among patients receiving dapagliflozin within the Cegedim Strategic Data (CSD) Longitudinal Databases (LPDs) electronic medical records (EMRs) in Europe. The assessment of combination use with loop diuretics or pioglitazone will incorporate longitudinal follow-up (See Figure 3.1-1). The study will describe the utilization pattern of dapagliflozin during the first three and a half years after marketing authorization and launch in Europe, specifically the countries reporting to this data source: Belgium, France, Germany, Italy, Spain and United Kingdom. This analysis will be repeated at 18, 30 and 42 months from the launch of dapagliflozin (T0, estimated as January 1, 2013).

The "prescription index date" for each patient included in the study will be defined as the date a patient is first prescribed dapagliflozin.

Follow-up of Subjects

The follow-up period will be used to identify concomitant use of dapagliflozin with loop diuretics or pioglitazone. Follow-up will begin on the date a patient is prescribed dapagliflozin, which will represent the beginning of follow-up (i.e., "index date"). Follow-up will continue until the discontinuation of dapagliflozin (i.e., the final day of the days' supply for the last prescription for dapagliflozin). Once a new user of dapagliflozin discontinues his/her treatment at any time (could also include an interruption in treatment), he/she will not be identified for future analyses.

Figure 3.1-1: Study Period



3.2 Study Population

The study population will be new users of dapagliflozin who meet the inclusion and exclusion criteria noted below. For research question 1-3 and 5 only, cross sectional data in new users of dapagliflozin will be used and therefore treatment interruptions are not of concern in these analyses. For research question 4, a follow-up will be performed and treatment interruptions will be dealt with as described in Section 3.1, *Follow-up of Subjects*.

Eligible study subjects will be selected from the study databases (described in section 3.2.1 and Figure 3.1-1). All subjects meeting inclusion and exclusion criteria and prescribed dapagliflozin within each database will be selected for inclusion for each analysis dataset:

- Patients newly prescribed dapagliflozin between T0 and 18 months meeting the inclusion criteria will be included in analysis 1 dataset.
- Patients prescribed dapagliflozin between T0 and 30 months meeting the inclusion criteria will be included in analysis 2 dataset. This second analysis dataset will include patients newly treated with dapagliflozin between 19 and 30 months and patients from analysis 1 still on treatment with dapagliflozin during the analysis 2 time period. The existing patients from analysis 1 will only contribute to the follow-up period (see section 3.1 for follow-up definition).
- Patients prescribed dapagliflozin between T0 and 42 months meeting the inclusion criteria will be included in analysis 3 dataset. This third cohort will include patients newly treated with dapagliflozin between 30 and 42 months and patients from analysis 2 still on treatment with dapagliflozin during the analysis 3 time period. The existing patients from analysis 2 will only contribute to the follow-up period (see section 3.1 for follow-up definition).

3.2.1 Inclusion Criteria

1) Enrolled in the Cegedim databases for at least one year prior to the first prescription of dapagliflozin (baseline time period),
2) All patients identified in the database who newly received at least one dapagliflozin prescription during the study period.

3.2.2 Exclusion Criteria

Not applicable.

3.3 Data Source/Data Collection Process

Data Collection Process

This study requires data sources that longitudinally capture patient demographics, prescription information, diagnosis codes, and available laboratory data.

Data Source

The Longitudinal Patient Databases (LPDs) owned by Cegedim Strategic Data include data from 5.8 million active patients in Belgium, France, Germany, Italy, UK and Spain. In the UK, the database is known as The Health Improvement Network (THIN). For this study, data from all six databases/countries will be used assuming dapagliflozin is prescribed in the country.

Physicians using Cegedim proprietary software (Appendix 1) for management of their practice are asked to take part in LPDs. If the practice agrees to participate, they are asked to provide additional services from Cegedim and in return received a discount on the price for their software subscription. Nationally representative panel in terms of age, gender and geographical region are extracted from the pool of all participating physicians using the quota sampling method. All LPD data are strictly anonymous and there is no direct interaction possible with the physician, who is totally free to prescribe any product he/she deems necessary for his/her patient.

The LPDs collect medical information from proprietary practice management software used by the physician during patients' office visits for recording their daily patient interactions in electronic medical records. In each country, a panel of physicians using this software volunteer to make available anonymized, patient-level information from their practices for clinical research purposes. Since these data are being collected in a non-interventional way, they reflect routine clinical practice in these countries. The panel of contributing physicians is maintained as a representative sample of the primary care physician population in each country according to age, sex, and geographical distribution. Whenever a physician leaves the panel, he/she is replaced by another one with a similar profile. Additionally, in most countries, the patient population is representative of the respective country population according to age and sex distribution, as provided by national statistic authorities (Table 3.3-1).

Table 3.3-1: Physician and Patient Populations in the CEGEDIM GP's LPD by Country (2012)

	Belgium	France	Germany	Italy	UK	Spain
Number of physicians in the panel	300	1200	550	700	1780	190
Average number of patients who consulted GP at least once in a year	360,000	1,800,000	620,000	700,000	3.200,000	130,000
% of the national population (Eurostat :http://epp.eurostat.ec.europa.eu)	3.24 %	2.76%	0.8%	1.18%	5.04%	0.28%

3.3.1 Content of the Cegedim LPDs

Data are entered during usual patient care and submitted daily to the Cegedim coordinating center, cleaned and de-identified and then made available in the LPDs for research no later than one month after submission. The following anonymized patient data collected from GPs' practice will be extracted from the databases:

- Demographics (Year of birth, gender, registration dates),
- Medical History (Event dates, diagnosis, symptoms, risk factors, co-morbidities, referrals,
- Treatments (indication, molecule/brand, dosage, posology, date and length of prescription),
- Clinical Data (Height, weight, BP, life habits),
- Clinical laboratory tests, X-ray and other investigations (not available in Germany).

The median follow-up duration of a patient in the data base is 27 months in Germany, 36 months in France, 44 months in Spain, 46 months in Belgium, 112 months in Italy and 117 months in UK.

Patient data collected by Cegedim in each country participating in the LPD varies to some extent to accommodate local needs. However, all countries collect data on medical co-morbidities and outcomes, prescriptions, demographics, and physician characteristics. Key metrics of the GPs' LPDs are provided for each country in Appendix 2 (2012 data).

In the LPDs, diagnoses of clinical events are recorded as diagnostic codes (*e.g.* ICD-9, READ). These codes from the different participating countries will be harmonized based on pre-specified algorithms developed prior to the analysis and listed in the statistical analytic plan. Table 3.3.1-1 summarizes the coding conventions for each of the databases.

Table 3.3.1-1: Data Coding Conventions in the Different Countries

	Belgium	France	UK	Italy	Germany	Spain
Drug code dictionary	Association Pharmacologique Belge	Claude Bernard	Multilex	Farmadati	Abdata	Vademecum
Therapy classification	European Pharmaceutical Marketing Research Association (EphMRA)		British National Formulary (BNF)	ATC	ATC	ATC
Disease classification	Cegedim proprietary thesaurus (mapped to ICD-10)		Read Codes	ICD-9	ICD-10	CIAP mapped to ICD-9

ATC = Anatomic, Therapeutic, Chemical
ICD = International Classification of Disease
CIAP = International Classification of Primary Care

3.3.2 Content of the Cegedim LPDs for T2DM patients

Table 3.3.2-1 gives approximate estimates of the proportion of T2DM patients captured using the Cegedim LPD in each country.

The total number of T2DM patients in the respective countries during the same period is not known, but may be roughly estimated based on information from the Diabetes Atlas 2012[1] and assuming that this population represents approximately 90 % of the total diabetic population.

Table 3.3.2-1: Estimates of the proportion of T2DM patients captured in Cegedim LPDs (year 2012)

	Source	France	Germany	Italy	Spain	Belgium	UK
Diabetic Population (in thousands)	Diabetes Atlas http://www.idf.org/diabetesatlas	3,493	5,249	3,903	3,249	547	3,336
Estimates of T2DM patients Total Population (in thousands)	0.9 x Diabetic population	3,115	4,724	3,523	2,925	493	3,002
T2DM patients in CegedimLPDs (in thousands)	Cegedim LPD	65	37	42	14	13	118
% of T2DM population in Cegedim LPDs:	T2DM patients in Cegedim LPDs/Estimates of T2DM patients Total Population	2.1%	0.8%	1.2%	0.5%	2.6%	3.9%

Key metrics for the T2DM patients in GP's LPDs are provided for each country in Table 3.3.2-2.

Information collected during the course of one year (2012), on serum creatinine and eGFR for T2DM patients is detailed in Table 3.3.2-3.

These data, when insufficiently documented in the course of one year, can be extracted over an extended period of time (two years) in patients' medical record in order to add information and increase possibilities for meaningful analyses.

In addition, this information can be complemented in each country by a search of patient's medical records with a list of diagnostic codes identifying chronic kidney disease, renal failure, end-stage renal disease and related diagnoses (according to Wan J. et al[2] and Denburg M.R., et al[3].)

Table 3.3.2-2: Key Data Metrics of T2DM Patients in GP's LPDs

	France	ITALY	UK	GERMANY	SPAIN	BELGIUM
GENERAL DATA						
Number of patients having visited a physician	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Of which Male	58,0%	51,9%	55,2%	51,8%	53,8%	53,4%
Female	42,0%	48,1%	44,8%	48,2%	46,2%	46,6%
BIOMETRIC DATA						
Number of patients with a weight recorded (1 year history)	56,0%	39,9%	83,5%	29,1%	28,3%	37,3%
Number of patients with height recorded (1 year history)	20,6%	39,6%	30,8%	28,3%	23,2%	13,9%
Number of patients with BMI recorded (1 year history)	20,2%	39,4%	29,0%	28,2%	20,7%	13,2%
Number of patients with BMI recorded (3 years history + calc)	40,0%	62,7%	54,0%	32,1%	77,5%	24,4%
Number of patients with waist recorded (1 year history)	1,2%		0,4%			
Number of patients with Blood Pressure recorded (1 year history)	64,8%	54,5%	91,6%	28,2%	40,9%	73,4%
average records by year by patient	*3,34*	*2,73*	*2,80*	*5,72*	*2,44*	*4,42*
BIOLOGICAL EXAMS						
Number of patients with a fasting blood glucose test	29,5%	67,7%	45,6%	54,5%	44,1%	68,5%
Number of patients with a Hba1c rate	43,2%	66,4%	88,0%	77,1%	27,5%	53,6%
Number of patients with Total cholesterol	26,7%	61,2%	83,8%	57,2%	0,0%	56,4%
Number of patients with LDL	26,8%	55,4%	55,9%	54,1%	26,5%	53,1%
Number of patients with HDL	26,0%	58,3%	74,7%	52,8%	27,7%	55,9%
Number of patients with Triglycerid	27,2%	60,3%	63,9%	56,6%	41,8%	56,1%

Table 3.3.2-3: Information collected on serum creatinine and eGFR for T2DM patients in Cegedim LPDs during the course of one year (year 2012)

	FRANCE		GERMANY		ITALY		SPAIN		BELGIUM		UK	
Number of patients in the panel	1 800 000		620 000		700 000		130 000		360 000		3 200 000	
T2DM patients (% of panel population)	65 000	4%	37 000	6%	42 000	6%	14 000	11%	13 000	4%	118 000	4%
T2DM patients:												
with creatinine values recorded	20 500	32%	26 600	72%	25 000	60%	5 700	41%	7 350	57%	105 000	89%
with eGFR values recorded or calculated*	12 300	20%	19 300	52%	20 850	50%	4 200	30%	1 100	9%	93 300	79%

*eGFR estimated using the Cockcroft-Gault (Appendix 5) or MDRD estimation formula (Appendix 4)

3.4 Definitions of Study Variables

Demographic, medical history, treatment, clinical and clinical laboratory data on patients included in this study will be collected from each database.

3.4.1 Outcomes/Endpoint Variables

Table 3.4.1-1 includes the characteristics and definitions of how dapagliflozin users will be categorized and described.

Table 3.4.1-1: Categorization of Dapagliflozin Users

Characteristic	Definition
Patient Demographics, at initiation of dapagliflozin use:	
Age categories	< 45, 45-59, 60-74, >= 75
Sex	Male or Female
Country	Belgium, France, Germany, Italy, Spain, United Kingdom
Concomitant medications before, at initiation of and during dapagliflozin use[1]	Loop diuretics (ATC code C03Cx) Pioglitazone (Actos, Glustin) (ATC code A10BG), Biguanides (ATC Code A10BA) Sulfonamide derivates (ATC code A10BB and A10BC) Dipeptidyl peptidase 4 inhibitors (DDP4) (ATC code A10BH) Glucagon-like peptide 1 (GLP-1, exenatide, liraglutide, ATC code A10BX), Alpha glucosidase inhibitors (ATC code A10BF). Insulin (ATC code A10A)
Concomitant co-morbidities, at initiation of dapagliflozin use:	
Baseline history of moderate renal impairment	CrCl or eGFR[2] value between 30-60 within 12 months prior to the dapagliflozin dispensing
Baseline history of severe renal impairment	CrCl or eGFR value < 30 within 12 months prior to the dapagliflozin dispensing
Baseline history of renal failure	Diagnosis codes for end stage renal disease or dialysis (see Appendix 3)
Congestive heart failure	Diagnosis codes for congestive heart failure (see Appendix 3)
Hypertension	Diagnosis codes for hypertension (see Appendix 3)
Type 2 Diabetes Mellitus, at initiation of dapagliflozin use:	Diagnostic code for T2DM within 12 months prior to the dapagliflozin dispensing (see Appendix 3) OR Prescription for an antiglycemic medication within 12 months prior to the dapagliflozin dispensing

Table 3.4.1-1: Categorization of Dapagliflozin Users

Characteristic	Definition
Dapagliflozin Dose, at initiation of dapagliflozin use:	10 mg 5 mg
Off-Label Use	Patients meeting one of the following criteria: • 17 years old or younger on index date • 75 years of age or older on index date • No diagnostic codes for Type 2 Diabetes Mellitus, at or before the initiation of dapagliflozin use (see definition above) • Baseline history of moderate renal impairment (see definition above) • Baseline history of severe renal impairment (see definition above) • Baseline history of renal failure (see definition above)

[1] Pioglitazone sales are suspended in France and Germany

[2] eGFR will be estimated by the creatinine clearance calculated by means of the Cockroft & Gault formula if it is not available in the data base.

Estimated creatinine clearance (CrCl) will be calculated by means of the Cockcroft-Gault formula if it is not available in the database. (Appendix 5)

3.4.2 Exposure/Independent Variables of Interest

For eligible patients, use of dapagliflozin will be defined by the date of first dapagliflozin prescription in the database.

3.4.3 Other Co-variates/Control Variables

Not applicable.

4 STATISTICAL ANALYSIS

4.1 Statistical Analysis Methods

The details of the data analysis will be documented in a Statistical Analysis Plan (SAP) document, which will be finalized before the first extraction of data from the databases. The SAP will elaborate on the analytic plans outlined in the protocol; any major modifications will be reflected in a protocol amendment.

The goal of the analyses is to describe the characteristics of patients who use dapagliflozin.

Since the study is purely descriptive, missing data will not be imputed but the amount of missing data will be reported for each variable.

4.1.1 Analysis Plan for Primary Objective

Descriptive statistics will be calculated to describe baseline characteristics in dapagliflozin initiators. These characteristics include age group, sex, dapagliflozin dose, country, selected co-morbidities, selected concomitant medications, frequency of laboratory testing, and available results of laboratory testing (see Table 3.4.1-1).

We will also describe dapagliflozin use within and outside (off-label use) the labeled indication of type 2 diabetes.

Categorical variables will be summarized by frequencies and proportions, and continuous variables will be summarized by means and standard deviations or medians and interquartile ranges. Because this is a descriptive drug utilization study, formal comparative tests of significance will not be performed.

4.1.1.1 Missing Data on Serum Creatinine and eGFR

In order to assess the risk of informative missingness the impact of missing key variables (e.g. BMI and eGFR) will be checked by describing patients with and without missing values respectively regarding basic characteristics available for all or most patients including age, gender, country of residence, co-medication and co-morbidity.

However, all analyses requiring information on serum creatinine and eGFR will be based primarily on information from the THIN database in the UK and from the German LPD database, which are the best documented.

Analyses will be performed also for the other countries covered by this study but the interpretation will take into account the loss of information. (See section 4.1.1.2: Sensitivity analysis.)

4.1.1.2 Sensitivity analysis

In order to assess the effect of excluding patients prescribed dapagliflozin but not included because of enrollment less than one year before index date, these patients will be counted and their main characteristics at index date (age, gender, eGFR) will be described together with the characteristics of patients included in the study.

To study patients that may be prescribed dapagliflozin in a specialist setting (such as patients with kidney disease and/or patients older than 75 years, paediatric patients), we will primarily use the UK THIN database in which references to a specialist are recorded.

4.1.2 Analysis Plan for Secondary Objectives

Not applicable.

4.1.3 Analysis Plan for Exploratory Objectives

Not applicable.

4.2 Power/Sample Size

This is a descriptive study that will quantify dapagliflozin use according to selected patient characteristics. Estimates of projected dapagliflozin exposure, based on commercial forecasting, are shown in Tables 4.2-1 and 4.2-2 below.

Table 4.2-1: Estimated Annual Number of Dapagliflozin Users in the Cegedim Databases

				Estimated Annual Number of Dapagliflozin Users[1]						
	Number of Type II Diabetics, 2010[2]	**% of T2DM Patients Receiving OAD Treatment[3]**	**Estimate Annual Number of OAD Users**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**Total**
France	65000	82,70%	53755 [53566 ; 53944]	-	269 [236 ; 302]	753 [697 ; 809]	1344 [1268 ; 1420]	2957 [2843 ; 3071]	3924 [3792 ; 4056]	9246 [8836 ; 9658]
UK	85000	57,20%	48620 [48246 ; 48903]	49 [35 ; 63]	146 [121 ; 170]	292 [256 ; 327]	535 [486 ; 583]	778 [718 ; 837]	924 [858 ; 988]	2723 [2474 ; 2968]
Italy	45000	64,50%	29025 [28777 ; 29224]	-	261 [228 ; 295]	319 [282 ; 356]	435 [391 ; 479]	784 [723 ; 843]	929 [862 ; 994]	2728 [2486 ; 2967]
Germany	70000	55,30%	38710 [38363 ; 38968]	77 [60 ; 95]	387 [345 ; 428]	542 [492 ; 591]	735 [676 ; 793]	968 [899 ; 1035]	1161 [1085 ; 1235]	3871 [3557 ; 4177]
Belgium	15000	71,90%	10785 [10658 ; 10893]	11 [4 ; 17]	86 [67 ; 105]	129 [106 ; 153]	183 [155 ; 212]	313 [275 ; 350]	388 [346 ; 430]	1111 [953 ; 1267]
Spain	18000	52,80%	9504 [9323 ; 9635]	-	105 [83 ; 126]	95 [74 ; 115]	86 [66 ; 105]	162 [134 ; 189]	200 [169 ; 230]	646 [526 ; 765]
			TOTAL	**137 [99 ; 175]**	**1254 [1080 ; 1426]**	**2130 [1907 ; 2351]**	**3318 [3042 ; 3592]**	**5962 [5592 ; 6325]**	**7526 [7112 ; 7933]**	**20325 [18832 ; 21802]**

[1] Dapagliflozin uptake was estimated from AZ forecasting models; country-specific estimates were not available for Belgium. The overall EU5 update was applied to Belgium. Please see Table 4.2-2 for more information.

[2] Data extracted from the Diabetes Atlas, 3rd Edition, http://da3.diabetesatlas.org/index-2.html - When multiple estimates were presented, the smallest proportion was chosen.

[3] OAD use alone or in combination with insulin.

Table 4.2-2: Estimated Percentage Uptake of Dapagliflozin by Country[1]

	Dapagliflozin Uptake - Percentage of T2D Treated Patients					
	2012	2013	2014	2015	2016	2017
France	N/D	0.5%	1.4%	2.5%	5.5%	7.3%
German	0.2%	1.0%	1.4%	1.9%	2.5%	3.0%
Italy	N/D	0.9%	1.1%	1.5%	2.7%	3.2%
Spain	N/D	1.1%	1.0%	0.9%	1.7%	2.1%
UK	0.1%	0.3%	0.6%	1.1%	1.6%	1.9%
UE5	0.1%	0.8%	1.2%	1.7%	2.9%	3.6%

[1] Dapagliflozin uptake was estimated from AZ forecasting models.

4.3 Milestones

Descriptive statistics will be conducted 18 months after dapagliflozin has been on the market (T0 = December 2012), and annually thereafter through 3.5 years. These analyses will be shared with the CHMP. The proposed drug utilization study can be extended beyond 3.5 years if the use of dapagliflozin in the Cegedim database is less than what is currently projected; the decision to extend the study would be made in collaboration with the CHMP and would be formalized in a protocol amendment.

5 STUDY LIMITATIONS/STRENGTHS

5.1 Limitations Linked to Data Sources

As the study is based on fully anonymized real life, electronic medical records data, there are limitations inherent to the impossibility to link data from different databases and to get additional information from the physicians.

Data are collected by GPs: Even if some patients with type 2 diabetes mellitus may be treated exclusively by specialists, most of them will be treated by their GP at least for the treatment maintenance period. This study will not capture any use of dapagliflozin that is prescribed by specialists in Europe.

Data are collected in routine care: Cegedim's LPDs collect real life clinical practice information from the actual medical patients' records. A study specific CRF will not be used to collect supplemental information not entered in the medical records as part of the patient's routine clinical care. Also, as a non-interventional study, this protocol does not mandate any study specific visits or procedures.

Potential for missing data: No individual patient identifiers will be available. It is therefore impossible to query the physicians providing the data for any missing information.

Data not recorded correctly in the databases: Recording of the indication of each prescribed treatment is mandatory in the Cegedim Software but the GP is free to enter any diagnosis and can for instance enter the reason of visit (*e.g.* flu) as indication for all treatments prescribed at the visit, and an oral antidiabetic drug may thus be recorded with an indication 'flu'.

Incomplete data: Clinical laboratory results are entered by the GPs whenever they deem these results relevant. The patient's renal function is therefore uncertain in the absence of recent creatinine. Similarly, weight may not be recorded by the physician at each consultation, which may impact the calculation of creatinine clearance calculation.

Absence of systematic recording of referrals to specialists: Except for the UK database, there is no systematic recording of referrals. If dapagliflozin is a secondary care initiation, this will not be recorded, leading to misclassification bias for newly initiated patients.

5.2 Selection Bias

In order to more precisely describe the characteristics of European patients prescribed dapagliflozin (specifically baseline renal impairment), a minimum of 12 months (365 days) of medical history in the LPD is required prior to index date. Health care utilization patterns are best described when they include data from all potential prescribers of the drug. In this instance, the LPD data source does not capture prescriptions written in the specialist setting/hospitals; therefore, selection bias is possible if GPs prescribe dapagliflozin to a different patient population than a physician in a specialist setting. In the UK, the GP is a 'gatekeeper' of information for the patient and thus may be aware of and report health care visits or events taking place outside the GP setting. Thus, in the UK this selection bias may be less pronounced and UK THIN database can be used to estimate that selection bias in other databases.

5.3 Misclassification Bias

Misclassification bias can result if study subjects are not categorized correctly with regards to exposure or selected patient characteristics. We expect minimal misclassification with respect to exposure, since this will be determined from each database's prescribing records. However, actual adherence to dapagliflozin or other ADs cannot be confirmed. Further, misclassification as to whether the patient is a new initiator could exist (1) if providers supplied samples of dapagliflozin for varying duration to patients, at no cost, and with no record in the database and (2) if dapagliflozin was initiated by a specialist. This will vary by country and database, and could result in varying results across countries.

5.4 Generalizability

The study results will be generalizable to patients meeting the inclusion and exclusion criteria. However, by including data from the several countries in E.U., the study has the potential to maximize the populations to which these findings can be generalized.

6 STUDY CONDUCT

This study will be conducted in accordance with International Society for Pharmacoepidemiology (ISPE) Guidelines for Good Pharmacoepidemiology Practices (GPP) and applicable regulatory requirements.

6.1 Ethics Committee Review and Informed Consent

6.1.1 Ethics Committee Review

The investigator must ensure that the required approvals from Ethics Committees, Independent Review Committees, Regulatory Authorities, and/or other local governance bodies are obtained before study initiation at the site.

6.1.2 Informed Consent

This study does not require that informed consent is obtained from patients.

6.2 Responsibilities within the Study

The study shall be conducted as described in this approved protocol. All revisions to the protocol must be discussed with, and be prepared by BMS/AZ. The authors should not implement any deviation or change to the protocol without prior review.

6.3 Confidentiality of Study Data

The confidentiality of records that could identify subjects within the database must be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s). Documentation and archiving of the database may be implemented.

6.4 Quality Control

Quality controls employed will be outlined in the final observational study report.

6.5 Database Retention and Archiving of Study Documents

Location of database and supporting documentation will be outlined in the final observational study report.

6.6 Registration of Study on Public Website

This study will be registered on the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP).

7 ADVERSE EVENT REPORTING

This study does not meet the criteria for adverse event reporting.

8 GLOSSARY OF TERMS AND LIST OF ABBREVIATIONS

8.1 Glossary of Terms

Not applicable.

8.2 List of Abbreviations

Term	Definition
ATC	Anatomic Therapeutic Chemical
CHMP	Committee for Medicinal Products for Human Use
CIAP	International Classification of Primary Care
eGFR	Estimated Glomerular Filtration Rate
EU	Europe
GEP	Good Epidemiology Practices
ICD-9	International Classification of Diseases, 9th edition
ICD-10	International Classification of Diseases, 10th edition
ISPE	International Society for Pharmacoepidemiology
OAD	Oral Antidiabetic Drug
SAP	Statistical Analysis Plan
SGLT-2	Sodium glucose co-transporter
SmPC	Summary of Product Characteristics
T2DM	Type 2 Diabetes Mellitus

9 REFERENCES

1 Diabetes Atlas: http://www.idf.org/diabetesatlas

2 Wan J. et al, BMJ, 2013; 347:f5961

3 Denburg M.R., et al; 2011; 20(11): 1138-1149

APPENDIX 1 CSD DATA CAPTURE SYSTEM (PHYSICIAN SOFTWARE)

A Software package is installed in the physician's office for patient file follow-up and practice management.

The package includes the computer, software (see table below) Internet access, national health insurance institution transfer system, additional documentation (free drug prescription database). Also offered to participating physicians: training at no cost, priority in information system support (computer, Internet, software hotline), newsletter and a possible recruitment proposal in clinical trials or observational surveys.

Complete computerized clinical records are gathered in the Electronic Medical Records (EMR).

	Software for Physicians	CEGEDIM Software Company / Preferred Partners
BELGIUM	Heallth[one]	HDMP
FRANCE	Doc'Ware®	CEGEDIM Logiciels Médicaux
GERMANY	Software „Albis on Windows® aSoftware TurboMed® MCS-ISYNET ® DOCconcept®	CompuGROUP Holding MCS AG DOCexpert
ITALY	Millewin®	Millenium
UNITED KINGDOM	Vision®	In Practice Systems
SPAIN	OMI-AF®	Stacks

The data are uploaded by modem on a regular basis with a simple validation green light provided by the physician.

Once the MD has accepted to upload the data, they are transmitted to CSD every time the MD provides the agreed signal in Belgium, France, Italy, UK and Spain, whereas in Germany the transmission to CSD occurs twice a month.

The most recent data collected from CSD Databases can be viewed daily in Belgium, France, Italy, UK and Spain, and biweekly in Germany; providing a high level of quality of information.

Data collected are voluntarily recorded by the General Practitioner in his usual routine practice into the patient EMR. Since data are collected directly by the physicians and uploaded in an anonymized way, it is not possible to refer back to patients' files and perform any site quality control or any modification by asking more detailed information on any points that our client would like to know more about.

APPENDIX 2 KEY DATA METRICS OF THE GP'S LPD

		France	ITALY	UK	GERMANY	SPAIN	BELGIUM
GENERAL DATA							
Number of patients having visited a physician		100,0%	100,0%	100,0%	100,0%	100,00%	100,00%
Of which	Male	46,5%	46,7%	46,9%	45,9%	43,4%	46,4%
	Female	53,5%	53,3%	53,1%	54,1%	56,6%	53,6%
	Average age (years)	38,5 yo	50,2 yo	40,8 yo	49,2 yo	52,3 yo	42,8 yo
BIOMETRIC DATA							
Number of patients with a weight recorded (1 year history)		39,9%	15,1%	28,2%	15,3%	16,5%	14,4%
Number of patients with height recorded (1 year history)		20,0%	14,9%	13,9%	15,1%	12,9%	6,8%
Number of patients with BMI recorded (1 year history)		19,2%	14,8%	12,8%	14,9%	12,2%	6,5%
Number of patients with BMI recorded (3 years history + calc)		32,2%	30,7%	30,1%	17,4%	30,1%	12,0%
Number of patients with waist recorded (1 year history)		0,3%		0,2%			
Number of patients with Blood Pressure recorded (1 year history)		33,4%	26,9%	38,9%	14,6%	25,0%	35,3%
	average records by year by patient	*2,63*	*2,13*	*2,04*	*4,10*	*1,76*	*2,80*
BIOLOGICAL EXAMS							
Number of patients with a fasting blood glucose test		7,8%	39,4%	20,9%	32,7%	27,3%	35,1%
Number of patients with a Hba1c rate		2,7%	11,3%	8,9%	44,5%	6,3%	7,7%
Number of patients with Total cholesterol		7,3%	33,4%	23,7%	34,8%	0,0%	27,4%
Number of patients with LDL		6,9%	27,7%	15,7%	27,8%	14,6%	24,8%
Number of patients with HDL		6,8%	29,3%	20,5%	26,0%	19,5%	26,1%
Number of patients with Triglyceride		7,3%	32,0%	17,3%	30,5%	22,8%	26,4%
Number of patients with Transaminases SGOT (ASAT)		4,9%	26,5%	4,3%	22,4%	13,0%	29,0%
Number of patients with Transaminases SGPT (ALAT)		5,0%	29,8%	31,2%	30,1%	18,5%	31,1%
Number of patients with Gamma-GT		3,0%	17,7%	11,2%	38,4%	20,3%	28,7%

APPENDIX 3 CODE LISTS

	ICD9	ICD10
T2DM	250.x0	E10.x
Acute renal failure	584.x	N17.x
Severe chronic renal failure	585.x	N18.3, N18.4, N18.5
U failure	586.x	N19.
Dialysis	V45.11, V56.x	Y84.1
Renal transplantation	E878.0	Y83.0
Congestive heart failure	428.0	I50.0
Hypertension	401.x	I10.x, I11.x, I12.x, I13.x, I15.x

APPENDIX 4 MDRD GFR EQUATION

GFR = 175 x SerumCr-1.154 * age-0.203 * 1.212 (if patient is black) * 0.742 (if female)

APPENDIX 5 COCKROFT & GAULT FORMULA

Cl_{cr} = Estimated creatinine clearance (ml/min.)

Age = Age in years

Weight = Body weight (Kg)

[Cr] = Serum creatinine (μmol/L)

K = 1. $Cl_{Cr} = \frac{(140 - Age) \times Weight}{7.2 \times [Cr]} \times k$ 23 (men) or 1.04 (women)