Study Protocol

Pan European Multi-Database Bladder Cancer Risk Characterisation Study:

Additional causes of death analysis study

Study number: ER-9515

Version 1.0

Date: 21 April 2016

Sponsor: Takeda Development Centre Europe Ltd

TABLE OF CONTENTS

1 APPROVALS

I have carefully read this document and approve it with my signature.

Principal investigator EPID Research Oy

PASI KORHONEN
PhD, Adj. Prof of Biostatistics

Date 29.4.2016

Written by

Solomon Christopher, Statistician, EPID Research

Reviewed by

Fabian Hoti, Head of Statistics, EPID Research

2 LIST OF ABBREVIATIONS

ACE inhibitor	Angiotensin converting enzyme inhibitor
ARB	Angiotensin receptor blockers
ATC code	Anatomical therapeutic chemical classification system code
BPH	Benign prostatic hypertrophy
CHMP	Committee for Medicinal Products of Human Use
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CPE	Centre for Pharmacoepidemiology
CPRD	Clinical Practice Research Datalink Group
DDD	Defined daily dose
EMA	European medical agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
FIN	Finland
GP	General practitioner
HR	Hazard ratio
ICD-9; ICD-10	International classification of diseases, 9th revision; 10th revision
ICD-O-3	International classification of diseases for oncology, 3rd edition
ICPC	International classification of primary care
ISPE	International Society for Pharmacoepidemiology
KI	Karolinska Institute
NCSP	NOMESCO classification of surgical procedures code
NL	The Netherlands
NL GP	PHARMO General Practice Dataset
PASS	Post authorisation safety study
PHARMO	Pharmo Institute for Drug Outcomes Research
PUNLMP	Papillary urothelial neoplasia of low malignant potential
SWE	Sweden
UK	United Kingdom

3 RATIONALE

In order to obtain information on a potential association between pioglitazone use and bladder cancer, the Medicinal Products of Human Use Committee (CHMP) of the European Medicines Agency (EMA) requested Takeda Global Research & Development Centre Europe Ltd (Takeda) to conduct a Pan European post-authorization safety study (PASS) on pioglitazone use and occurrence of bladder cancer. The Pan-European bladder cancer risk characterization study was conducted pooling six datasets from four countries (Finland, Netherlands, Sweden and the United Kingdom) and the final study report was submitted to Global Regulatory Agencies in July 2015. As part of this report, the risk for total mortality and bladder cancer specific mortality were also presented in relation to pioglitazone use among type 2 diabetes patients. The finding from this study suggested a lower all-cause mortality rate on using pioglitazone. As a further step, it is planned to investigate and characterize the cause-specific mortality risk in relation to pioglitazone use among type 2 diabetes patients by using the same several datasets from the Pan-European bladder cancer risk characterisation study which was conducted pooling six datasets from four countries.

4 OBJECTIVES OF THE STUDY

The objective of this study is to assess the association between pioglitazone use and the cause-specific mortality risk among patients with type 2 diabetes mellitus through pooled patient level analysis using data from Finland (FIN), Sweden (SWE) and the United Kingdom (UK) databases.

The study will have the following main objectives:

- To estimate and compare the absolute and relative risk of cause-specific mortality in patients with type 2 diabetes who are ever exposed to pioglitazone vs. never exposed to pioglitazone.
- To estimate and compare the absolute and relative risk of cause-specific mortality in patients with type 2 diabetes with increasing duration of pioglitazone treatment.
- To estimate and compare the absolute and relative risk of cause-specific mortality in patients with type 2 diabetes with increasing cumulative dose of pioglitazone treatment.

5 PARTICIPATING CENTRES AND DATA SOURCES

5.1 Participating centres

Principal investigator Pasi Korhonen, PhD; Adj. Professor Research Director EPID Research Oy Metsänneidonkuja 12 FI-02130 Espoo Finland	Co-Investigator Edith M. Heintjes, PhD Scientific Research Manager Pharmo Institute for Drug Outcomes Research (PHARMO) Van Deventerlaan 30-40 3528 AE Utrecht The Netherlands (PHARMO will not provide data for cause-specific mortality analysis)
Co-Investigator Helle Kieler, MD, PhD Associate Professor Head, Centre for Pharmacoepidemiology (CPE) Karolinska Institutet (KI) T2, Karolinska University Hospital SE 171 76 Stockholm Sweden	Co-Investigator Rachael Williams, MSc Research Programme Manager The Clinical Practice Research Datalink Group (CPRD) The Medicines and Healthcare products Regulatory Agency 151 Buckingham Palace Road London SW1W 9SZ England

Other co-investigators who have contributed to the design of the bladder cancer study protocol:

Shahram Bahmanyar, Associate Professor, CPE; Paul Dolin, Director Epidemiology, Takeda; Susan Eaton, MSPH, CPRD; Fabian Hoti, PhD, EPID Research; Leanne Houweling, MSc, PHARMO; Marie Linder, PhD, CPE; Timothy Williams, Dr, CPRD; Helen Strongman, MSc, CPRD; Solomon Christopher MSc, EPID Research; Maila Majak, MSc, EPID Research

5.2 Data sources/pooled analysis datasets

This new cause specific mortality study will utilize the same study cohorts constructed in the original Pan European Multi-Database Bladder Cancer Risk Characterisation Study for estimating the risk of bladder cancer in association with pioglitazone use (original study protocol attached as appendix).

The new study will use linkage of drug prescribing/dispensing databases to country-specific selected databases including (i) cancer registries, (ii) general practitioner (GP) records and/or hospital discharge records, (iii) death records, (iv) reimbursement decisions, and (v) immigration and emigration records in FIN, SWE and UK.

The study will contain two pooled datasets, a primary and a secondary pooled dataset. In the *primary* pooled dataset all country-specific datasets are combined into a single dataset containing common variables from all datasets. In the *secondary* dataset only SWE and UK datasets with information on smoking, BMI, and HbA1C will be used.

6 STUDY POPULATION

The target population consists of patients with type 2 DM whose antidiabetic treatment at cohort entry is modified to include pioglitazone or another antidiabetic medication. In each dataset these patients are identified with the following inclusion / exclusion criteria.

Inclusion criteria:

- Treatment with any oral antidiabetic drugs at any time in the available medication records.
- Baseline is modified (cohort entry point) to include pioglitazone (exposure group) or another antidiabetic medication (reference group).
- Age ≥ 40 years at cohort entry.
- At least 12 months of medication database membership during baseline period prior to cohort entry.

Exclusion criteria:

- Diagnosis of type 1 diabetes, gestational diabetes, or secondary and other types of diabetes mellitus prior to cohort entry using available data sources,
- Patients who are entitled to special reimbursement for type 1 diabetes, diabetes with malnutrition, other diabetes or gestational diabetes (specific to FIN dataset only), and
- Diagnosis or history of bladder cancer, in-situ bladder cancer or benign neoplasm of bladder prior to cohort entry.
- Diagnosis of secondary malignant neoplasm of bladder prior to cohort entry
- Diagnosis of neoplasms of uncertain or unknown behaviour of bladder prior to cohort entry
- History of cystectomy at cohort entry
- History of resection or removal of bladder tumour (benign or malignant) at cohort entry
- History of biopsy of bladder tumour or lesion a cohort entry

6.1 Study cohorts

In the original study, to minimize channelling bias, each pioglitazone-exposed patient (exposure group) was matched (Li et al 2012; Spreeuwenberg 2010) with up to 10 pioglitazone-unexposed diabetic patients (reference group). Each country performed the matching of country specific datasets according to the main study protocol, version 2.0, dated 20th June 2013.

Matching was performed in two stages (i) exact matching based on groups defined by prior use of other TZD, antidiabetic treatment prior to matching and type of

treatment modification (ii) matching based on propensity score (probability to being prescribed pioglitazone) calipers using a list of variables associated with choice of treatment and bladder cancer risk.

It should be noted that the exact matching variables are related to treatment only (and not bladder cancer outcome). The variables used in the propensity score were indicators of advanced diabetes and cardiovascular risks that are associated with mortality, in general, as with bladder cancer outcome itself in addition to the treatment choice. Therefore, the matched study cohorts can be used to study the risk of mortality associated with use of pioglitazone.

6.2 Cohort entry date and start of follow-up

Cohort entry date (CED) and start of follow-up are defined as the date when the baseline therapy (including no pharmacotherapy) was modified to include pioglitazone (exposure group) or another antidiabetic medicine (reference group).

6.3 Baseline observation period

Each patient must have at least one year of membership in the relevant study database prior to cohort entry. This period is used to characterise baseline therapy, baseline medical history, and baseline covariates. This baseline observation period will start from January 1, 1998 onwards in Finland, from July 1, 2005 onwards in Sweden, and from start of database membership for the UK datasets. In the Finnish and Swedish datasets the coverage for each patient is assured by verifying the place of domicile not being abroad 12 months prior to the index date (FIN, SWE).

6.4 End of follow-up

In the mortality analyses, the follow-up time for each patient will start at cohort entry and will end at death, start of thiazolidinediones (other than pioglitazone) at or after cohort entry, end of membership in database, end of database coverage, or end of study period, whichever occurs first. The end of study period is 30 June, 2011 for Sweden, and 31 December, 2010 for UK and Finland.

6.5 Censoring of follow-up

The UK CPRD GOLD-HES datasets have definable entry and exit dates for each person. Exit date will be used as a follow-up censoring date. The FIN Population Information System register and the SWE Total Population Register (maintained by Statistics Sweden) contain immigration and emigration date information and will be used if available to censor follow-up. In addition date of death will be used as an endpoint for the mortality analyses.

In the analysis of a particular cause of death in cause-specific mortality analyses, death from any other cause is treated as censored observation.

7 STUDY VARIABLES

7.1 Outcome variables

The following cardiovascular disease (CV) related causes of death are of particular interest and will be analyzed separately. In addition CV and non-CV causes of death will be analysed as two separate groups.

1. Cardiovascular disease CV
    1. Myocardial infarction (MI)
    2. Stroke
    3. Sudden death
    4. Heart failure or cardiogenic shock
    5. Renovascular disease
    6. Peripheral vascular disease
    7. Arrhythmia
    8. Other CV (Aortic aneurysm, other)
2. Non-CV

For completeness the composite of the non-CV causes of deaths will be summarized on subgroup level based on the first 3 characters of the ICD-10 codes.

Cause-specific mortality

Information on date and causes of death of all deaths among cohort members will be obtained from the national death registries for FIN, SWE, and UK CPRD GOLD-HES datasets.

7.2 Exposure variables

Continuous drug use periods are constructed for the following drug groups.

Table 1: Diabetes drugs

Group number	Group name*
1	Pioglitazone
2	Other thiazolidinediones (including rosiglitazone)
3	Biguanides/Metformin
4	Sulphonylureas
5	DDP-4 inhibitors
6	Alpha glucosidase inhibitors
7	GLP-1 receptor agonists
8	Meglitinides
9	Amylin analogues
10	Other oral diabetic medications
11	Insulin
** Combination products included into multiple groups.*	

The following time dependent variables will be calculated for each of the diabetes drug groups of interest at any given time during the follow-up:

- Current exposure indicating whether the patient is in the treatment group at the current time point (needed for all drug groups to identify changes in treatment)
- Ever vs. never exposure taking the value 1 (ever) as soon as one prescription of the treatment group has been purchased / prescribed, and 0 (never) otherwise.
- Duration of exposure indicating how long the patient has been exposed to a treatment group at the current time point (only for pioglitazone and insulin group).
- Cumulative dose indicating the cumulated dose the patient has been exposed to at the current time point (only for pioglitazone group).
- Time since last exposed to treatment group (only for pioglitazone group).

For other drugs of interest only the date of the first purchase / prescription will be used to define the time dependent ever vs. never exposure variable. For pioglitazone the exposure variables listed in Table 2 are constructed.

Table 2: Pioglitazone exposure variables

Pioglitazone exposure variable (T = time dependent)	**Description**
Current exposure (T)	Indicator of current pioglitazone use either used alone or as a combination. This variable is not in the statistical analysis but only used in the calculation of duration of exposure and time since last dose.
Ever vs. never use (T)	Taking the value 1 (ever) as soon as one prescription with pioglitazone has been purchased, and 0 (never) otherwise.
Duration of exposure (T)	Time-dependent cumulative sum of durations of previous pioglitazone exposure periods. The duration of exposure categories are: Never, <12 months, 12-24 months, 24-48 months and >48 months.
Cumulative dose (T)	Time-dependent cumulative sum of drug consumption based on the daily dosage of pioglitazone containing prescriptions or dispensings since entry into the cohort. The cumulative dose categories are: Never, 1-10,500 mg, 10,501-28,000 mg, 28,000-40,000 mg and >40,000 mg.
Time since last dose (T)	Current time minus the time of the end of last current exposure to pioglitazone containing prescriptions since entry into the study cohort. Categorised into Current, <1 year, 1-2 years, 2-4 years, 4+ years, Never

7.3 Cohort matching variables

The variables used in constructing the country specific matched populations are described in Table 3.

Table 3: Matching variables

Variable (F= Fixed at cohort entry, T = time dependent)	**Description**
Propensity score variables	
Duration of treated diabetes mellitus at cohort entry (F)	Duration (years) of treated diabetes mellitus is approximated as the interval between the first diabetes therapy in the prescription records, and date of cohort entry. The categories used are <1 year, 1 to <2 years, 2 to < 4 years, 4 to <6 years and ≥ 6 years
History of diabetic complications at cohort entry (F)	History of diabetic complications at CED defined by the following 5 separate (No/Yes) variables: • Diabetic retinopathy or maculopathy • Ketoacidosis • Diabetic coma • Diabetic lower limb severe complications • Diabetic renal complications
History of myocardial infarction or stroke at cohort entry (F)	Classified as (Yes/No).
History of congestive heart failure at cohort entry (F)	Classified as (Yes/No).
Year of cohort entry (F)	Year of CED with categories 2000-2003, 2004-2007 and 2008-2011
Duration of database membership before cohort entry (F)	Duration (years) of membership in medication database prior to cohort entry with categories 1-2, 3-4, 5-6, 7+
Number of different antidiabetic drug classes prior to cohort entry (F)	Score from 0 to 10 with one point from each of the following classes used prior to cohort entry: metformin, sulphonylureas, other TZDs, alphaglucosidase inhibitors, DPP-4 inhibitors, GLP-1R agonists, meglitinides, amylin analogues, insulin, other. Combination products contribute separately to each drug class based on the active substances included in the product. The categories used are 0, 1, 2, 3 and >3.
Exact matching variables	
Use of other TDZs prior to cohort entry (F)	Use of other thiazolidinediones (other than pioglitazone) prior to cohort entry (N/Y)
Type of antidiabetic treatment prior to cohort entry (F)	Type of antidiabetic medication immediately prior to cohort entry classified as: • No treatment

	• Metformin only • Sulphonylurea only • Metformin + sulphonylureas only • Insulin with or without other antidiabetic medications • Any other antidiabetic medications or combinations
Type of modification in baseline therapy (F)	Type of modification in baseline antidiabetic therapy at CED classified as treatment switch or addition of new treatment to all prior treatment (if any existed).

7.4 Other relevant covariates

The relevant covariates to be adjusted for are given in Table 4.

Table 4: Other relevant covariates

Covariates (F = fixed at cohort entry, T = time dependent)	**Description**
Dataset identifier (F)	**Dataset membership categorized as: FIN, SWE, CPRD GOLD HES, NL GP, NL Pharmacy-Hospital**
Gender (F)	Male/Female
Age (F)	Age at cohort entry with categories 40-49, 50-59, 60-69 and ≥70
Cigarette smoking (F)	Cigarette smoking status at cohort entry. If not available before cohort entry the first available record after cohort entry is adopted. If no cigarette smoking data available, coded as 'unknown'. Smoking data is not available in FIN dataset. Categorized into: Never, Ever (current or former), Unknown
Body Mass Index at cohort entry (F)	Classified as missing, <30, 30-34.9 and ≥ 35. Used in those datasets where information is available. If not available at baseline, the first record within 12 months of cohort entry is adopted. If no BMI data is available, coded as 'missing'. BMI data is not available in FIN dataset.
HbA1C (F, T)	Classified as missing, <7.5%, 7.5-8.9%, ≥9.0%. Baseline HbA1C measurement will be the most recent record within 6 months prior to cohort entry. Persons with no baseline HbA1C will be coded as 'missing'. HbA1C will also be measured as a time varying covariate, using the most recent HbA1C record at a point in time.

	HbA1C data is not available in FIN dataset.
History of diabetic complications	
Diabetic retinopathy or maculopathy (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Ketoacidosis (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Diabetic coma (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Diabetic lower limb complications (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Diabetic Renal complications (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
History of relevant comorbidities	
Other urinary tract cancers (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Other cancers, excluding urinary tract (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Myocardial infarction or stroke (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Peripheral vascular disease (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Congestive heart failure (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
Chronic pulmonary obstructive disease (T)	Classified as never (0) or ever (1) with the condition evaluated at any given time during the follow-up; Starting at 1 if condition exists at cohort entry, 0 otherwise.
History of relevant medications	

Statin use (T)	Prior use of statins or statin combinations Classified as never (0) or ever (1) use evaluated at any given time during the follow-up; Starting at 1 if prior use at cohort entry, 0 otherwise.
ARB use (T)	Prior use of angiotensin receptor blockers (ARB) Classified as never (0) or ever (1) use evaluated at any given time during the follow-up; Starting at 1 if prior use at cohort entry, 0 otherwise.
ACE use (T)	Prior use of angiotensin converting enzyme (ACE) inhibitors Classified as never (0) or ever (1) use evaluated at any given time during the follow-up; Starting at 1 if prior use at cohort entry, 0 otherwise.
BPH drug use (T)	Prior use of drug for benign prostatic hypertrophy (BPH) Classified as never (0) or ever (1) use evaluated at any given time during the follow-up; Starting at 1 if prior use at cohort entry, 0 otherwise.
Use of other diabetic medications (T)	A separate variable for each of the following classes: metformin, sulphonylureas, other thiazolidinediones, alpha-glucosidase inhibitors, DPP-4 inhibitors, GLP-1 agonists, meglitinides, amylin analogues, insulins and other oral antidiabetic medications. Classified as never (0) or ever (1) use evaluated at any given time during the follow-up; Starting at 1 if prior use at cohort entry, 0 otherwise.

8 DATA DEIDENTIFICATION

Prior to transferring the data to EPID Research, the following precautionary measures are taken by each country to ensure non-identifiability of study individuals in the pooled analysis.

- Patients are anonymised and assigned a dummy study identification number (SID)
- Dates of birth are coded at year level only
- For each individual the cohort entry time is taken as time zero
- All other event times are defined relatively to cohort entry time
- The actual cohort entry time is coded at year level only
- Country specific dataset will not include within country location information (e.g. address, postal code, municipality)
- Reporting of the results is done on aggregated level only, according to the study protocol

9 DATA STORAGE

The datasets will only be used for the analysis outlined in the protocol. The datasets will be retained at EPID Research until the study report is finalized and the main results have been published. This is to ensure comments raised by the referees in the submission phase can be addressed in a timely fashion.

The country specific components of the datasets will be preserved by each country and in case new analyses of the pooled datasets is required a new request for the use of the data will be made according to the data source specific requirements.

10 DATA ANALYSIS

The pooled datasets will be analysed using the methods and approaches described here. The analyses are similar to those used in the analyses of the individual data sets. The Cox proportional hazards models will include a dataset identifier as a categorical covariate.

10.1 Pooled datasets

The following two pooled datasets will be constructed (see Table 5 for an overview)

Primary pooled dataset

The primary pooled analysis will include the largest number of patients as possible with common variables from all datasets. This analysis will not include information on smoking, BMI, and HbA1Cs.

For FIN and SWE -the hospital-based morbidity data sets will be utilized.

For UK -the CPRD GOLD-HES data will be used.

Secondary pooled dataset (sensitivity analysis)

The secondary pooled analysis will utilize datasets containing information on smoking, BMI, and HbA1C.

For FIN -no dataset will be included.

For SWE -the hospital-based morbidity dataset is complimented with data from the National Diabetes Register.

For UK -the CPRD GOLD-HES will be used.

Table 5: Pooled datasets

Country	Primary pooled analysis	Secondary pooled analysis
	Including data from as many patients, and the greatest number of sources for these patients. The limitation is that we will not include variables that are missing from some data sets (e.g. BMI)	Restricted to patients with the extra variables (e.g. BMI) available, but using the maximum number of sources for these patients
UK	CPRD-GOLD-HES (linkage dataset)	CPRD-GOLD-HES (linkage dataset)
Sweden	Hospital based morbidity dataset	Hospital based morbidity complimented with information from the National Diabetes Register
Finland	Hospital based morbidity dataset	Not available

10.2 Statistical hypothesis

With regard to cause-specific mortality the formal statistical hypothesis between any two treatment groups is as follows:

$$H_0 : HR = \text{Mortality rate}_{PioGroupX} / \text{Mortality rate}_{REF} = 1$$

against the alternative

$$H_1 : HR = \text{Mortality rate}_{PioGroupX} / \text{Mortality rate}_{REF} \neq 1$$

where

HR denotes the hazard ratio,

$\text{Mortality rate}_{PioGroupX}$ denotes the cause-specific mortality rate in a specific group of pioglitazone exposure for example in the group of ever exposed to pioglitazone or in the group of <18 months of exposure to pioglitazone, and

$\text{Mortality rate}_{REF}$ denotes the cause-specific mortality rate in the reference group of patients never exposed to pioglitazone.

10.3 Sample size

The numbers of patients who were ever exposed and who were never exposed to pioglitazone in the PS-matched cohort and estimated numbers of deaths due to causes of interest as expected during the follow-up are presented in Table 6 for each dataset.

Table 6: Presents the number exposed, follow-up time and mortality in each dataset

Datasets	Ever exposed to pioglitazone			Never exposed to pioglitazone		
	Exposed[1]	Follow-up[2]	Mortality	Unexposed[1]	Follow-up[2]	Mortality[3]
FIN	18,794			157,263		
SWE	3,712			31,408		
CPRD GOLD-HES	12,109			35,184		
Pooled-1	34,615	2.90		223,855	2.83	448.30
Pooled-2	15,821			66,592		

[1] number exposed / unexposed to pioglitazone in the PS-matched cohort with 1:10 ratio

[2] mean follow-up time (in years) up to death

[3] Mortality rate per 10,000 person-years

The proportions of the cause-specific mortality outcomes are expected to range from 5% -25%. Based on this and the observed all-cause mortality rate (Table 6) the power for the study is computed and presented in Table 7. The power calculations for the cohort study (1:1 and 1:10 matching) are presented. For the cohort study the power is calculated for the effect sizes 1.1 – 1.4 on the relative risk scale. A two-sided type 1 error rate of 5% is used.

Table 7: Power of the study for 10% and 20% of all-cause mortality with 1:1 and 1:10 matching ratios for Pooled-1 dataset

HR	Cause-specific mortality as % of total mortality and Matching ratio			
	20% ; 1:1	10% ; 1:1	20% ; 1:10	10% ; 1:10
1.1	54.8%	31.1%	79.6%	50.1%
1.2	98.2%	81.6%	99.9%	97.3%
1.4	> 99.9%	99.9%	> 99.9%	> 99.9%

The power calculations were carried out using the cpower function of the Hmisc package in the R-program (R Development Core Team 2008; Peterson et al. 1993; Lachin and Foulkes 1986; Schoenfeld 1983).

10.4 Missing data

If a variable (for example smoking and BMI) is not available in an individual data set, which is included in the pooled data set, the missing variable is excluded from the analysis of that pooled database. If a variable is missing only for some patients in an individual data set, a missing data category is added and used in the analysis of the pooled data set.

10.5 Population summary

The characteristics of the post-matching pooled study population will be described with descriptive statistics at the entry into the study cohort. The following baseline variables will be included in the population summary: country, age, gender, year of entry, duration of treated diabetes mellitus, history of relevant diabetic medications, history of other relevant medications, history of diabetic complications and history of other relevant comorbidities. The summary of time to death for the pre-specified causes of interest (see 7.1 Outcome variables) will also be presented.

The population summary will be stratified according to pioglitazone exposure using ever vs. never exposure variable.

In addition, the non-CV causes of death will be summarised in subgroups based on the first 3 characters of the ICD-10 code. The summaries will include counts and percentages and will not be stratified by pioglitazone exposure.

10.6 Descriptive evaluation

Comparisons for the various pioglitazone exposure definitions will be performed. Crude cause-specific mortality rates with 95% confidence intervals (CI) will be estimated for each pioglitazone exposure definition separately within the strata of gender, age, year of entry, duration of treated diabetes, history of relevant diabetic medications, history of relevant concomitant medications, history of diabetic complications and history of other relevant comorbidities.

10.7 Formal evaluation

The hazard ratio (HR) estimates with 95% CIs for each pioglitazone exposure definition (ever exposed to pioglitazone vs never exposed to pioglitazone; increasing duration of pioglitazone; and cumulative dose of pioglitazone) will be estimated using the conventional Cox's proportional hazards model with a counting process approach which enables the follow-up time of each patient to be split into several periods and thus allows adjustments for relevant baseline and time-dependent covariates in the model specification

In the matched cohort analysis three models will be used: i) the crude model with pioglitazone exposure as ever vs. never exposed included only, ii) the base model with age at cohort entry, sex and use of metformin, sulphonylureas, insulins or other antidiabetic drugs each classified as never vs. ever exposed added to the crude model, and iii) the adjusted model created by using the following forward and backward selection procedure:

1. Identify candidate covariates as those covariates with at least 5% prevalence in pioglitazone or reference group and with the p-value of the univariate association between the covariate and bladder cancer incidence <0.1.

2. Start with the base model
3. For each candidate covariate produce a new adjusted model with the candidate covariate one at a time added in the base model.
4. Consider each candidate covariate as a potential confounder if the relative change in the HR of pioglitazone exposure is at least a 10% relative when comparing to the base model. This process will generate a set of potential confounders.
5. Add all potential confounders simultaneously in the base model
6. Remove each potential confounder one at a time from the model and see if this results in a 10% relative change in the HR of pioglitazone exposure. If any potential confounder does not fulfil the 10% threshold, drop the one with the smallest relative change in HR. Repeat the process until no further changes are needed. This will be the adjusted model.

The covariates included in adjusted model will be used in the analysis of the effect of different pioglitazone exposure definitions (never/ever exposed; duration of exposure; cumulative dose; time since last dose) on cause specific-mortality.

The crude, base and adjusted models will include as covariates the study cohort dataset identifier and all matching variables (Bland 1994, Stuart 2010). The propensity score (at start of follow-up) will be included as a categorical variable based on quintiles (5-quantiles). Especially, the Cox's proportional hazards models will not be stratified using the matching sets (Austin 2013).

10.8 Stratified analysis

The risk models will be presented stratified, if there are sufficient events, by the following baseline variables:

- Duration of treated diabetes
- Use of other TZDs (other than pioglitazone) prior to cohort entry
- History of Chronic Kidney disease or renal impairment
- History of diabetic complications
- History of macrovascular disease

10.9 Sensitivity analysis

Analysis to assess the impact of adjusting/not adjusting for smoking status, BMI and HbA1C information in SWE and CPRD GOLD-HES datasets

Analysis comparing risk estimates from incident T2DM sub-cohort vs. prevalent-only T2DM sub-cohort will be performed. The incident sub-cohort will include only patients with at least 12 months of database membership before first diabetes treatment. The prevalent-only sub-cohort will include only patients with less than 12 months of prescription database membership before first recorded diabetic medication. This analysis will be performed using the first pooled dataset.

Analysis in which the pioglitazone exposure definition will be changed from at least one prescription to at least 2 prescriptions within the first 6-month period will be performed. To ensure immortal time bias is not introduced, the CED will be moved to

be the original CED plus 6 months. This analysis will be performed using the first pooled dataset.

11 LIMITATIONS

A possible limitation to the design of study in terms of estimating mortality risk is that some variables that are important to mortality are not considered in the matching procedure. This is because the study was primarily designed to estimate bladder cancer risk and the variables were selected for that outcome. However, additional variables that influence death due to other causes can be adjusted for at the analysis stage.

12 COMMUNICATION OF STUDY RESULTS

The principal and co-investigators will write the study report. The report is delivered to the Sponsor. Based on these results the principal and co-investigators will co-author scientific manuscript(s) of the results to be published. The publication strategy has been defined in the research agreement between the principal investigators and the Sponsor. The publications will include a large group of co-authors and each country can decide which contributors need to be included.
The study Sponsor is entitled to view the final results and interpretations thereof prior to submission for publication and to comment in advance of submission as agreed in the research contract and without unjustifiably delaying the publication.

13 APPENDICES

Protocol version 2.0, dated 20 Jun 2013, titled “Pan European Multi-Database Bladder Cancer Risk Characterisation Study”

14 REFERENCES:

Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med* 2013 Jul 20;32(16):2837-49

Bland JM, Altman DG. Statistics notes: Matching. *BMJ* 1994;309:1128

ENCePP. The ENCePP Code of Conduct for Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies. 21 November 2011, EMA/929209/2011.
www.encepp.eu/code_of_conduct/documents/CodeofConduct_Rev2.pdf

ISPE, International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiological Practices. Revision 2: April 2007.
www.pharmacoepi.org/resources/guidelines_08027.cfm

Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowances for non-uniform patient entry and losses to follow-up, non-compliance and stratification. *Biometrics.* 1986, 42: 507–519.

Li W, MacDonald TM, Mackenzie IS. Pioglitazone and bladder cancer: a propensity score matched cohort study. Br J Clin Pharmacol 2012.
http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2125.2012.04325.x/pdf

Peterson B, George SL. *Controlled Clinical Trials.* 1993, 14: 511–522.

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, 2008. http://www.R-project.org.

Schoenfeld D. Sample-Size Formula for the Proportional-Hazards Regression Model. *Biometrics.* 1983, 39: 499–503.

Spreeuwenberg MA et al. The multiple propensity score as control for bias in the comparison of more than two treatment arms: An introduction from a case study in mental health. *Medical Care.* 2010, 48: 166-174.

Stuart EA, Matching methods for causal inference: A review and a look forward. Stat. Sci. 2010 Feb 1; 25(1):1_21