1. TITLE

Pan European Multi-Database Bladder Cancer Risk Characterisation Study

2. MARKETING AUTHORIZATION HOLDER AND SPONSOR

Takeda Global Research & Development Centre (Europe) Ltd

3. RESPONSIBLE PARTIES

Principal Investigator:

Pasi Korhonen, PhD
Adjunct Professor of Biostatistics
EPID Research Oy
Tekniikantie 12
FI-02150 Espoo
Finland

Co-Investigators:

Edith M. Heintjes, PhD Scientific Research Manager Pharmo Institute for Drug Outcomes Research (PHARMO) Van Deventerlaan 30-40 3528 AE Utrecht The Netherlands	**Rachael Boggon, MSc** Research Statistician The Clinical Practice Research Datalink Group (CPRD) The Medicines and Healthcare products Regulatory Agency 151 Buckingham Palace Road London SW1W 9SZ England
Helle Kieler, MD, PhD Associate Professor Head, Centre for Pharmacoepidemiology Karolinska Institutet (KI) T2, Karolinska University Hospital SE 171 76 Stockholm Sweden	

Other co-investigators who have contributed to the design of the study protocol:
Shahram Bahmanyar, Associate Professor, KI; Paul Dolin, Director Epidemiology, Takeda; Susan Eaton, MSPH, CPRD; Fabian Hoti, PhD, EPID Research; Leanne Houweling, MSc, PHARMO; Marie Linder, PhD, KI; Timothy Williams, Dr, CPRD

4. ABSTRACT

Title of study: Pan European Multi Database Bladder Cancer Risk Characterisation Study

Rationale: The Committee for Medicinal Products of Human Use (CHMP) of the European Medicines Agency recently finalized its review on pioglitazone use and occurrence of bladder cancer. Evidence from several recent epidemiological studies suggests a small increased risk of bladder cancer with pioglitazone use in patients with type 2 diabetes [Lewis et al. 2011, CNAMTS 2011]. In the 4th interim analysis (8-years follow-up) of the Kaiser Permanente Northern California Study cohort, the hazard ratio (HR) of ever vs. never use of pioglitazone was 1.07 (95% CI 0.87-1.30) and a non-significant elevated risk among patients with longest >4 years duration of therapy (HR=1.30, 95% CI 0.91-1.86) [Lewis et al. 2012]. In the French CNAMTS study a similar risk for pioglitazone use was observed (HR=1.22, 95% CI 1.05-1.43), and a significant risk with cumulative doses over 28 000 mg (HR=1.75, 95% CI 1.22-2.50), and with duration of therapy between 12 to 23 months (HR=1.34, 95% CI 1.02-1.75) or over 24 months (HR=1.36, 95% CI 1.04-1.79) [CNAMTS 2011]. The CHMP agreed that further evidence is still needed. Most previous studies were limited in their control of channelling bias, detection bias or control for confounding by important bladder cancer risk factors such as BMI and smoking.

Limited information is currently available on whether the potential increase in risk is associated with a particular type or stage of bladder cancer. Data from the Kaiser Permanente Northern California Study cohort study showed any increased risk was largely limited to early stage tumours [Lewis et al. 2012].

This observational study is being undertaken to further assess the association between pioglitazone use and bladder cancer risk among patients with type 2 diabetes mellitus in four European countries: Finland, Netherlands, Sweden, and United Kingdom.

This study is being undertaken at the request of the European Medicines Agency. The study protocol has been reviewed by the European Medicines Agency.

Research Questions:

Primary objective:

1) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes who are ever exposed to pioglitazone vs. never exposed to pioglitazone.

Secondary objectives:

2) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with increasing duration of pioglitazone treatment,
3) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with increasing cumulative dose of pioglitazone treatment,
4) To examine the association between pioglitazone exposure and risk of bladder cancer with respect to possible channelling bias, detection bias and other sources of confounding.
5) To characterize the stage and grade of the bladder cancer cases at the time of diagnosis in patients with type 2 diabetes who are ever exposed to pioglitazone vs. who are never exposed to pioglitazone, and
6) To characterize all-cause mortality and bladder cancer-specific mortality pattern amongst patients ever exposed to pioglitazone vs. never exposed to pioglitazone.

Explorative objective:

To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with time since the last dose of pioglitazone use.

Study design:

- Retrospective Cohort Studies in 4 countries
- Individual patient level data meta analysis of the retrospective cohort study datasets

Matching & Channelling bias

The observed association between long term pioglitazone use and risk of bladder cancer could in part be due to channelling bias because pioglitazone is likely to be given to patients with more severe diabetes as compared to other oral diabetic medications, and patients with more advanced diabetes are likely to be at higher risk of bladder cancer.

To minimize channelling bias, each pioglitazone-exposed patient (exposure group) will be matched with up to 10 pioglitazone-unexposed diabetic patients (reference group) based on:

- Propensity score derived from the following variables:
  - Duration (years) of treated diabetes mellitus at cohort entry
  - History of diabetic complications at cohort entry
  - History of myocardial infarction or stroke at cohort entry
  - History of congestive heart failure at cohort entry
  - Year of cohort entry
  - Duration (years) of membership in medication database prior to cohort entry
  - Number of different antidiabetic drug classes prior to cohort entry
- Use of other thiazolidinediones (other than pioglitazone) prior to cohort entry
- Type of antidiabetic treatment immediately prior to cohort entry
- Type of modification in baseline antidiabetic therapy at cohort entry classified as treatment switch or addition of new treatment to all prior treatment (if any existed)
- Living in geographical area for which cancer data are available (yes/no, used in the Netherlands only)

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. 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
- Patients who are entitled to special reimbursement for diabetes with ICD-10 diagnosis starting with 'E10', 'E12', 'E13', 'O24.4' indicative of type 1 diabetes, diabetes with malnutrition, other diabetes or gestational

diabetes, respectively (specific to FIN dataset only), and

- Diagnosis or history of bladder cancer or in-situ bladder cancer prior to cohort entry using available data sources

Data Sources:

The cohort studies will be undertaken in the 4 European countries (FIN, NL, SWE, UK) using linkage of drug prescribing/dispensing databases to country-specific selected databases including (i) cancer registries, (ii) GP records and/or hospital discharge records, (iii) death records, (iv) reimbursement decisions, and (v) immigration and emigration records.

As some countries only have hospital-discharge morbidity databases while other have GP-based records, the datasets are categorized according to source of morbidity covariate data:

- 4 hospital-based morbidity datasets (FIN dataset, SWE dataset, NL Pharmacy-Hospital dataset, UK CPRD GOLD-HES dataset)
- 2 GP-based morbidity datasets (NL GP dataset, UK CPRD GOLD dataset)

Throughout the protocol, definitions of outcome and other explanatory variables are harmonized across the datasets within each category.

Definition of follow-up time and outcome

Period covered by the datasets

Dataset	Start of coverage	End of Follow-up
FIN	01 January 2000	30 June 2011
NL Pharmacy-Hospital	01 January 2000	30 June 2011
NL GP	01 January 2000	30 June 2011
SWE	01 July 2005	30 June 2011
UK CPRD GOLD	01 January 2000	30 June 2011
UK CPRD GOLD-HES	01 January 2000	31 December 2010

Cohort entry / Start of Follow-up

Cohort entry and start of follow-up is defined as the date when the baseline therapy was modified to include pioglitazone or another diabetic medicine.

Baseline observation period before cohort entry

Each patient must have at least 1 year of membership in the relevant medication database prior to cohort entry. This period is used to characterise baseline therapy, baseline medical history, and baseline covariates, and minimise the risk of left-censoring of exposure information.

End of Follow-up

In the bladder cancer analyses, each patient will be followed-up from cohort entry until the date of diagnosis of the first incident bladder cancer or censored at end of membership of the database, death, start of other thiazolidinediones (other than pioglitazone) at or after cohort entry, end of database coverage or 30 June 2011 whichever occurs first.

In the mortality analyses, each patient will be followed-up from cohort entry until date of death, start of other thiazolidinediones (other than pioglitazone) at or after cohort entry, end of membership of the database, end of database coverage or 30 June 2011, whichever occurs first.

Diagnosis of Bladder Cancer

In the FIN and SWE and CPRD GOLD-HES datasets, cancer registry-based information will be used for all patients.

The NL Pharmacy-Hospital and NL GP datasets only have partial linkage to the cancer registry data (20-30% in NL datasets). It is not possible to use cancer registry-based diagnosis for all subjects. Diagnosis of bladder cancer will be based on cancer registry data where available. For practices with no linkage, GP records data (NL GP dataset) and Hospital discharge data (NL Pharmacy-Hospital dataset) will be used to identify bladder cancers.

In the NL Pharmacy-Hospital, NL GP, UK CPRD GOLD and UK CPRD GOLD-HES datasets, the occurrence of bladder cancer in cancer registry data and GP/hospital records data will be compared to validate using GP or hospital discharge data for bladder cancer diagnosis.

The CPRD GOLD dataset is not linked to the cancer registry data. Diagnosis of bladder cancer will be based on GP records data only.

All cause mortality and bladder cancer mortality

In the FIN, SWE and UK CPRD GOLD-HES datasets, there is full linkage to death registries. Bladder cancer mortality will only include deaths where bladder cancer was recorded as the underlying cause of death. Date of death but not causes of death are available for the two NL datasets and UK CPRD GOLD dataset.

Definition of exposure:

Pioglitazone exposure variable (T = time dependent)	**Description**
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 the previous pioglitazone exposure periods. The duration of exposure categories will be taken identical to those used in the KPNC final (10-year follow-up) analysis.
Cumulative dose (T)	Time-dependent cumulative sum of drug consumption based on daily dosage of pioglitazone containing dispensings or prescriptions since entry into the cohort. The cumulative dose categories will be identical to those used in the KPNC final (10-year follow-up) analysis.
Time since last dose (T)	Current time minus the time of the end of last current exposure to pioglitazone containing dispensing or prescription since entry into the study cohort. Categorised into Current, <1 year, 1-2 years, 2-3 years, 3-4 years, 4+ years, Never

Other explanatory variables

Covariate to be assessed (where available in a dataset) in the cohort studies include:

- Gender
- BMI
- Smoking
- HbA1C
- History and concurrent relevant comorbidities, including cardiovascular disease, CHF, other cancers, chronic renal disease
- History and concurrent renal co-morbidity, including renal dialysis, transplant
- History and concurrent bladder comorbidities, including incontinence, UTI, pyelonephritis, urolithiasis, haematuria, retention, catheterization
- Concurrent diabetic complications, including retinopathy, peripheral neuropathy, nephropathy, proteinuria, microalbuminuria
- History and concurrent use of other medications, including statins, ARBs, ACEs, BPH drugs

In the 4 Hospital-based morbidity datasets some covariates will not be assessable, while others may be based on hospitalizations (i.e., more severe cases) or on use of specific medications.

In the 2 GP-based morbidity datasets covariates will be based on GP records or use of specific medications.

Where possible identical definitions are used across all six datasets. Where this is not possible, a single definition is used within the hospital-based morbidity data sets, and a separate single definition is used in the GP-based morbidity datasets.

Statistical methods:

Analysis of individual cohort datasets Identical statistical methods, analyses and data categorizations will be used in each cohort study. The hazard ratio (HR) estimates with 95% CIs for each pioglitazone exposure definition will be estimated using the conventional Cox's proportional hazards model with adjustments for relevant baseline and time-dependent covariates. Analysis will be performed for each country separately.

Meta analysis of individual patient data The individual patient data from the 6 cohorts will be pooled into a single dataset, and the primary and dose response analyses repeated using the pooled dataset. The hazard ratio (HR) estimates with 95% CIs for each pioglitazone exposure definition will be estimated using Cox's proportional hazards model with cohort dataset included as a categorical covariate. The hazard modelling will include adjustments for relevant baseline and time-dependent covariates.

Sample size:

	Finland	**Sweden**	**NL Pharmacy-Hospital**	**UK CPRD GOLD-HES subset**	**NL GP**	**UK CPRD GOLD**
Ever exposed to pioglitazone	4 021	4 067	10 680	17152	2 276	31 185
Never exposed to	40 210	40 670	106 800	171 520	22 760	311 850

pioglitazone (10:1 matched)						
Total matched cohort	44 231	44 737	117 480	188 672	25 036	343 035

Statistical power of individual patient data meta-analysis

	Pooled dataset
Follow-up time	5.5 years
Cumulative probability of developing bladder cancer (%)	0.19
Power (%)	
RR=1.2	**34.4**
RR=1.3	**58.0**
RR=1.4	**76.3**
RR=1.5	**87.8**
RR=2.0	**99.7**

Ethics: All the information is obtained without any potential harm to patients. This is fully a register-based study and patients will not be contacted in any phase of the study. All patient data is anonymous which ensures the full data protection of patients. Approval from relevant Ethical Review Boards will be requested before conducting the study.

The investigators and the Sponsor and individuals acting on their behalf commit to adhere to the rules of the ENCePP Code of Conduct in their entirety. The study has been registered into the ENCePP's E-register of studies and will be submitted for the ENCePP Seal of Approval.

Results and publications: The principal investigator and co-investigators will jointly write the pharmacoepidemiologic analysis reports. The reports are delivered to Sponsor. Based on these results the investigators will co-author scientific manuscript(s) of the results to be published. A summary of the main results of the study, whether positive or negative and including results from prematurely terminated studies, will always be made available to the public. An abstract of the study findings will be provided through the ENCePP E-register of studies within three months following the final study report. The principal investigators may ask the ENCePP Secretariat to delay the publication of this abstract for a limited period pending response to peer-review comments. 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. The final study report may be shared with regulatory agencies as required.

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
CHF	Congestive Heart Failure
CHMP	Committee for Medicinal Products of Human Use
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink Group
DDD	Defined daily dose
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
FIN	Finland
HR	Hazard ratio
ICD-9	International classification of diseases, 9th revision
ICD-10	International classification of diseases, 10th revision
ICD-O-3	International classification of diseases for oncology, 3rd Edition
ICPC	International classification of primary care
ISPE	International Society for Pharmacoepidemiology
NL	The Netherlands
NL GP	PHARMO General Practice Dataset
PGJ2	Prostaglandin J2
PID	Personal identification number
PPAR	Peroxisome proliferators-activated receptors
PUNLMP	Papillary urothelial neoplasia of low malignant potential
RXR	Retinoid X receptor
SID	Study identification number
SWE	Sweden
TZD	Thiazolidinediones
UK	United Kingdom
UTI	Urinary tract infection
Vnr	Nordic article number

5. AMENDMENTS & UPDATES

Version 1.0 (dated 22 Jan 2013)

- None

Version 2.0 Current version (dated 20 June 2013)

- Objectives
  - The original seven objectives classified as one primary objective and six secondary objectives
- Co-investigators
  - The list of co-investigators broadened to include all persons involved in the design of the study protocol.
- Inclusion and exclusion criteria
  - Criteria consolidated across databases to include patients over 40 years using any antidiabetic medication and further exclude patients with diagnosis of type 1 diabetes, gestational diabetes, or drug induced diabetes prior to cohort entry.
  - The use of other thiazolidinediones but pioglitazone prior to cohort entry allowed. Preliminary data on exposure from PHARMO and CPRD records suggest that up to 30% of patients starting pioglitazone have been exposed earlier to other TDZs. To increase the power of the study these patients will be included, and the history of prior TDZ use used in cohort matching. Stratified analysis will be performed to explore the effect of prior TDZ use on the association between pioglitazone exposure and bladder cancer incidence. Follow-up time will be censored at start of other TDZs after cohort entry
- Bladder cancer
  - Bladder cancer as an underlying cause of death added as source of information on bladder cancer
  - Diagnosis codes specified in detail in the definition of bladder cancer
- Matching
  - Variables used for the matching within the propensity score and outside of the propensity score reorganised and defined.
- Other antidiabetic drugs
  - Categories grouped into most common regimens
- Co-morbidities, drugs and exposure algorithm
  - Code lists updated
  - Details moved into appendices
- Analysis
  - Definitions of the crude, base and adjusted models defined in detail. Fully adjusted model not used as a term anymore.
  - Stratified analyses added.
  - The sensitivity analysis of the impact of exclusion of unmatched pioglitazone-exposed patients described in more detail. Sensitivity analysis with cumulative duration of insulin added. Sensitivity analysis where pioglitazone exposure group is changed from at least one prescription or dispensing of pioglitazone to at least 2 prescriptions or dispensing within a 6 month period added. Analysis of the association of ever vs. never use of pioglitazone with bladder cancer incidence when insulin use is included as a cumulative duration in the adjusted model. Sensitivity analysis for incidence of neoplasm of uncertain or unknown behaviour added.

- A separate meta-analysis added by using a subset of the pooled dataset including those datasets where information on smoking, BMI, HbA1C and urinary/renal markers is also available.

- Data sharing
  - Study to be registered for the ENCePP Seal of approval
  - Criteria and process for sharing the analytical country specific datasets and meta-analysis dataset given in appendix

6. MILESTONES

Activity	**Date**
Final updated protocol submitted to EMA	31 Jan 2013
Research Ethics approval for data linkage & access[1]	30 Apr 2013
Datasets available[2]	
NL, UK	30 July 2013
FIN, SWE	30 Oct 2013
Pooled dataset for individual patient data level meta-analysis	30 Dec 2013
Analyses Completed[3]	
NL, UK cohorts	30 Apr 2014
FIN, SWE cohorts	30 Apr 2014
Individual patient data level meta-analysis	30 Jun 2014
Draft Study Report to Sponsor	
6 Individual cohort studies	30 Jun 2014
Individual patient data level meta-analysis	30 Aug 2014
Study reports to EMA	30 Sep 2014

[1] Research Ethics approval typically takes 3 months in each country. All countries require the finalised protocol before submission for research ethics approval.

[2.] Dataset availability largely dependent on delivery of linked data from cancer registries and other sources in each country. Best estimates are that research dataset should be available 3 months after research ethics approval in NL and UK, and 6 months for FIN and SWE.

[3.] Analysis of NL and UK datasets, including sensitivity analyses (1-3, 5-7) is estimated at 9 months from availability of research dataset. Analysis of SWE and FIN datasets, with sensitivity analyses (1-4) is estimated at 6 months from availability of research dataset.

7. RATIONALE AND BACKGROUND

Peroxisome proliferators-activated receptors (PPARs) are members of the nuclear hormone receptor super family of transcription factors whose activities are regulated by high-affinity binding of small, lipophilic ligands such as steroid hormones, vitamin D, retinoids, and thyroid hormone. PPAR alpha, delta and a third subtype called gamma are related sufficiently to be considered members of a subfamily, and have similar properties including DNA binding specificity and heterodimerization with retinoid X receptor (RXR), whose ligands also activate the PPAR/RXR heterodimers. Ligands selective for PPAR gamma include prostaglandin J2 (PGJ2) derivatives, such as 15-deoxy-Δ12,14-PGJ2, and anti-diabetic thiazolidinediones (TZD) compounds, including troglitazone, rosiglitazone, and pioglitazone.

The Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency recently finalized its review on pioglitazone use and occurrence of bladder cancer. Evidence from several recent epidemiological studies suggests a small increased risk of bladder cancer with pioglitazone use in patients with type 2 diabetes [Lewis *et al*. 2011; CNAMTS 2011]. In the 4th interim analysis (8-years follow-up) of the Kaiser Permanente Northern California Study cohort, the hazard ratio (HR) of ever vs. never use of pioglitazone was 1.07 (95% CI 0.87-1.30) and a non-significant elevated risk among patients with longest >4 years duration of therapy (HR=1.30, 95% CI 0.91-1.86) [Lewis et al. 2012]. In the French CNAMTS study a similar risk for pioglitazone use was observed (HR=1.22, 95% CI 1.05-1.43), and a significant risk with cumulative doses over 28 000 mg (HR=1.75, 95% CI 1.22-2.50), and with duration of therapy between 12 to 23 months (HR=1.34, 95% CI 1.02-1.75) or over 24 months (HR=1.36, 95% CI 1.04-1.79) [CNAMTS 2011]. The CHMP agreed that further evidence is still needed. Previous observational studies were limited in their ability to control for potential channelling bias, detection bias and confounding by important bladder cancer risk factors such as BMI and smoking.

Limited information is also currently available on whether the potential increase in risk is associated with a particular type or stage of bladder cancer. Data from the KPNC cohort study showed any increased risk was largely limited to early stage tumours (Lewis et al 2012). This finding has yet to be confirmed in another study.

8. RESEARCH QUESTION / STUDY HYPOTHESIS

The objectives of the study are:

Research Questions:

Primary objective:

1) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes who are ever exposed to pioglitazone vs. never exposed to pioglitazone.

Secondary objectives:

2) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with increasing duration of pioglitazone treatment,
3) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with increasing cumulative dose of pioglitazone treatment,
4) To examine the association between pioglitazone exposure and risk of bladder cancer with respect to possible channelling bias, detection bias and other sources of confounding,
5) To characterize the stage and grade of the bladder cancer cases at the time of diagnosis in patients with type 2 diabetes who are ever exposed to pioglitazone vs. who are never exposed to pioglitazone, and

6) To characterize all-cause mortality and bladder cancer-specific mortality pattern amongst patients ever exposed to pioglitazone vs. never exposed to pioglitazone.

In addition an exploratory objective of the study is:

7) To estimate and compare the absolute and relative risk of bladder cancer in patients with type 2 diabetes with time since the last dose of pioglitazone use.

9. RESEARCH METHODS

9.1 Study design

Retrospective observational cohort studies will be undertaken using six datasets in four EU countries: Finland, Sweden, the Netherlands, and United Kingdom.

Each dataset will be analysed separately, then pooled into a single data set and this pooled anonymous individual patient level dataset will be used in meta-analysis.

The cohort studies

The previously observed weak association between long term pioglitazone use and risk of bladder cancer could in part be due to channelling bias because pioglitazone is more likely to be prescribed to patients with more advanced diabetes or who are obese. Advanced diabetes and obesity are both independent risk factors for developing bladder cancer, and the preferential prescribing of pioglitazone to such patients is could lead to biased estimates of risk.

To minimize channelling bias, each pioglitazone-exposed patient (exposure group) will be matched with up to 10 pioglitazone-unexposed diabetic patients (reference group) based on:

- Propensity score (the probability to initiate pioglitazone therapy) will be evaluated with a logistic regression model using the following proxy measures of severity of diabetes and variables affecting the prescription of pioglitazone. The number and exact definitions of variables included in the propensity score may vary between the databases.
    - Duration (years) of treated diabetes mellitus at cohort entry
    - History of diabetic complications at cohort entry
    - History of myocardial infarction or stroke at cohort entry
    - History of congestive heart failure at cohort entry
    - Year of cohort entry
    - Duration (years) of membership in medication database prior to cohort entry
    - Number of different antidiabetic drug classes prior to cohort entry
- Use of other thiazolidinediones (other than pioglitazone) prior to cohort entry
- Type of antidiabetic treatment immediately prior to cohort entry classified as
    - no treatment
    - metformin only
    - sulphonylureas only
    - metformin and sulphonylureas
    - insulin with or without any other antidiabetic medication

  - other (including use of other thiazolidinediones)
- Type of modification in baseline antidiabetic therapy at cohort entry classified as treatment switch or introduction of an add-on treatment
- Living in geographical area for which cancer data are available (used in the Netherlands only)

Patients in the pioglitazone-exposed group will be matched to patients in the reference group to within ± 0.05 of the propensity score and exactly within the classes of the other matching variables. If any pioglitazone-exposed patients do not have any match, the matching will be repeated for these patients by loosening the matching range upto ± 0.1 for the propensity score. Caution is needed not to loosen the matching range too much because this would reduce the comparability of the groups.

Channelling bias due to baseline BMI and HbA1C level is also likely. BMI is a moderately strong independent risk factor for bladder cancer (RR≈5) and the EU SmPC specifically recommends pioglitazone use among overweight T2DM patients. As BMI and HbA1C data are only available in some datasets, these are included as covariates in the data analyses rather than as cohort matching variables.

9.2 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. 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. For detailed ICD-10, ICD-9, ICPC and READ codes see Appendix 3.
- Patients who are entitled to special reimbursement for diabetes with ICD-10 diagnosis starting with ‘E10’, ‘E12’, ‘E13’, ‘O24.4’ indicative of type 1 diabetes, diabetes with malnutrition, other diabetes or gestational diabetes, respectively (specific to FIN dataset only).
- Diagnosis or history of bladder cancer or in-situ bladder cancer prior to cohort entry.

9.3 VARIABLES

9.3.1 Outcome variable

Bladder cancer

Date of diagnosis of the first incidence of bladder cancer after the entry into the study cohort will be used as the primary outcome date. The bladder cancer definition will include

- malignant neoplasm of the bladder and
- carcinoma in situ of the bladder.

Sensitivity analysis will also look the following additional outcomes:

- neoplasm of uncertain or unknown behaviour of the bladder.

In the FIN and SWE datasets information from cancer registries will used to define incident bladder cancers. Other datasets will use hospital discharge and GP records to find incident bladder cancers. In the cancer registries, malignant bladder cancer is coded to ICD-O-3 code C67 with behaviour code "3". In the Hospital discharge data, malignant bladder cancer is coded to ICD-9-CM 188 or ICD-10 code C67. In NL GP, bladder cancer is coded to ICPC code U76, and in the UK CPRD is coded to READ codes B49.

The specific codes malignant neoplasm of the bladder, carcinoma in situ of the bladder and neoplasm of uncertain or unspecified behaviour of the bladder are given in the following table.

	ICD-10 CM	ICD-9 CM	ICD-O-3	ICPC	READ
Malignant neoplasm of the bladder	C67	188	C67 with behaviour code "3"	U76	B49
Carcinoma in situ of the bladder	D09.0	233.7	C67 with behaviour code "2"	*	B837
Neoplasm of uncertain or unknown behaviour of the bladder	D41.4	236.7 239.4	C67 with behaviour code "1" (PUNLMP with morphology code M-8130/1	*	B917, BB4111, BA04

* Text mining of the database contents used to identify in situ carcinomas and neoplasms of uncertain or unknown behaviour of the bladder

In the FIN, SWE and UK CPRD GOLD-HES datasets, there is 100% linkage to cancer registries. The data collection on all cancer cases is compulsory and the informants submitting data on cancer patients to the registry include all hospitals, physicians, pathological, cytological and haematological laboratories, and dentists. The data collected by these registers include demographic information on the patient, medical data such as the primary site of the tumour (International Classification of Diseases for Oncology code), date of diagnosis, and tumour type and grade. In the FIN, SWE and UK CPRD GOLD-HES datasets, bladder cancer will be based on cancer registry-based information.

The NL Pharmacy-Hospital and NL GP datasets only have partial linkage to the cancer registry data (20-30% in NL Pharmacy-Hospital and 10% NL GP datasets). It is thus not possible to use cancer registry-based diagnosis for all subjects. GP records data (NL GP dataset) and Hospital discharge data (NL Pharmacy-Hospital dataset) will be used to identify bladder cancers. For patients living in geographical areas that can be linked to the

cancer registry, this method of outcome detection will be validated and supplemented using the cancer registry data.

The CPRD GOLD dataset is not linked to the cancer registry data. Diagnosis of bladder cancer will be based on GP records data only.

All cause mortality and bladder cancer mortality

Information on date of all deaths among cohort members, and deaths where bladder cancer is recorded as the cause of death will be obtained from the national death registries for FIN, SWE, and UK CPRD GOLD-HES datasets. A recent validation study in the UK CPRD Gold demonstrated 99% concordance between GP's mortality record and national death registry entries. Cause-specific mortality is not available for the two NL datasets or the UK CPRD GOLD dataset. Bladder cancer deaths will be identified as deaths with malignant bladder cancer (e.g. ICD-10 code C67) as the underlying cause of death.

9.3.2 Drug exposure

In FIN and SWE datasets, exposure is based on purchasing. In the NL GP, UK CPRD GOLD and UK CPRD GOLD-HES datasets, exposure is based on GP prescriptions. In the NL Pharmacy-Hospital dataset, exposure is based on dispensing.

Diabetic medications are classified into the following antidiabetic drug exposure groups: pioglitazone, metformin, sulphonylureas, other thiazolidinediones, alpha-glucosidase inhibitors, DPP-4 inhibitors, GLP-1 agonists, meglitinides, amylin analogues, insulins and other oral antidiabetic medications. Based on the active substances contained in combination products, they are allocated to several exposure groups. Further details of the diabetic medications are given in Appendix 1.

The amount of a specific drug is given as the number of defined daily doses (DDDs) or as actual amount dispensed or prescribed depending on the database. DDD describes how long the prescription would last if taken as one DDD per day and can converted into total amount contained using the Nordic article number which identifies the medicine package. Based on total amount of drug contained in each prescription will be modified into a drug use period by defining the length of the prescription plus an additional 50% treatment gap period [Nielsen et al. 2008, 2009]. The purpose of the treatment gap period is to avoid unnecessary breaks between two prescriptions. The resulting consecutive drug use periods with possible gaps between them are used to define a time-dependent current exposure periods indicating in which treatment categories the patient is at any given time during the follow-up.

Based on the prescriptions/purchases/dispensing of pioglitazone and the subsequent consecutive drug use periods the following time dependent variables for the pioglitazone exposure will be constructed:

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 will be identical to those used in the KPNC final (10-year follow-up) analysis. A secondary analysis will use the duration of exposure category from the KPNC 3rd interim (5-year) analysis: Never, <12 months, 12-24 months, and more than 24 months.
Cumulative dose (T)	Time-dependent cumulative sum of drug consumption based on the daily dosageof pioglitazone containing prescriptions or dispensings since entry into the cohort. The cumulative dose categories will be identical to those used in the KPNC final (10-year follow-up) analysis. A secondary analysis will use the exposure category from the KPNC 3rd interim (5-year) analysis: Never, 1-10,500mg, 10,501-28,000mg, and more than 28,000mg.
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, 2-3 years, 3-4 years, 4+ years, Never†
† The category levels are preliminary, and will be based on the actual exposure distribution	

Further details about calculation of the exposure are given in Appendix 2.

9.3.3 Propensity score and other cohort 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.
History of diabetic complications at cohort entry (F)	History of any of the following: • Diabetic retinopathy or maculopathy (N/Y) • Peripheral neuropathy (N/Y) • Chronic kidney disease (N/Y) • Proteinuria (micro or macro) (N/Y) • Diabetic nephropathy (N/Y) • Ketoacidosis (N/Y) • Diabetic coma (N/Y). Definitions of the individual variables are given in Appendix 3.
History of myocardial	Classified as (Y/N), for detailed definition see Appendix 3.

infarction or stroke at cohort entry (F)	
History of congestive heart failure at cohort entry (F)	Classified as (Y/N), for detailed definition see Appendix 3.
Year of cohort entry (F)	Calendar year of cohort entry
Duration of database membership before cohort entry (F)	Duration (years) of membership in medication database prior to cohort entry.
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 TDZs, alphaglucosidase inhibitors, DPP-4 inhibitors, GLP-1 agonists, meglitinides, amylin analogues, insulin, other. Combination products contribute separately to each drug class based on the active substances included in the product.
Other 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 pharmacotherapy • 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 cohort entry classified as treatment switch or addition of new treatment to all prior treatment (if any exisited). See Appendix 2.
Geographical area (F)	Living in geographical area for which cancer data are available (Y/N, used in the Netherlands only)

9.3.4 Other explanatory variables

In the hospital-based morbidity datasets (FIN, SWE, NL Pharmacy-Hospital and UK CPRD GOLD-HES dataset) some covariates are not available, while others may be based on hospitalizations (more severe cases) or on use of specific medications.

In the two GP-based morbidity datasets (NL GP dataset and UK CPRD GOLD) covariates are based on GP records or use of specific medications.

Further details of the variable definitions for each dataset are given in Appendix 3 and Appendix 4.

Covariates (F = fixed at cohort entry, T = time dependent)	**Description**
Gender (F)	M/F
Age (F)	Age at cohort entry
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'. No smoking data is available in FIN, SWE & NL Pharmacy-Hospital datasets. 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 available, for detailed definition see Appendix 3. 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'. No BMI data is available in FIN, SWE & NL Pharmacy-Hospital datasets.
HbA1C (F, T)	Classified as missing, <7.5%, 7.5-8.9%, ≥9.0%. Baseline HbA1C measurement will be 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. No HbA1C data is available in FIN, SWE & NL Pharmacy-Hospital datasets
PSA elevated (T)	PSA elevated at any given time during the follow-up. Classified as Never vs. ever elevated; and Never elevated vs. Elevated vs. Not elevated. No PSA data in FIN, SWE & NL Pharmacy-Hospital datasets.
Number of PSA tests (T)	Cumulative number of PSA tests during the follow-up period.
Number of urine protein tests (T)	Cumulative number of urine protein tests during the follow-up period
History of diabetic complications (For details see Appendix 3)	
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.
Peripheral neuropathy (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.
Proteinuria (micro or macro) (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 nephropathy (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 kidney 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.
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.
History of relevant comorbidities (For details see Appendix 3)	
Other 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.
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.
Other 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 (for details see Appendix 4)	
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.
History of bladder comorbidities (For details see Appendix 3)	
Urinary incontinence (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.
Urinary tract infection (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.
Pyelonephritis (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.
Urolithiasis (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.
Hematuria (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.
Urinary Retention (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.
Neurogenic bladder (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.
Catheterization (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.

9.4 FOLLOW-UP PERIOD

Period covered by the datasets

Dataset	Start of coverage	End of Follow-up
FIN	01 January 2000	30 June 2011
NL Pharmacy-Hospital	01 January 2000	30 June 2011
NL GP	01 January 2000	30 June 2011
SWE	01 July 2005	30 June 2011
UK CPRD GOLD	01 January 2000	30 June 2011
UK CPRD GOLD-HES subset	01 January 2000	31 December 2010

Cohort entry / Start of Follow-up

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

Baseline observation period before cohort entry

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 1.1.1998 onwards in Finland, from 1.7.2005 onwards in Sweden, and from start of database membership for the NL and 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).

End of Follow-up

In the bladder cancer analyses, each patient will be followed-up from cohort entry until the date of diagnosis of the first incident bladder cancer, start of other thiazolidinediones (other than pioglitazone) at or after cohort entry, end of membership of the database, end of database coverage, death or 30 June 2011, whichever occurs first.

In the mortality analyses, each patient will be followed-up from cohort entry until death, start of other thiazolidinediones (other than pioglitazone) at or after cohort entry, end of membership of the database, end of database coverage or 30 June 2011, whichever occurs first.

9.5 CENSORING OF FOLLOW-UP

The NL GP, NL Pharmacy-Hospital, UK CPRD GOLD and UK CPRD GOLD-HES datasets have definable entry and exit dates for each person, so mortality data is not required for censoring. Exit date will be used as a follow-up censoring date.

The FIN and SWE death registers provide data on dates and causes of death with personal identification number, sex, age, place of residence and causes of death reported by ICD-10 diagnosis codes, and will be used to censor follow-up. Date of death will be used as a censoring event in the bladder cancer incidence analyses,

and as an endpoint for the mortality analyses. 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 to censor follow-up. The NL PHARMO Database Network contains dates of death from hospital, pharmacy and GP records, which are supplemented through linkage to national death registries, which provide date of death, but not cause of death.

9.6 DATA SOURCES

Finland and Sweden both have well-developed population-wide register systems with longitudinal follow-up data. The persons are identified in the registers with a unique personal identification number (PID) and thus the records can be linked for research purposes on subject level between the registers. Permissions to use the data may be received upon providing a research plan to the authorities responsible for the registers. The protocol is also subjected to relevant Ethical Committee´s review and approval. The data permit processes and timelines for data permit granting by the authorities responsible for the registers vary between different registers and countries, as well as the requirements for the local research collaboration and the publication of results. Registers that are used in the study are described below in detail. The starting years of the relevant data sources for the proposed study are shown in the Table below.

In Finland, the nationwide prescription registers contain information on outpatient medication purchases in pharmacies [Furu *et al.* 2010]. The register contains information on all purchased prescribed medicines within the reimbursement scheme excluding, for example, some very cheap medicines and contraceptives. In addition, patients who are entitled for special refunds (72% or 100% reimbursement) of medicine expenses based on their chronic condition can be identified using the Registry for Reimbursed Medications. For example, the special reimbursement category 103 is for diabetes and the category 203 for "Chronic asthma and similar chronic obstructive pulmonary disease" includes COPD patients with the International Classification of Diseases (ICD-10, 10th revision) diagnose code 'J44'. This database is used for identification of comorbidities.

In Sweden, the nationwide prescription registers contain information on outpatient medication purchases in pharmacies [Furu *et al.* 2010]. All prescribed medicines purchased in community pharmacies are included irrespective of reimbursement status.

The study data will be extracted from the relevant registers. The extraction process, subject to country specific adaptations, is as follows. First, patients in the study cohort with the unique personal identification (PID) number are identified from the prescription register of each country. A unique dummy study identification number (SID) is created for each PID by the prescription register holder. The list with the PID-SID pairs will be provided from the prescription register holders to the other register holders. Each register holder extracts the relevant data according to the study protocol and links the data to the PID - SID list. Each register holder decodes the data by destroying the key between the PID and SID permanently. Finally, each register holder provides the de-identified study data to EPID Research. Thus, only de-identified data will be provided to EPID Research for performing the analysis of the study data.

Table Starting years of the relevant registers in Finland and Sweden

Data	Finland	Sweden
Drug exposure (antidiabetic drugs and	The Finnish Prescription Register/ The Registry for	The Swedish Prescribed Drug Register

other relevant drugs)	Reimbursed Medications **1994**	**July 1, 2005**
Cancer	Finnish Cancer Register **1953**	The Swedish Cancer Registry **1958**
Mortality	The Cause of Death Register **1969**	The Causes of Death Register **1952**
Co-morbidities	The Finnish Prescription Register/ The Registry for Reimbursed Medications **1994** The Finnish Hospital Care Register **1967**	The Swedish Prescribed Drug Register **July 1, 2005** The National Patient Register **1987**
Immigration and emigration dates	The Finnish Population Information System Register **1950**	The Swedish total population register
Institutionalization	Institutional Care Register (Social HILMO) **2000**	

The PHARMO database network is a population-based patient-centric data tracking system which started in 1986 that includes high quality information of patient demographics, drug dispensing, hospital morbidity, clinical laboratory, pathology and general practitioner and other information. The PHARMO databases are linked to form a database network on a patient level and contain information required for the study. The PHARMO Pharmacy-Hospital dataset is part of the PHARMO database network and contains information for 3.2 million community-dwelling inhabitants of 65 municipal areas in the Netherlands. Some databases (Cancer registry, GP, clinical laboratory, in hospital pharmacy data) only overlap partially with the 3.2 million patients included in the PHARMO Pharmacy-Hospital dataset. To gather the most information possible the PHARMO GP dataset is analyzed separately from the PHARMO Pharmacy-Hospital dataset.

The PHARMO Community Pharmacy database is a nationally representative database that was started in 1986 and includes dispensing information from community pharmacies in the Netherlands. The database includes all GP or specialist prescribed and pharmacy dispensed healthcare products on the Dutch market. Of each dispensed drug, the Anatomical Therapeutic Chemical (ATC) code, dispensing date, prescriber, prescribed dosage regimen, dispensed quantity, costs and the estimated duration of use are available. The costs of drugs are estimated from a third payer perspective and represent reimbursement costs per dispensing record including VAT. All drug dispensing data in the pharmacy database are encoded according to standards based

upon the Z-Index drug database (http://www.z-index.nl/). It is possible to identify and classify drug use followed over time, both on the basis of national and international classification schemes and on the basis of a drug's individual active ingredients and administration forms. The database covers approximately 12% of the total Dutch population each calendar year, is available in-house, and does not require separate permission for use in each study.

The Hospitalization database (Dutch National Medical Register (LMR) (www.dutchhospitaldata.nl) is a national database comprising all hospital admissions in the Netherlands, i.e., admissions for more than 24 hours and admissions for less than 24 hours for which a bed is required. These records include detailed information concerning the primary (mandatory) and secondary (optional) discharge diagnoses, procedures and dates of hospital admission and discharge. All diagnoses are coded according to the International Classification of Diseases, 9th edition (ICD-9-CM). Procedures are coded according to the Dutch classification of procedures ("CvV - Classificatie van Verrichtingen"). The part of the database that is linked to the PHARMO community pharmacy database is available in-house at PHARMO and does not require separate permission for use in each study. All hospitalizations of patients represented in the community pharmacy database are covered by the linkage. The linked dataset is referred to as NL pharmacy-hospital dataset throughout this document.

The Eindhoven Cancer Registry (ECR) is a population-based registry (covering a demographic region with 2.4 million inhabitants) which is maintained by the Comprehensive Cancer Centre South and collates records on all newly diagnosed cancer patients in the south-eastern part of the Netherlands. The ECR is notified for new cases of cancer by six pathology departments, ten general hospitals and two radiotherapy institutes. Trained registry personnel subsequently actively collect on site data on patient characteristics, diagnosis, tumour staging, co-morbidity at diagnosis, socioeconomic status at diagnosis and treatment received directly after diagnosis (e.g. chemotherapy (yes/no), radiation therapy and surgery). The ECR is linked with the PHARMO Pharmacy-Hospital dataset. The overlap is approximately 30% (1 million patient lives). This database is not a PHARMO in-house database, but an external (partnership) database and permission of a scientific review board is needed to access the data.

The PHARMO GP database is part of the PHARMO database network. The database is a longitudinal observational database that contains data from computer-based patient records of general practitioners (GPs) throughout the Netherlands, who voluntarily chose to supply data to the database. GPs have complete control usage of their data, through the Steering Committee and are permitted to withdraw data for specific studies. Collaborating practices are located throughout the Netherlands and the collaborating GPs are comparable to other GPs in the country based on age and gender.

The PHARMO GP database currently includes about 1.5 million patients. This is the cumulative number of patients who have ever been part of the dynamic cohort of patients who were registered. Turnover occurs as patients move and transfer to new practices. The records of 'transferred out' patients remain in the database and are available for retrospective studies with the appropriate time periods.

Detailed clinical information and prescription drugs data are captured in this database, including among other things demographic data (age, sex, patient identification, GP registration information), diagnoses, physician-linked indications for therapy, comorbidities, drug prescriptions, laboratory values (e.g., potassium, creatinine), doctor in attendance, referrals to specialists, and a 'medical chart' containing free-text as noted by the general practitioner. Diagnoses and symptoms are recorded based on the International Classification of Primary Care (ICPC), which can be mapped to ICD-9 codes, but diagnoses and complaints can also be entered as free text, prescription data such as product name, quantity dispensed, dosage regimens, strength, and indication are

entered into the computer. Prescriptions include medications prescribed by the GP (please note that specialist prescriptions are generally not included in the GP data). The National Database of drugs, maintained by the Royal Dutch Association for the Advancement of Pharmacy, enables the coding of prescriptions, according to the Anatomical Therapeutic Chemical classification scheme recommended by the World Health Organization. The system complies with European Union guidelines on the use of medical data for medical research and has been validated for pharmaco-epidemiological research. For confidentiality reasons, the database is strictly anonymized. The database is available in-house and no separate permission is required for use in each study. The dataset for this study is referred to as NL GP dataset throughout this document.

The NL GP dataset is linked with the Eindhoven Cancer Registry through the pharmacy dataset. Overlap between GP dataset, pharmacy dataset and the Eindhoven cancer registry is approximately 150.000 patient lives.

UK CPRD GOLD -HES contains the anonymised longitudinal medical records managed by GPs working the NHS primary care setting. Approximately 8% of all GPs contribute to the CPRD database, which has collected data prospectively from 1987. The data contains demographic information, diagnoses, prescriptions, tests and referrals. Approximately 56% of CPRD GOLD practices take part in the English linkage program (CPRD GOLD-HES subset), which includes patient level linkage to the inpatient Hospital Episode Statistics (HES), allowing for hospitalization diagnoses and procedures to be captured. Records in both settings include the unique NHS number, which is used for linkage purposes. In a similar way, CPRD GOLD data is linked to Cancer Registry (including histology, stage and grading) and Death Certificate data (including date and cause). Protocols using CPRD GOLD and linked data are subject to approval by an Independent Scientific Advisory Committee.

9.7 SAMPLE SIZE

The estimated number of patients who were ever exposed and who were never exposed to pioglitazone and estimates number of bladder cancers expected during the follow-up are presented in the table below for each dataset.

Power calculations for

i) each of the six datasets separately,
ii) pooling all morbidity datasets for meta-analysis

are presented for the comparison of ever vs. never exposed to pioglitazone. In pooling of all datasets the UK CPRD-HES dataset has been dropped, as it is a subset of the UK CPRD GOLD dataset. In addition overlap of the NL datasets must be removed (~10%). Power calculations for the cohort study (1 to 10 matching) are presented. For individual datasets the assumed data specific background rates and follow-up times given in the first table, and for pooled datasets the average values provided are presented in the second table.

For the cohort study the power is calculated for the effect sizes 1.2 – 2.0 on the relative risk scale. A two-sided type 1 error rate of 5% is used. 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].

Table Estimated patients exposed and unexposed to pioglitazone and power calculations for the cohort design for individual datasets and pooled datasets.

	Finland	**Sweden**	**NL Pharmacy-Hospital**	**UK CPRD GOLD-HES subset**	**NL GP**	**UK CPRD GOLD**
Ever exposed to pioglitazone	4 021	4 067	10 680	17152	2 276	31 185
Never exposed to pioglitazone (10:1 matched)	40 210	40 670	106 800	171 520	22 760	311 850
Mean follow-up time (years)	10.5	4	5.2	5.2	5.2	5.2
Total person years	464 426	178 948	610 896	981 094	130 187	1 783 782
Annual Incidence rate (Incidence / 100 000)	30.6	50.5	33.5	32.8	33.5	32.8
Cancer cases (total)	142	90	205	322	44	585
Power (%) Individual Cohort Studies						
RR=1.2	8.9	7.4	10.6	13.9	6.2	21.5
RR=1.3	12.5	9.7	16.0	22.4	7.3	36.6
RR=1.4	16.7	12.4	22.0	31.8	8.5	51.8
RR=1.5	21.1	15.1	28.2	41.1	9.8	64.8
RR=2.0	40.5	27.9	54.1	73.5	15.9	93.7

Table Power calculations for the cohort design (1 to 10 matching) for pooled dataset.

	Pooled dataset
Follow-up time t (years)	5.5
Annual incidence rate **λ** (Incidence/100 000)	33.7
Cumulative probability of developing bladder cancer (%)	0.19
Power (%)	

Cohort studies	
RR=1.2	**34.4**
RR=1.3	**58.0**
RR=1.4	**76.3**
RR=1.5	**87.8**
RR=2.0	**99.7**

9.8 Data Management

FIN and SWE:

The FIN component of the study will be undertaken by EPID Research and the Swedish part by CPE. Full audit trail starting from raw data obtained from register holders, and ending with statistical tables and graphs in reports will be maintained. In Finland data management, tabulations, graphics, and statistical modelling in carried out with R data-analyses language (http://www.r-project.org). R language is described more detailed in report "R: Regulatory Compliance and Validation Issues: A Guidance Document for the Use of R in Regulated Clinical Trial Environments" (http://www.r-project.org/doc/R-FDA.pdf). In Sweden all analyses and data management are carried out using SAS software version 9.3 (SAS Institute Inc. Cary, NC, USA). Source code of data management and data analyses is kept for inspection for five years after publication of results. The study may be inspected by the sponsor's study independent representative(s) or inspected by the competent authorities. Finland and Sweden will share a common statistical analysis plan (SAP).

NL:

The PHARMO Institute for Drug Outcomes Research will coordinate and conduct the study with the NL databases. PHARMO has direct access to the in-house de-identified data, but not the cancer registry data. The study is funded by Sponsor who has only access to the study results but not to the original data. All data analysis and reporting is carried out according to the study protocol. A team of researchers will conduct the study, which is coordinated by the local project leader. All programming will be independently reviewed by a trained researcher and all results will be Quality reviewed by a senior researcher.

All data will be analysed using SAS Enterprise Guide version 4.3 (SAS Institute Inc., Cary, NC, USA). All analyses will be conducted under Windows using SAS version 9.2. Source code of data management and data analyses is kept for inspection.

UK:

The CPRD GOLD and CPRD GOLD-HES components of the study will be undertaken by CPRD, and results provided to the Sponsor. All data analysis and reporting will be carried out according to the study protocol and subsequent data specification document. The study will be overseen by the Head of Research, managed and conducted by a senior researcher, with appropriate quality review. All data will be analysed using Stata v11.2 and coding kept and archived for inspection.

Pooled data set for meta-analysis

The pooled dataset will be held by the principal investigator in Finland, and data management will be handled as described above for Finland. The pooled dataset will contain the core set of variables that are available in all datasets. Smoking, BMI, HbA1C and other laboratory-based variable are not available in all datasets and will

thus not be included in the meta-analysis. As there is overlap of patients within the 2 NL PHARMO datasets and within the 2 UK CPRD datasets, individual patients will only be included once in the pooled dataset.

Data retention and archiving

All study data and supporting documents will be retained for a minimum of five (5) years after the study ends. Secure archives will be maintained for the orderly storage and retrieval of all study related material. An index shall be prepared to identify the location of the archived contents. Access to the archives will be controlled and limited to authorised personnel only.

9.9 DATA ANALYSES

The principles of the statistical analysis are outlined below. More detailed statistical analysis plans will be written separately for each dataset and for the meta-analysis.

Missing data

If a variable is totally missing from a database (for example smoking and BMI), it is excluded from the analysis in that database. If a variable is missing for only some of the patients within a database a missing data category is added and used in the analysis.

Population summary

The characteristics of the study population will be described with descriptive statistics at the entry into the study cohort. The following variables will be included in the population summary: age, gender, year of entry, duration of diabetes mellitus, history of relevant diabetic medications, history of relevant concomitant medications, history of diabetic complications, history of bladder comorbidities and history of other relevant comorbidities.

The population summary will be stratified according to pioglitazone exposure using never use vs. ever use as the exposure variable. Summaries for each country will be provided separately. Summaries pooling the country specific results will be provided for those variables that are available from each country.

Descriptive analysis

Comparisons for the various pioglitazone exposure definitions will be performed.

Crude bladder cancer incidence and 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 disease, history of relevant diabetic medications, history of relevant concomitant medications, history of diabetic complications, history of bladder comorbidities and history of other relevant comorbidities.

Crude incidence rates for each country will be provided separately. Summaries pooling the country specific results will be provided for those variables that are available from each country.

The incident bladder cancer cases occurring during the follow-up period are described in detail using the following information where available:

- patient's age at diagnosis

- tumour staging (behaviour, morphology code 5th digit),
- histology grade and differentiation (morphology code 6th digit),
- treatments used and their temporal association with the time of diagnosis.

Emphasis will be made to present the temporal association between the time of diagnosis and exposure to pioglitazone (ever vs. never; cumulative exposure; time since start, time since last dose).

Formal analysis of individual cohorts

The hazard ratio (HR) estimates with 95% CIs for each pioglitazone exposure definition 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. Separate analyses of bladder cancer incidence, bladder cancer mortality and all-cause mortality will be performed for each dataset.

A propensity score matched cohort analysis (Li et al 2012, Spreeuwenberg 2005) will be performed.

In the cohort analysis two models will be used: i) the crude model with pioglitazone exposure as ever vs never exposed included only and ii) the base model with age at cohort entry, sex and use of metformin, sulphonylureas, insulins or other antibiabetic drugs each classified as ever vs. never exposed included in the crude model. In addition an adjusted model will be created 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 bladder cancer incidence, all cause mortality and bladder cancer mortality as well as in the sensitivity analyses.

Stratified analyses

The risk models will be presented stratified, if there are sufficient event, by baseline variable used in the matching or propensity score:

- 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 proteinuria (micro or macro)
- History of diabetic nephropathy

Sensitivity analysis

The following sensitivity analyses will be undertaken:

1. An analysis to assess the impact of exclusion of unmatched pioglitazone-exposed patients in the propensity score matched cohort. The characteristics of the unmatched pioglitazone-exposed patients are compared to those of the matched pioglitazone patients. The bladder cancer incidence rates are compared between these two patient groups. If the incidence rates do not differ then no further analysis is required. If the incidence rates are different, then an incidence density sampling from the reference group will be performed to form a new reference group with similar distribution of the year of cohort entry as in the pioglitazone exposed patients. The resulting cohorts are analysed using a Cox model to assess the differences between the matched and unmatched analyses. This analysis will only be undertaken if more than 2.5% of the pioglitazone patients are not matched.
2. An analysis excluding all bladder cancers occurring within 12 months after cohort entry to allow for bladder cancer latency.
3. An analysis excluding adenocarcinomas and squamous cell carcinomas will be performed to look at risk within transitional cell urothelial tumours.
4. An analysis where follow-up time is censored when the first gap of at least 4 months occurs between two prescriptions will be performed to assess the impact of removing patients who have gone to nursing homes (FIN and SWE datasets only).
5. Analysis to assess the impact of adjusting/not adjusting for smoking status, BMI and HbA1C information in NL and UK datasets (NL and UK datasets only).
6. Analysis of the impact of using first smoking and BMI record after cohort entry for persons with no baseline smoking or BMI record on the adjusted risk estimates (NL and UK datasets only).
7. Analysis comparing risk estimates from full cohort vs. an incident sub-cohort. The incident sub-cohort will only include patients with at least 12 months of database membership before first diabetes treatment (NL and UK datasets only). Note: the SWE and FIN cohorts are incident cohorts.
8. Analysis of the association of ever vs. never use of pioglitazone with bladder cancer incidence when insulin use is included as a cumulative duration in the adjusted model.
9. Analysis in which the pioglitazone exposure group is changed from at least one prescription or dispensing of pioglitazone to at least 2 prescriptions or dispensing within a 6 month period. To ensure immortal time bias is not introduced, the cohort entry date will be amended for both exposed and unexposed patients to be the original cohort entry date plus six months.
10. Analysis in which the definition of incident bladder cancer is broadened to include neoplasms of uncertain and unknown behaviour. The bladder cancer risk models will be reanalysed, if there are at least 5% increase in the bladder cancer event count after inclusion of neoplasm of uncertain behaviour or unknown behaviour of the bladder in the definition of an incident bladder cancer event. This will assess the specificity of any observed association.

Sensitivity analyses will also be undertaken on any significant unplanned data handling decisions.

The following validation analysis will be undertaken:

- In the NL Pharmacy-Hospital, NL GP and UK CPRD GOLD-HES datasets, the cancer registry data, GP record data and hospital records data will be compared to assess the validity of using GP or hospital discharge data for bladder cancer diagnosis.

Meta-analysis

The pooled data set will be analysed using the same methods and approach as the individual cohort analyses. The Cox models will include a cohort dataset identifier as a categorical covariate. The modelling approach will be similar to the step wise modelling described above.

Two separate meta-analyses will be performed. The first meta-analysis will use a pooled dataset containing common variables from all datasets. If smoking, BMI, HbA1C and urinary/renal marker information can be obtained for the SWE dataset, then the second meta-analysis will use a subset of the pooled dataset including only datasets where information on these variables is also available (i.e. excluding FIN dataset and NL Pharmacy-Hospital dataset). If these variables cannot be obtained for the SWE dataset, then only the first meta-analysis will be performed.

9.10 Quality Control

The study will be conducted as specified in this protocol. All revisions to the protocol shall be properly documented as protocol amendments and when necessary such protocol amendments are delivered to register holder(s) whenever amendment(s) to the data permissions are required.

The study protocol has been written by following the Code of Conduct by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance [ENCePP 2011] that provides a set of rules and principles for post-authorization studies with regard to the best practices and transparency, thereby promoting scientific independence of such studies. The ENCePP is a project led by the European Medicines Agency to further strengthen the post-authorization monitoring of medicinal products in Europe by facilitating the conduct of multi-centre, independent, post-authorization studies focusing on safety and on benefit/risk. The study has been registered to the ENCePP's E-register (http://encepp.eu/encepp/viewResource.htm?id=3627). The results will also be published on the same site.

The study protocol also follows the key elements of the Guideline for Good Pharmacoepidemiology Practices by International Society for Pharmacoepidemiology [ISPE 2007], and the recent draft Guidance for Industry and FDA Staff "Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets" [FDA 2011].

9.11 Strengths and limitations of the research methods

Statistical Power

Several previous studies have observed that ever use of pioglitazone is weakly associated with bladder cancer (pooled relative risk= 1.2). The six individual datasets in this protocol are each too small to detect a RR<1.5. However, the meta-analysis of the pooled dataset has a predicted power of 85% to detect a RR= 1.5, and a power of 73% for detection of a RR=1.4. The meta-analysis is not sufficiently powered (32%) to detect very weak associations of RR= 1.2. This study has included the main European data sources that contain information to enable the controlling of potential channelling bias, detection bias and confounding. It is possible that there are some other small European datasets, but their inclusion would only have minimal impact on improving the study's power to detect very weak associations.

Study design

A potential limitation of using matching is that any pioglitazone-exposed patient without a matched control would be excluded from the analyses. A risk minimisation strategy has been included, whereby a wider matching criterion would be applied propensity score of to any unmatched cases after the first round of matching. In addition, if the unmatched pioglitazone users change the incidence rates of the outcomes, a sensitivity analysis will be undertaken to compare the matched analysis to an unmatched analysis with the propensity score included as a covariate.

Confounding and bias

While a weak association may be real and causal in nature, they are equally as likely to simply reflect differences between the exposed and reference group. Potential sources of difference between pioglitazone-exposed and pioglitazone-unexposed groups that could lead to misleading results include channelling bias, detection bias and uncontrolled confounding.

Pioglitazone-treated patients are likely to have more severe diabetes than patients treated with other antidiabetic medications, and severity of diabetes is believed to be an independent risk factor for bladder cancer. Pioglitazone is also recommended for use in obese patients, and obesity is another independent risk factor for bladder cancer. Channelling bias is probably the most important issues in pharmacoepidemiology and observational research studies. Previous observational studies have been limited in their ability to control for channelling bias. In this protocol, particularly attention has been given to channelling bias and its control by matching on propensity scores based on several proxy measures of diabetic severity.

Few of the previous studies have considered detection bias. Microalbuminuria and proteinuria may be more common in patients treated with pioglitazone. Diabetics with microalbuminuria or proteinuria are likely to undergo thorough renal and bladder examination and be subject to further surveillance than patients without microalbuminuria or proteinuria. This would lead to a bias with pioglitazone-exposed patients being more likely to have a bladder cancer detected because of urinary investigations. This study will examine the occurrence of microalbuminuria, proteinuria and other urinary bladder/renal signs and symptoms as possible sources of detection bias. The UK and NL datasets will have data on these variables, but the FIN and SWE hospital based discharge records do not contain laboratory data but include only diagnoses for proteinuria.

We attempt to control for the confounding bias by using baseline and time-dependent variables of potential confounders in the statistical analyses. Although HbA1C, smoking and BMI are not available in the hospital discharge datasets, sensitivity analyses in the UK and NL datasets will investigate the role they play on pioglitazone related risk, after adjusting for the channelling bias.

Exposure

Prescriptions do not contain information on the diagnosis. It is assumed that patients using oral antidiabetic medications (ATC code A10B) are of type 2 diabetic patients. A further criterion of age > 40 years has been added to avoid inclusion of patients with type 1 diabetes. Type 1 diabetes, diabetes with malnutrition, other diabetes, or gestational diabetes are considered as exclusion criteria.

Prescription databases do not contain information on medication use in hospitals or nursing homes. This may causes miss-classification of the current exposure and underestimation of the cumulative exposure and cumulative dose in some patients. The effect of such bias is evaluated in a sensitivity analysis where follow-up

time is censored when the first gap of at least 4 months occurs between two prescriptions. In Finland the information on institutionalization (other than hospitalization) is used to avoid such gaps.

The Finnish prescription database contains information on all medications purchased provided that their cost exceeds the threshold for basic reimbursement that applies to all prescribed medications. In this study bias due to possibly missing some of the purchased medications is not expected, because the patients are entitled to special reimbursement and therefore they are expected to have all their diabetes medications registered.

This study will use the same exposure categories as used in the final (10-year) follow-up of the KPNC cohort study. This will enable comparisons to be made between this study and the KPNC cohort study's 10-year follow-up analyses. The primary analyses will also be repeated using the same exposure categories as he KPNC 3rd interim (5-year) analysis. These same exposure categories were also used in the French cohort study (CNAMTS 2011).

Cancer data

The cancer registers have nation-wide coverage, except in NL where linkage is to a regional registry that contains more detailed information than the national registry. The data collection on all cancer cases is compulsory and the informants submitting data on cancer patients to the registry include all hospitals, physicians, pathological, cytological and haematological laboratories, and dentists. The completeness and accuracy of the register data are considered good for scientific research [Teppo et al. 1994; Korhonen et al. 2002; Barlow et al. 2009; van Herk-Sukel et al. 2011]. Therefore, reporting bias for the outcomes of interest in the current study is not considered to have major impact on the results.

Since 2007 all three morphology types of cancers have been registered into the Finnish Cancer Registry. Before that only malignant cancers and *in situ* cancers were included lacking PUNLMPs.

Loss to follow-up from nationwide registers (FIN and SWE) means emigration to other countries. In the UK and NL datasets attrition may be due to persons moving to a geographic area not covered by the PHARMO / CPRD Gold databases. It is unlikely that such attrition would be different between the treatment groups. Therefore attrition bias is not considered to have major impact on the results.

10. PROTECTION OF HUMAN RIGHTS

To ensure the full data protection of patients, all the research data in each country is anonymized. Approval from relevant Ethical/Research Review Boards will be required before conducting the study.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/REACTIONS

The study design is a retrospective cohort based on secondary use of existing health care records. Therefore, no adverse event or serious adverse reporting will be performed beyond the defined study deliverables.

12. COMMUNICATION OF STUDY RESULTS

The principal and co-investigators will write the study reports. 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 study has been registered to the ENCePP's E-register (http://encepp.eu/encepp/viewResource.htm?id=3627). The results will also be published on the same site. The principal investigator will seek ENCePP Seal of Approval. The criteria and process for sharing the analytical country specific datasets and meta-analysis dataset for third parties are described in Appendix 5. According to the ENCePP Code of Conduct, the principal investigator is responsible of publication of the results.

A summary of the main results of the study, whether positive or negative and including results from prematurely terminated studies, will always be made available to the public. An abstract of the study findings will be provided through the ENCePP E-register of studies within three months following the final study report. The principal investigator may ask the ENCePP Secretariat to delay the publication of this abstract for a limited period pending response to peer-review comments. The outcome of a study will always be presented in an objective and truthful manner providing a comprehensive and accurate description of the findings. In no way shall the interpretation and presentation of the results be aimed towards any commercial, financial or personal interests.

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

APPENDIX 1. ANTIDIABETIC MEDICATIONS

APPENDIX 2. CALCULATION OF EXPOSURE

APPENDIX 3. VARIABLE DEFINITIONS

APPENDIX 4. OTHER MEDICATIONS

APPENDIX 5. CRITERIA AND PROCESS FOR SHARING THE ANALYTICAL COUNTRY SPECIFIC DATASETS AND META-ANALYSIS DATASET FOR THIRD PARTIES

14. REFERENCES

Barlow L, Westergren K, Holmberg L, Talbäck M. The completeness of the Swedish Cancer Register – a sample survey for year 1998. *Acta Oncologica* 2009, 48: 27-33.

CNAMTS, Caisse nationale de l'assurance maladie. Risk of blabber cancer in people with diabetes treated with pioglitazone in France: a group study on SNIIRAM and PMSI data. Final report 7/06/2011. http://www.mhlw.go.jp/stf/shingi/2r9852000001hbq8-att/2r9852000001hd4o.pdf

Cancer Incidence in Sweden 2010. OFFICIAL STATISTICS OF SWEDEN, Statistics – Health and Medical Care. http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/18530/2011-12-15.pdf. 15 December 2011

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

Finnish Cancer Registry. [Online] Cancer Statistics at www.cancerregistry.fi, updated on 07.05.2012

FDA, Food and Drug Administration. Guidance for Industry and FDA Staff. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets. February 2011. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf

Furu K, Wettermark B, Andersen M et al. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol*. 2010, 106: 86-94.

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

Korhonen P, Malila N, Pukkala E et al. The Finnish Cancer Registry as Follow-Up Source of a Large Trial Cohort. *Acta Oncologica.* 2000, 41: 381-388.

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.

Lewis JD *et al*. Risk of bladder cancer among diabetic patients treated with pioglitazone. Interim report of a longitudinal cohort study. *Diabetes Care.* 2011, 34: 916-922.

Lewis JD *et al*. Risk of bladder cancer among diabetic patients treated with pioglitazone. 4th Interim analysis (8-year) report with data from January 1, 1997 to December 31, 2010. Unpublished data, 30 May 2012. http://www.encepp.eu/encepp/openAttachment.htm;jsessionid=ac10290dd114e55b0b300e44bfeba88c399468fa486.rlmOahaInljyqA4IpR9BcxaNb3i?field=documents.otherDocument%5B0%5D&id=2903

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

Nielsen, L H et al. Using prescription registries to define continuous drug use: how to fill gaps between prescriptions. *Pharmacoepidemiology and Drug Safety.* 2008, 17: 384-388.

Nielsen, L H et al. Estimating the effect of current, previous and never use of drugs in studies based on prescription registries. *Pharmacoepidemiology and Drug Safety.* 2009, 18: 147-153.

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.* 2005, 48: 166-174.

Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. *Acta Oncologica.* 1994, 33: 365-369.

van Herk-Sukel MP, Lemmens VE, Poll-Franse LV, Herings RM, Coebergh JW. Record linkage for pharmacoepidemiological studies in cancer patients. *Pharmacoepidemiol Drug Saf*. 2012;21(1):94-103. Epub 2011/08/04.

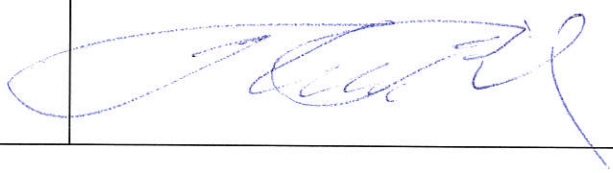
APPROVALS

We have reviewed this study protocol (Version 2.0, dated 20 June 2013) and agree to its terms by signing it.

On behalf of EPID Research Oy		
Principal investigator	**Signature**	**Date**
Pasi Korhonen, Adjunct Professor of Biostatistics		25 June 2013
On behalf of Pharmo Institute for Drug Outcomes		
Principal investigator	**Signature**	**Date**
Edith M. Heintjes Scientific Research Manager Pharmo Institute for Drug Outcomes Research	Heintjes	25 June 2013
On behalf of Clinical Practice Research Datalink		
Principal investigator	**Signature**	**Date**
Rachael Boggon, Research Statistician Clinical Practice Research Datalink		25 JUNE 2013
On behalf of Karolinska Institute		
Principal investigator	**Signature**	**Date**
Helle Kieler, Associate Professor Head, Centre for Pharmacoepidemiology		

APPROVALS

We have reviewed this study protocol (Version 2.0, dated 20 June 2013) and agree to its terms by signing it.

On behalf of EPID Research Oy		
Principal investigator	**Signature**	**Date**
Pasi Korhonen, Adjunct Professor of Biostatistics		
On behalf of Pharmo Institute for Drug Outcomes		
Principal investigator	**Signature**	**Date**
Edith M. Heintjes Scientific Research Manager Pharmo Institute for Drug Outcomes Research		
On behalf of Clinical Practice Research Datalink		
Principal investigator	**Signature**	**Date**
Rachael Boggon, Research Statistician Clinical Practice Research Datalink		
On behalf of Karolinska Institute		
Principal investigator	**Signature**	**Date**
Helle Kieler, Associate Professor Head, Centre for Pharmacoepidemiology		June 25 2013