SAFEGUARD
SAFETY EVALUATION OF ADVERSE REACTIONS IN DIABETES

Safety Evaluation of Adverse Reactions In Diabetes

HEALTH-282521

Common Study Protocol for Drug Utilisation Studies

WP4 – Observational Studies

Final
V 1.7

Document Information

Grant Agreement Number	Health- 282521	**Acronym**	SAFEGUARD
Full title	Safety Evaluation of Adverse Reactions in Diabetes		
Project URL	http://safeguard-diabetes.org/		
EU Project officer	Iiro Eerola (Iiro.EEROLA@ec.europa.eu)		

Deliverable	Common Study Protocol for Observational Studies – Drug Utilisation studies			
Work package	**Number**	4	**Title**	WP4: Drug Utilisation Studies

Status	Version 1.7	final ☐
Dissemination Level	Public ☐ Confidential ☑	

Document History

Name	Date	Version	Description
C. de Vries	30-05-2012	V1.1	First draft for internal revision.
I. Leal, S. Romio, C de Vries	05-06-2012	V1.2	Review
I. Leal, S. Romio	08-06-2012	V1.3	Integration of comments
M. Sturkenboom, S. Romio	12-06-2012	V1.4	Review and integration of comments
C de Vries	30-06-2012	V1.5	Review and integration of comments
C Sammon	30/10/2012	V1.6	Added sections 4 and 5
C Sammon	29/07/2013	V1.7	Updated lists in ANNEX I and III

Definitions

- Partners (also named as beneficiaries) of the SAFEGUARD Consortium are referred to herein according to the following codes:

  EMC - Erasmus Universitair Medisch Centrum Rotterdam (Netherlands) - Coordinator
  SYNAPSE - Synapse Research Management Partners S.L. (Spain) - Beneficiary
  PHARMO - PHARMO Coöperatie UA (Netherlands) - Beneficiary
  F-SIMG - Fondazione Scientifica SIMG-ONLUS (Italy) - Beneficiary
  UBATH - University of Bath (UK) - Beneficiary
  AEMPS - Agencia Española de Medicamentos y Productos Sanitarios (Spain) - Beneficiary
  CMNS - Consorzio Mario Negri Sud (Italy) - Beneficiary
  DSRU - Drug Safety Research Trust (UK) - Beneficiary
  CUNI - Univerzita Karlova v Praze (Czech Republic) - Beneficiary
  VUA - Vereniging Voor Christelijk Hoger Onderwijs Wetenschappelijk Onderzoek en Patientenzorg (Netherlands) - Beneficiary
  BWH - The Brigham and Women's Hospital, Harvard Medical School (US) - Beneficiary
  UNIMIB - University of Milano-Bicocca (Italy) - Beneficiary
  Uni-HB - Universitaet Bremen (Germany) - Beneficiary
  RTI-HS - RTI Health Solutions (US) - Beneficiary

- **Grant Agreement**: The agreement signed between the beneficiaries and the European Commission for the undertaking of the SAFEGUARD project (HEALTH-282521).
- **Project**: The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan**: Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium**: The SAFEGUARD Consortium, conformed by the above-mentioned legal entities.
- **Consortium Agreement**: Agreement concluded amongst SAFEGUARD participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.
- **Deliverable review**: An evaluation procedure by one or more reviewers, which precedes the distribution of a deliverable (as defined in the work plan) to the European Commission.
- **Quality assurance**: All the planned and systematic activities implemented to provide adequate confidence that an entity will fulfil requirements for quality.
- **Quality policy**: A set of principles on which quality assurance procedures are based.
- **Risk**: Uncertainty that may have a significant impact on the execution or outcome of the project, and which effect may be negative – a *threat* risk - or positive – an *opportunity* risk.
- **Foreground:** Means the results, including information, whether or not they can be protected, which are generated by activities in the Project. Such results include rights related to copyright; design rights; patent rights; plant variety rights or similar forms of protection.
- **Background:** Means information which is held by participants prior to their accession to the Grant Agreement, as well as copyrights or other intellectual property rights pertaining to such information, the application for which has been filed before their accession to the Grant Agreement, and which is needed for carrying out the indirect action or for using the results of the indirect action.

Abbreviations

The following abbreviations are used in this report:

- **ATC** – Anatomical therapeutic chemical classification system
- **BMI** – Body Mass Index
- **DDD** – Defined Daily Dose
- **DDP**-4 – Dipeptidyl peptidase 4
- **DM** – Diabetes Mellitus
- **DUS** – Drug Utilisation Study
- **EU –** European
- **GP** – General Practitioner
- **ICD-9-CM** – International Classification of Disease, 9th rev., Clinical Modification
- **ICD-10-GM** – International Classification of Disease, 10th rev., German Modification
- **ICPC** – International Classification of Primary Care
- **IPCI** – Integrated Primary Care Information
- **IV** - Intravenous
- **AMI** – Acute Myocardial Infarction
- **NIBGLD**: Non-insulin blood glucose lowering drugs
- **OR** – Odds ratio
- **OTC** – over-the-counter medication
- **RX** – prescription
- **SCD –** Sudden Cardiac Death
- **SUD** – Sudden Unexpected Death
- **T2DM** – Type 2 Diabetes Mellitus
- **TM –** Total Mortality
- **UK** – United Kingdom
- **VA –** Ventricular Arrhythmia
- **VF** – Ventricular Fibrillation
- **VT –** Ventricular Tachycardia
- **WHO** – World Health Organization

1. INTRODUCTION

The health outcomes evaluated in Work Package 4 (WP4) of SAFEGUARD are all, directly or indirectly, associated with presence of type II diabetes mellitus (T2DM) and the extent of glycaemic control. For instance, the risk of myocardial infarction (MI) is increased in people with diabetes and this increase in risk is higher when diabetes has been present for longer. The risk may be modified by improving glycaemic control. The increased risk of MI is thought to be a direct consequence of chronic hyperglycaemia although it is acknowledged that lifestyle factors predisposing people to T2DM also contribute to the increase in MI risk[1-4]. A more diffuse example is bladder cancer, the risk of which is also thought to be increased in people with T2DM although it is debated whether this increase in risk is a direct consequence of chronic hyperglycaemia, or whether it is an indirect consequence, e.g. of associated lifestyle factors, underlying genetic make-up, or of diabetes treatment [5, 6].

In clinical practice, medicine prescribing to individual T2DM patients is largely guided by the level of glycaemic control obtained. Usually, the advice is to attempt to achieve glycaemic control through dietary changes and physical activity in the first instance. When that fails, sulfonylureas or biguanides are recommended, with ADA and EASD guidelines in 2012 expressing a preference for metformin as the optimal first-line agent unless contraindicated. When these no longer achieve sufficient glycaemic control they are followed by acarbose, meglitinides or prandial glucose regulators, TZDs, incretin mimetics, DPP4 inhibitors and insulin [7]. As a patient with T2DM ages, glycaemic control tends to deteriorate, resulting in additions or changes to the prescription medicine regimen. Generally, those on sulphonylureas may be considered to be in the early stages of disease whereas T2DM patients who receive insulin are likely to have been suffering from the disease for longer. However, to a large extent this also depends on patient and prescriber preferences: on the one hand, the fact that insulin is available as an injectable medicine only is reason for some patients to resist changing to insulin treatment despite poor glycaemic control, and on the other hand, some prescribers believe in switching to insulin treatment relatively early on in the disease process in an attempt to achieve better cardiovascular outcomes in the long term.

Therefore, when evaluating whether the risk of the health outcomes captured in WP4 of SAFEGUARD differs between products, confounding by lifestyle factors, duration of T2DM, comorbidities and patient and prescriber preferences is expected to have a substantial impact on the risk estimates. This has been made explicit in the guidelines published in 2012, in which it is proposed that stringency of treatment and HbA1c values strived for depends on a range of patient characteristics such as the likelihood of adherence to treatment, the risk associated with hypoglycaemia, life expectancy, important comorbidities, and established cardiovascular disease [7]. This drug utilisation study aims to understand the extent of this potential confounding by evaluating utilisation patterns for the different antidiabetic medicines is potential. In addition, the results will provide insight into whether the confounding can be addressed through logistic regression, whether the use of instrumental variables would be appropriate, or whether confounding needs to be handled by creating propensity or disease risk scores. This drug utilisation study aims to evaluate patterns of use of the different medications to treat T2DM. In addition, the utilisation study will provide insight into any changes herein over calendar time. The basis of this protocol originated from the SOS project and was modified for the purpose of the SAFEGUARD project, with approval from the initial authors.

2. OBJECTIVES

Antidiabetic medicine prescribing will be identified and analysed on three levels:

Population level

1. Annual and monthly NIBGLD, insulins and insulin analogs user rates;
2. The impact of the choice of minimal required duration for a treatment-naïve period on the ability to identify incident users reliably;
3. The total volume of use (prescription rates per annum, person-days of exposure) by age, sex and type of diabetes treatment. A subgroup analysis will focus specifically on the use of diabetes medicines in young adults with T2DM (18-40 years old). Changes in total volume of use over calendar time will be evaluated.

Person level

1. Volume of use per user (for each type of T2DM medicine, the number of prescriptions, the number of prescriptions per annum, and duration of use);
2. Patterns of switching between products (switching between products in the same drug class as well as, how long does it take for a person to switch 'up' to the next level of T2DM treatment for better glycaemic control – and which patient characteristics are the strongest determinants for switching);
3. Persistence (duration of continuous use);
4. Concomitant drug use (overlapping days of T2DM medicine use, including insulin if applicable, in the context of total days of T2DM medicine use).

Prescription level

1. Patient determinants at each treatment initiation, channelling and change in channelling over time;
2. Gaps between prescriptions;
3. Durations of prescriptions and differences between different methods of calculation.

3. METHODOLOGY

3.1. Study design

The drug utilisation study is a descriptive retrospective observational study using routinely collected automated healthcare data.

3.2. Source data

The underlying population will consist of all eligible persons retrieved from the participating databases:

3.2.1. IPCI

Database description

In 1992 the Integrated Primary Care Information Project (IPCI) was started by the Department of Medical Informatics of the Erasmus University Medical School. IPCI is a longitudinal observational database that contains data from computer-based patient records of a selected group of general practitioners (GPs) throughout the Netherlands, who voluntarily chose to supply data to the database. GPs receive a minimal reimbursement for their data and completely 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 according to age and gender.

The database contains information on about 1.2 million patients. This is the cumulative amount of patients who have ever been part of the dynamic cohort of patients who have been 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.

The system complies with European Union guidelines on the use of medical data for medical research and has been validated for pharmaco-epidemiological research. Approval for this study will be obtained from the 'Raad van Toezicht' an IPCI specific ethical review board.

Database updates and data time lag

The database is updated continuously, every 3 months a data draw down is made for research purposes.

Data subsets and variables

The database contains identification information (age, sex, patient identification, GP registration information), notes, prescriptions, physician-linked indications for therapy, physical findings, and laboratory values (e.g. potassium, creatinine).

The International Classification of Primary Care (ICPC) is the coding system for patient complaints and diagnoses, 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. 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 (ATC) classification scheme recommended by the WHO.

Limitations of the database

Limitations of the databases are that a lot of information is available in narratives, especially information from specialists and symptoms. Also specialist medications are not complete if the GP does not enter them. It is known, however, that this proportion is minor.

3.2.2. PHARMO Database

Database description

The PHARMO medical record linkage system is a population-based patient-centric data tracking system that includes high quality and complete information of patient demographics, drug dispensings, hospital morbidity, clinical laboratory, and date of death of 3.2 million community-dwelling inhabitants of 65 municipal areas in the Netherlands. The drug dispensings originate from out-patient-pharmacies. This core dispensing database is linked on a patient level with different databases, among which the hospital morbidity data The Dutch National Medical Register (LMR). This register comprises all hospital admissions in the Netherlands, i.e., admissions for more than 24 hrs and admissions for less than 24 hours for which a bed is required. Only hospital admissions for the out-patient-pharmacy patients are collected in the PHARMO database. Clinical laboratory tests are available for a subset of the out-patient-pharmacy patients in a completely computerized format. Dates of death are available from the Central Bureau of Genealogy (CBG). The CBG is the Dutch information and documentation

centre for genealogy, family history and related sciences. Data are collected since October 1994 and include mortality. The CBG returns date of death for the out-patient-pharmacy patients. The linkage method used for individuals of the separate databases is probabilistic.

Database updates and data time lag

The linked databases in the PHARMO database network are updated every year. Databases are linked when the hospital admission data of the preceding calendar year become available; the updated database becomes available in the second half of the year. In between the outpatient pharmacy data is updated every month. Dates of death returned from the CBG have a lag time of 2 years.

The PHARMO database network covers the period 1998-2010.

Data subsets and variables

The PHARMO databases contain the following information:

- **Socio-demographic data:**
  Unique anonymous person identification number
  Gender
  Birth year
  Last known ZIP-code
  Date first contact
  Date last contact
  Reason last contact
  Date of death

- **Outpatient dispensing drug data:**
  Unique person identification number
  Unique pharmacy identification number
  Type prescriber (GP, specialist)
  ATC
  Molecule name
  Dispensed quantity (number of units)
  Type of unit (fluid, tablets etc.)
  Dispensation date
  DDD (number of DDD in one unit)
  Duration of dispensing
  Number of units to take each day (free text in Dutch)
  Strength of one unit

- **Hospital data:**
  Unique person identification number
  Unique hospital identification number
  Main diagnoses are coded in ICD9-CM
  Main diagnostic/surgical procedure
  Side diagnoses

Dates of hospital admission and discharge
Type of care (day/clinical)

Limitations of the database

- Date of first entry, last entry in the population might be subject to misclassification.
- Linkage is highly sensitive and specific but does not exclude a small percentage of linkages as misclassified
- CBG data have a lag time of 2 years compared to 1 year or less for the other sources of data. Clinical lab tests are only available for a subset of the PHARMO database

3.2.3. Health Search Database/CSD Longitudinal Patient (HSD)

Database description

The Health Search/Longitudinal Patients Database (HSD) is a longitudinal observational database that is representative of the general Italian population. It was established in 1998 by the Italian College of General Practitioners. The HSD contains data from computer-based patient records from a select group of GPs (covering a total of 1.5 million patients) located throughout Italy who voluntarily agreed to collect data for the database and attend specified training courses. 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. The HSD complies with European Union, guidelines on the use of medical data for research. The HSD has been the data source for a number of peer-reviewed publications on the prevalence of disease conditions, drug safety and prescription patterns in Italian primary care. Approval for use of data is obtained from the Italian College of Primary Care Physicians. Data are in house, no ethical approval needed.

Data subset and variables:

The database includes information on the age, gender, and identification of the patient, and GP registration information, which is linked to prescription information, clinical events and diagnoses, free text patients diary, hospital admission, and death. All diagnoses are coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). Drug names are coded according to the ATC classification system. To be included in the study, GPs must have provided data for at least 1 year and meet standard quality criteria pertaining to: levels of coding, prevalence of well-known diseases, and mortality rates. At the time in which this study will initiate, 700 GPs homogenously distributed across all Italian areas, covering a patient population of around million patients, reached the standard quality criteria.

Database updates and data time lag:

The database is updated continuously, every 6 months a data draw down is made for research purposes.

Limitations of the database:

The main limitation is the difficulty to provide additional information from GPs since in such a case an ethical approval from all the local health authorities of the respective GP practice is needed.
Medication not reimbursed from the NHS are incomplete, as well as those prescribed by the specialists. Symptoms and diagnostic instrumental results are in free text form and are not necessarily complete.

3.2.4. Regional Database Puglia

Database description
The regional databases of Puglia include hospital discharge records, prescription databases, and the civil registry, for the period 2002-2009. Shortly data on 2010 will be available. In addition the archive of physicians for 2005 is available.

Data subsets and variables
Prescription databases (last update: January 2009)
Prescription databases provide data on the community prescriptions reimbursed by the NHS with information on type and quantity of dispensed drug (generic and brand names) and dispensing date with drugs coded according to ATC classification system. This database provides the following information for each reimbursed prescription:

- Number of prescription
- Date of prescription
- Date of dispensing
- Identification number of the dispensed product (MINSAN)
- ATC code
- Number of packages
- Cost of prescription
- Pharmacy code
- City code
- First name and surname of the patient
- Date of birth of the patient
- Gender of the patient
- Fiscal or sanitary code (Tax code [alphanumerical code for personal identification, SSN cultural equivalent])
- Prescriber identification number

Hospital discharge records (last update: January 2009)
Hospital discharge records include information on primary diagnoses and up to five co-existing conditions, performed procedures (diagnostic and therapeutic interventions), date of hospital admission and discharge, and in-hospital death. All diagnoses are coded according to the ICD-9 CM.
In particular, this database provides the following information for each hospital admission:

- Hospital
- Hospital code
- Hospital discharge record
- Family name
- First name
- Date of birth
- Gender
- Place of birth (code)
- Citizenship
- Residence Region (code)
- Residence Local Health Authority (LHA code)
- Residence city (code)

- Fiscal or sanitary code (Tax code (alphanumerical code for personal identification, SSN cultural equivalent))
- Marital Status (Single; married; legally separated; divorced; widowed; not declared)
- School education (Primary/None; Middle school; Secondary school; University degree)
- Type of admission (Ordinary Hospitalization; Day hospital)
- Type of ordinary hospitalization (Scheduled not urgent; Urgent; Involuntary psychiatric treatment; Scheduled with pre-hospitalization)
- Birth weight
- Reason for day hospital (Diagnostic procedure; Day Surgery; Therapeutic procedure; Rehabilitation)
- Number of days in day-hospital
- Admission date
- Admission Unit
- Origin of the patient (GP; direct access; other region; other private institute; other public institute…)
- Traumatisms or poisoning (Industrial accident; domestic injury; car accident; suffer acts of violence; self-injury; other)
- Discharge date
- Discharge Unit
- Discharge modality (Death; ordinary; voluntary; transfer to other structure…)
- Transfer Unit and date
- ICD-9 underlying (main) discharge diagnosis code
- Other ICD-9 underlying discharge diagnosis codes
- Principal Procedure Code and Date
- Other Procedure Codes and Dates
- DRG

Civil registry (last update: January 2011)
Population registry with patients' demographics information, as gender, date of birth, fiscal code, as subjects' identifiers of all patients of LHA.

Limitations:

- No information on: Race/Ethnicity; Laboratory values; Dosing regimen; Symptoms
- It is not possible to distinguish between type 1 and type 2 diabetes mellitus
- It is not possible to assess the burden of diabetes in terms of mortality, since regional death registries are not uniformly available in all areas and are updated with a substantial delay.

3.2.5. Regional Database Lombardy (SISR)

Database description
In the Sistema Informativo Sanitario Regionale (SISR) database, data are obtained from the electronic healthcare databases of the Lombardia region. Lombardia is the largest Italian region with about nine million inhabitants, about 16% of the population of Italy. This population is entirely covered by a system of electronically linkable databases containing information on health services reimbursable by the National Health Service, including hospital admission and outpatient prescriptions of drugs free of charge.
The SISR database has a full population coverage (i.e. the population covered is not selected by any criteria) and the available information is related to drug prescriptions and to hospital admissions for the period 2000 -2010.

Database updates and data time lag
The SISR database is updated yearly. However, the possibility of access to new data is not guaranteed.

Data subsets and variables
The SISR database contains the following information:

Patient register:

- Patient ID: unique person identification number used for record linkage
- Sex
- Birth date
- Date of transferring out: the date on which a person leaves the database
- Cause of transferring out: the cause of exit from the database (death or migration)

Prescription**:** Contains all outpatients prescriptions of drugs reimbursable by the NHS

- Patient ID: unique person identification number used for record linkage
- ATC code of the drug
- AIC (Marketing Authorization): Unique code, released by AIFA (Italian Drug Agency), used to identify each box of each drug in commerce
- Prescription Date
- Quantity: number of prescribed boxes
- Using the AIC code it is possible to link drug prescriptions to a drug register which contains information on the commercial name of the drug, the quantity of active principle of the drug contained in one box, defined daily doses (DDDs) of the active principle, and the estimated coverage of one box.

Hospitalization: Contains all hospitalisations occurring in the public and private hospitals in Lombardy

- Patient ID: unique person identification number used for record linkage
- ICD-9-CM codes for diagnoses: there are 6 fields (one for the main diagnosis and 5 for the secondary diagnoses) containing ICD-9 codes
- Diagnostic procedures/surgery code: there are 6 fields (each field corresponds to a different procedure)
- Hospitalization Date: date of hospital admission
- Discharge Date: date of discharge from the hospital
- Procedures Date: there are 6 fields containing the date of the associated procedures

Limitations of the database

- The DB does not contain information on over-the-counter (OTC) medication.
- The DB does not contain information on outpatient care.
- The DB does not contain information on anthropometric measures and lifestyle (such as weight, being a smoker).
- DDDs and the amount of drug prescribed are available in the database, but there is no information on the prescribed dose.
- It is not possible to inspect hospital medical charts for validation through the regional database.

3.2.6. CPRD

CPRDThe Clinical Practice Research Datalink (CPRDCPRD) contains anonymised data from general practice for between 4 and 7% (depending on calendar year) of the UK population. The calendar period covered is 1987 – current although data from before ~2000 can only be reliably used for purposes of incidence & prevalence calculation if they are combined with the locked CPRDCPRD data (data lock April 2002). This is a consequence of the changeover of the GP practice software from a DOS-based (VM6) to a Windows-based (Vision) program and the opportunity at the time to erase historic data for patients no longer registered with the practice. As a result, the numerator & denominator data for the 1990s are unreliable for the calculation of incidence and prevalence rates.

Data are updated approximately bimonthly although this varies by practice. There are no data subsets as such. All anonymised data recorded by the GP deemed relevant for clinical management of the patient are available to researchers and comprises of demographic data including the dates the patient was registered with the practice, diagnoses and symptoms, prescriptions, hospital referral and discharge data as well as some information on major procedures carried out in hospital. For outpatient visits, hospital consultants are not allowed to issue prescriptions and hence an important proportion of prescribing initiated in hospital is captured in the CPRDCPRD (because hospital consultants will ask the GPs to prescribe the medication needed). Covariate information is available therefore on body mass index, height, weight, alcohol intake, smoking status, age, sex, socio-economic status, and comorbidity. No or extremely limited information is available on ethnicity, diet, physical activity level.

Limitations of the database include incompleteness of information on the indication for prescribing, on lab test results, and family history, as well as virtually complete lack of data on ethnicity, diet, physical activity levels, environmental and occupational exposures, medicines received and procedures undergone in hospital, and non-compliance. In addition, for data confidentiality reasons researchers do not have routine access to the free text fields; this is available at an additional cost of £0.05 per word plus £1000 administrative fee.

3.2.7. BIFAP (AEMPS)

Database description

BIFAP (Base de Datos para la Investigación Farmacoepidemiologica en Atención Primaria) database is a longitudinal population-based database of anonym zed computer based medical records of general practitioners (GPs) throughout Spain (Salvador-Rosa A, 2002). BIFAP is a non-profit research project, kept by the Spanish Medicines Agency (AEMPS), a public agency belonging to the Spanish Department of Health. The project started in 2003, including anonymized information from 2001 onwards, and the database covers data from approximately 1,260 GPs from 9 different autonomous communities in Spain. From those, 1,045 are GPs and 215 pediatricians. The database captures data on 3,948,464 patients corresponding to 17,735,987 person-years of follow-up.

In the Spanish health care system, patients are registered with a single GP who acts as a gatekeeper for and receiver of information from primary and secondary care.

The dataset is comparable with the Spanish population with respect to its age and sex distribution. Downloads are made periodically and the information is sent to the gatekeeper who de-identifies all information before further access is provided.

The GPs' electronic medical records contain coded and anonymous data on patient demographics, prescription details, clinical events, specialist referrals, laboratory test results. The International Classification of Primary Care (ICPC-1) is the coding system for patient complaints and diagnoses, although this information can also be

entered as free text[43]. Prescription data information in BIFAP includes product name, quantity dispensed, dosage regimens, strength and indication[44]. Prescriptions are coded according to the Anatomical Therapeutic Chemical (ATC) classification scheme recommended by the WHO[45]. The system complies with European Union guidelines on the use of medical research and has been proven valid for pharmaco-epidemiological research[46].

Database updates and data time lag

The out-patient-pharmacy database is updated every year with a time lag of approximately 3 months and covers the period 2001-2009.

From 2012 (for SAFEGUARD), BIFAP database will contribute with data up to the end of 2011.

Data subsets and variables

The BIFAP database contains the following information:

- **Socio-demographic data:**
  Unique anonymous person identification number
  Date of birth (dd/mm/yyyy)
  Gender
  Geographic Region (include information of GPs from 9 out of 17 Regions in Spain)
  Weight
  Height
  Smoking status
  Prescribing physician location/practice
  Prescribing physician code
  Registration status
  Start data in database system
  End date in database system
  Reason for end date recorded
  Date of death; cause of death not registered
  Start date with a specific practice
  Transfer/end date with a specific practice
  Start date for practice in system
  End date for practice in system

- **Outpatient dispensing drug data:**
  Unique anonymous person identification number
  Prescriber (GP, paediatrician)
  Drug prescribed (unique product name)
  Drug coding system (ATC codes)
  Therapeutic class
  Prescription date
  Prescribed dosage and quantity (number of units)
  Type of unit (fluid, tablets etc.)
  Strength of one unit

Duration of prescription
Number of units to take each day

- **Outpatient visits (GP):**
  Unique anonymous person identification number
  Diagnosis coded using (ICPC-2); unlimited number of diagnosis allowed
  Free text comments attached to the diagnosis.
  Referral to specialist and reason
  Tests ordered by GPs
  Name and date of tests ordered

Limitations of the database

- The dispensing of drugs is based on the prescriptions registered by the GPs, dispensings without prescription or prescriptions of other physicians are not included. In addition, it is not possible to know if a patient did actually take a dispensed drug or not.
- There is limited information available on laboratory test results, on diagnostic test results such as X-ray, MRI, etc., on results of referral visits, on hospital admission and discharge diagnosis, and on cause of death. This information is only available if the GPs include it in the electronic medical records, either in a structured way or as free text.

Information in the BIFAP database is anonimized and no personal identification is included. Consequently, in BIFAP it is not possible to access detailed information stored in other levels of primary and secondary care, and besides this, BIFAP database can not be linked to other data sources (such as hospitalary or mortality registries, etc).

3.2.8. GePaRD

Database description

The German Pharmacological Research Database (GePaRD) consists of claims data from four German statutory health insurance (SHI) providers. It covers about 14 million insurants throughout Germany who have at any time since 2004 been enrolled in one of the four SHIs. The database population represents approximately 17% of the German population of 82 million inhabitants.

Membership in an SHI is compulsory in Germany for employees with an annual income up to approximately 47.000 €. Subjects with higher incomes can choose private health insurance providers instead of an SHI and are probably underrepresented in SHIs. However, some of these higher-income subjects are voluntary members of SHIs, most often because SHIs provide free health insurance for unemployed family members (children and spouse) whereas in private health insurance plans all family members have to be paid for. About 70 million people (85% of the German population) are SHI members, including children and insurants who are retired or unemployed and about five million voluntary members.

Three of the four SHIs contributing to the database are so called ‘Ersatzkassen’ which are more likely to insure people of middle to higher socio-economic status. The database also includes data from one ‘Allgemeine Ortskrankenkasse’, an SHI which has traditionally insurants of lower socio-economic status. Two large SHIs contributing to the database together insure more than 13 million subjects all over Germany. We therefore expect the data to be adequately representative with respect to age, sex, and region of residence.

Since German SHIs pay the costs for ambulatory physician visits, hospital stays and prescription drugs for their enrolled members, information on these health services are contained in the database.

An advantage of data from German SHIs is the stability of their membership which makes long term follow-up studies feasible. In the BIPS database membership is stable in about 75% of all subjects from 2004 to 2006. However, insurants leaving a specific SHI and entering one of the other three participating SHIs cannot be identified as the same individual (synonym error).

Database updates and data time lag

At the moment the database includes about 14.3 million subjects covering the years 2004 until 2008. Usually, the database is updated annually and data from the most recent year should be available in the autumn of the following year. After data delivery another two months for in house preparation and validation are needed before updates of the database are finalised

Data subsets and variables

The SHI database contains the following information:

- **Socio-demographic data:**
  Unique person identification number: allows longitudinal analysis and linkage between the subsets
  Family identification number: identifies members of a family who are insured together
  Year of birth
  Sex
  Region of residence
  Nationality (German/other)
  Indicators for social status
  Dates of insurance coverage (entry and exit)
  Reasons for end of coverage (including death)

- **Hospital data:**
  Unique person identification number
  Unique hospital identification number
  Hospital diagnoses are coded in ICD-10-GM (at least 4 digits). Diagnosis at admission, main diagnosis at discharge, and a variable number of accessory diagnoses are available
  Dates of hospital admission and discharge
  Reason for admission
  Reason for discharge (including death)
  Diagnostic and surgical procedures (OPS Codes)

- **Outpatient prescription drug data:**
  Unique person identification number
  Unique pharmacy identification number
  Unique physician identification number: allows identification of speciality of prescribing physician
  PZN (Pharmazentralnummer): a pharmaceutical reference identification number
  Prescribed quantity (number of packages)

Prescription date
Dispensation date

A central pharmaceutical reference database with all PZN on the German market has been built up by BIPS. It contains information on generic name, brand, manufacturer, packaging size, strength, defined daily dose (DDD), pharmaceutical formulation, and ATC code. Information from the central pharmaceutical reference database is linked to the SHI database via the PZN.

- **Outpatient medical treatment data:**
  Unique person identification number
  Unique physician identification number: allows identification of specialty of consulted physician
  Ambulatory diagnoses are coded in ICD-10-GM (at least 4 digits). These diagnoses are not linked to a definite date, but refer to a quarter, as physicians' claims are collected quarterly.
  Diagnostic certainty: coded as certain, suspected, excluded, status post
  Dates of treatment / visits
  Types of treatment / diagnostic procedures with exact date (EBM codes, developed for payment of physicians for the outpatient treatment of German SHI patients)

Limitations of the database

- Exact date of birth is not known, only birth year available.
- Database contains no information on hospital or OTC medication.
- Only prescribed quantity, not prescribed dose available for medication data.
- Exact date for outpatient diagnoses is not known, only quarter available, however ambulatory diagnostic or therapeutic procedures (EBM codes) come with exact date.
- No laboratory values are contained in the database, but ordering of lab values is contained with exact date.
- The diagnostic certainty is missing for some ambulatory diagnoses, mostly 2004.
- No information on diagnoses, treatments, and prescriptions for occupational accidents and during rehabilitation is available as they are insured by a different carrier.

3.2.9. Medicare

General description
The Caremark-Medicare linked dataset provides healthcare transaction data on community-dwelling patients 65 years and older who receive their health insurance through Medicare and have prescription drug coverage through Caremark (a pharmacy benefits management company). The Caremark portion of the dataset includes pharmacy claims for medications dispensed to people who have this form of coverage, while the Medicare portion includes data on inpatient and outpatient services (Medicare Part A and B), as well as enrollment and demographic data. For research purposes, we link the data between these data sources using multiple identifies. The date range is from 2005 through 2008.

Type of database (GP/claims)

Administrative claims from pharmacies, inpatient facilities, and outpatient offices/clinics/ancillary medical services.

Database updates and data time lag

The Caremark data is updated frequently with an approximate 2-day delay from the date of prescription dispensing until its incorporation in the data source. The Medicare data is updated less frequently with a lag between service and availability in the data source for medical services of between 1-1.5 years.

To be usable for research at BWH, there is a lag of 1.5 – 2 years to obtain the most recent data and conduct the linkage between Medicare and Caremark. Thus, the year 2009 data is anticipated to be available for research in mid-2012.

Data subsets and variables

- Prescriptions: drug name (brand and generic entity), dose, formulation, days supply, number of units dispensed, dispensing date
- Inpatient and outpatient services: medical procedures (ICD-9 procedure codes), outpatient procedures (CPT-4 codes), inpatient and outpatient diagnoses (ICD-9 diagnosis codes), acute inpatient hospitalizations, emergency room visits, skilled nursing facility stays, hospice data, durable medical equipment
- Demographics, including race and ethnicity, as well as vital status
- Payments: reimbursement for inpatient and nursing home, copayments for prescriptions

Limitations of the database

Variables such as vital signs, lab test results, body mass index, and smoking are not captured; there is no cause of death recorded in the data source; diagnosis code correspond to the reason for the service and not the findings of the service; diagnosis codes could be misapplied

3.3. Study period

The study period for the drug utilisation study will comprise all the data available in the databases. It will be from from January 1st, 1998 (the earliest data) to the last data drawn down in each DB (2010-2011).

3.4. Study population

The study population will comprise all persons in the databases during the study period who have at least 365 consecutive days of valid data.

3.5. Follow-up period

Patients' data will be right-censored at the earliest of the following dates:

- Last data draw down (database-specific but generally 2010-2011);
- Transfer out of the database / end of registration / end of membership / interruption of registration or membership (depending on type of database);
- Death.

3.6. Data collection

The following data will need to be prepared by the database owners into a common data input model, in a similar fashion as will be done for the safety studies. These will be the input variables and files for Jerboa, the software tool that was developed previously in EU-ADR (FP7-ICT-2007-215847) and that has been adapted in subsequent pan-European studies funded under FP7 or by the ECDC. Jerboa is used to extract, prepare and aggregate data from a common data input model and subsequently to calculate the proposed drug use parameters in a systematic and uniform approach. To achieve this, database owners will elaborate the data locally, run Jerboa locally, and only aggregated, completely de-identified data will be shared centrally. Jerboa is used to standardise local work across SAFEGUARD partners and to avoid differences in definitions across databases that might impact study findings. Instructions for running Jerboa will accompany each request for data. All Jerboa modules have been or will be cross-validated in SAS.

3.6.1. Patient file

The source population comprises all patients (not just those diagnosed with T2DM) in the database who meet the criterion of having at least 365 days' worth of valid data. For the source population, the following data will be extracted as input files for Jerboa:

PatiendID	Patient identifier
Birthdate	Date of birth
Sex	Sex. Canc be either M or F, for male or female respectively.
Startdate	Date from which the patient is eligible to be included in the study. This is typically the date the patient is entered into the registration system.
Enddate	Date after which the patient is no longer eligible for study inclusion, as defined earlier in the end of data follow-up.

If patients appear multiple times in the dataset with different entry and exit dates, only the first time the patient is entered will be used. This applies to claims databases. The population file will not be separated into subcategories locally; instead, this will be done centrally in Jerboa where required.

3.6.2. Prescription file: prescriptions for treating T2DM

The following information on T2DM medicine use (ATC code A10 – ANNEX I) but also all other drugs will be retrieved from the databases. Underlined fields have to be filled in to keep the prescription: if any of these fields is left empty, the prescription is deleted.

PatientID	Patient ID.
ATC	Anatomic Therapeutic Chemical code of the drug (7 characters).
Date	Date of prescription or dispensing (dispensing preferred)
Formulation	Formulation (o=oral/solid, os=oral suspension/solution, r=rectal, p=parenteral (i.m/i.v, sc), inhal = inhalation, n=nasal, d=dermal, td=-transdermal, sl = subligual, v=vaginal, ocu=ocular, dep = parental depot injection, ves = intravescial, ch = chewing gum). If missing leave empty.
Strength	Indicates the strength (amount of active substance) per single entity of the formulation. For solid formulation listed it in mg. For liquid formulations listed it in mg per ml. Convert to mg or mg/ml if required. If missing leave empty.
TotalUnits	Total number of single entities of the formulation: tablets, capsules, suppositories, number of injections. In case of an oral suspension/solution indicate the total number of ml of the Rx. If missing leave empty.
UnitsPerDay	Number of prescribed units per day (if available as primary information). In case of an oral suspension/solution indicate the number of mls per day. In case of an injection indicate the number of injections per day. If missing leave empty.

DDDtotal	Total number of DDDs of the prescription for the specific formulation and ATC. Please use local values and otherwise the WHO standard. Do not calculate. If missing leave empty.
Duration	Prescribed duration of the prescription in case it is specified. If the information above is available together with the prescribed duration, please provide all. Provide also the type of information available on duration (i.e., if it is legend duration, duration by DDD, etc). This field has to be filled in for all drugs of interest, drugs with ATC code A10. For all others, leave empty if missing.

Based on the information provided, Jerboa will calculate, through 3 different modules, the duration of each prescription: please see ANNEX II for calculations.

3.6.3. Measurements file

A wide range of risk factors that are possibly also determinants of T2DM medicine prescribing has been identified for the risk assessment studies; these covariates are listed in ANNEX III. To inform our assessment as to whether these covariates act as confounders in practice, in the utilisation study for each patient diagnosed with T2DM the covariate status will be determined as well as whether they are predictive of drug choice or any other utilisation patterns. To this end, a data file will be created that comprises the following variables:

PatientID	Patient identifier
Date	Start date when the measurement first applies. For diagnoses this would be the date of diagnosis
CovariateType	Type of measurement. This table should include all covariates (potential confounders) and risk factors that are not included in prescription tables or patient tables, such tables as smoking habits, diabetes mellitus etc.
StartOrEnd	Indicates the START or END of the measurement (optional, can be left empty)

3.7. Outcome measures

3.7.1. Population level

Prevalence of T2DM medicine use

Monthly prevalence of use, overall and by type of T2DM medication (ATC7, 5, and 4 digit), will be reported as the number of patients using the medication divided by the total number of people present in the population at that time.

Table 1. Parameters to estimate measures of prevalence of use of drugs for the treatment of T2DM

Measure	**Unit of time**	**Unit**	**Numerator**	**Denominator**	**Output Jerboa**
Prevalence of specific T2DM treatment product	Calendar month	Number / 100,000 persons / month	Number of individuals having at least one day exposure to that specific T2DM treatment product in that calendar month	Total number of person months in that month	Prevalence of use per T2DM treatment product (ATC 7 level) by calendar month and calendar year by age and sex
Prevalence of specific T2DM treatment class	Calendar month	Number / 100,000 persons / month	Number of individuals having at least one day exposure to that specific T2DM drug treatment class in that calendar month	Total number of person months in that month	Prevalence of use per T2DM treatment (ATC 5 level) by calendar month and calendar year by age and sex
Prevalence of any pharmacological T2DM treatment	Calendar month	Number / 100,000 persons / month	Number of individuals having at least one day exposure to any T2DM drug treatment in that calendar month		Prevalence of use per T2DM treatment (ATC 4 level) by calendar month and calendar year by age and sex

The output files from Jerboa will include the numerator, denominator and outcome measure by different ATC levels, 5-year age categories, sex and calendar year (see Table). These datasheets will be sent from the local sites to the remote research environment (RRE).

3.7.2. New users and incidence of T2DM medicine use

A 'new user' is defined as a patient receiving the first prescription for any pharmacological T2DM treatment during the study period without having had any prescription to treat T2DM in the 12 months leading up to that prescription. Once a patient has been started on any form of pharmacological T2DM treatment they will no longer be considered eligible for classification as a new user, even if a period of non-use ensues. The latter will be considered potential non-compliance with treatment or a return to management with diet and exercise alone.

Table 2. Parameters to estimate incidence of use of drugs for the treatment of T2DM

Measure	**Unit of time**	**Unit**	**Numerator**	**Denominator**	**Output Jerboa**
Incidence of specific T2DM treatment product initiation	Calendar month	Number of users / 100,000 person months	Number of new users in that calendar month	Total number of person months contributed in that calendar month (excluding any person time from the run-in period)	Incidence of use for each pharmacological T2DM treatment (ATC 7 level) by calendar month and calendar year stratified by age and sex
Incidence of T2DM treatment class initiation	Calendar month	Number of users / 100,000 person months	Number of new users of any drug from a T2DM treatment class in that calendar month	Total number of person months contributed in that calendar month (excluding any person time from the run-in period)	Incidence of use for each pharmacological T2DM treatment (ATC 5 level) by calendar month and calendar year stratified by age and sex
Incidence of any T2DM treatment initiation	Calendar month	Number of users / 100,000 person months	Number of new users any T2DM medicine in that calendar month	Total number of person months contributed in that calendar month (excluding any person time from the run-in period)	Incidence of use for each pharmacological T2DM treatment (ATC 4 level) by calendar month and calendar year stratified by age and sex

The length of the required naïve period can vary depending on prescribing patterns locally or associated with the drug treatment. Gaps between prescriptions will be identified and the distribution of the length of these gaps will be evaluated to determine at what point (after how many days) 75%, 80, 85%, 90% and 95% of patients have collected a new prescription for further T2DM treatment. This will inform our best estimate of a minimally required run-in period to reliably identify truly treatment-naïve users.

The output files of Jerboa that will be shared will comprise the numerator, denominator and outcome measures by different ATC levels, calendar month/year, age and sex (see Table). These data sheets will be sent from the local sites to the RRE.

3.7.3. Volume of T2DM treatment use

The volume of T2DM treatment use will be expressed as the number of prescriptions and the person-time exposed during follow-up in the study population. For all products evaluated at ATC code level ATC 7, the Jerboa file that will be sent to the RRE will comprise the number of prescriptions, the number of person months exposed, the denominator (i.e. the total number of person months contributed excluding the run-in period) and outcome measures (total number of prescriptions by calendar month and calendar year as well as the total number of person days exposed by calendar month and calendar year) for all products stratified by age, sex and calendar year.

3.8. Person level use analyses

For each individual patient we will calculate several parameters to estimate duration of use, persistence with treatment, as well as treatment discontinuation or switching over the study period. These parameters will allow for the analysis of utilisation patterns.

3.8.1. Duration of T2DM use

For new users, we will calculate how long people remain on the first type of treatment before discontinuing, receiving additional T2DM treatment, or switching to different treatment altogether. The parameters to be calculated for duration of use are given in Table 3 below:

Table 3. Parameters to estimate the duration, persistence and volume of T2DM treatment per patient

Measure	**Output Jerboa**
Volume of T2DM medicine use	Total number of prescriptions following study entry
Duration of follow-up (by DDDs)	Total number of DDDs covered by prescriptions
Duration of follow-up (by days)	Total number of days covered by prescriptions
Volume of T2DM prescriptions in first year	Total number of prescriptions in first year after study entry
Total DDDs covered by T2DM prescriptions in first year	Total number of DDDs covered by prescriptions
Duration of T2DM treatment in first year	Total number of person days covered by treatment with T2DM medicines
Persistence with T2DM treatment in first year	Total number of person days covered by treatment with T2DM medicines divided by the Total number of person days contributed by study participants' first year of person time following initiation of T2DM treatment.

The Jerboa output file will be at the level of an individual patient and comprise the following variables: ATC code (level 7) for first T2DM treatment, year of study entry, age at study entry, sex, days of follow-up after the first prescription, and all covariates as listed in ANNEX III.

3.9. Prescription level analyses

3.9.1. Characteristics of prescriptions for T2DM treatment

The following characteristics of T2DM medicine use will be assessed for each product at ATC 7 level: duration of treatment as covered by the prescription, strength, formulation, daily dose, amount prescribed / dispensed, calendar year and month of prescribing, age & sex of the patient for whom the medicine was prescribed, as well as all covariates (listed in ANNEX III) valid for that patient on the date the T2DM medicine prescription was issued / dispensed (depending on the data source).

3.9.2. Overlap of multiple T2DM treatments

In case of two or more (partly) overlapping prescriptions, either a person has switched medication or more than one product is prescribed concomitantly. For the purposes of this study, if prescribing for one product is discontinued whereas a second product is added or continued, this will be considered a treatment switch. If however, multiple products are being prescribed continuously together – not necessarily on the same day, but such that the prescriptions result in treatment overlap, concomitant use will be assumed and utilisation will be mapped accordingly. Our analysis of the distribution of treatment gaps for each product at ATC level 7 will inform our decision as to whether the gap is large enough for us to assume a product is discontinued or whether we assume the gap represents less than 100% treatment persistence.

3.9.3. Channelling

Channelling occurs when interventions and/or drugs are prescribed to subjects who have a greater risk *a priori* of an adverse event or who are perceived to have a particularly high likelihood of benefiting from that specific therapy over more conventional treatment. Channelling is a type of allocation bias and it is likely to occur for newer T2DM medicines, because they will be prescribed to people who are insufficiently responsive to their existing treatment. This may have occurred with the TZDs, DPP4 antagonists, incretin-based therapies as well as analogue insulins. We will assess the extent of such channelling in different databases, in different countries between products and over time by comparing the patient characteristics of users of different products.

3.9.4. Analyses of T2DM treatment patterns

All data elaboration will be carried out locally using Jerboa scripts. The output can be used locally but it will also be submitted to the RRE for central use in the analyses. Utilisation will be described on a population, person and prescription level and compared between databases and countries. Characteristics of T2DM medicine users will be described and compared between different products by database and by country. Predictors of early or later treatment switches or discontinuation will be identified using survival analyses in which failure will be defined as a treatment switch in the first, or discontinuation in the second case. Predictors of persistence with treatment will be identified using logistic regression, in which a cut-off for 'persistence' will be identified based on the distribution of persistence. Sensitivity analyses regarding the choice of cut-off point will be carried out. Interrupted time series will be carried out using an ARIMA(p,d,q) model. In order to take into account and evaluate the regulatory interventions such as the removal of rosiglitazone from the market or licensing of new therapies (e.g. incretin based therapies), an intervention model with a complex transfer function will be used. In the time-series analysis, identification, estimation and diagnosis are necessary to model the data pattern. Identification of the model will be performed via autocorrelation and partial autocorrelation functions. Estimation of parameters p and q will be done using standard maximum likelihood approach. Diagnosis will be done analyzing residual scores; residuals will be studied using residual plots.

3.10. Limitations

This study will use data originating from different types of health care sources. These sources each have different levels of accuracy and reliability; this will impact on comparability of the data across participating

centres. In addition, information regarding treatment dosages, duration, the amount prescribed, as well as primary and secondary non-compliance is often missing or incomplete. We will impute values where necessary and carry out sensitivity analyses to evaluate the impact of assumptions made.

4. Quality Control

The studies will be conducted according to the Good pharmacoepidemiology Guidelines (GPP) issued by the International Society of Pharmacoepidemiology [8] following the Guide on Methodological Standards in Pharmacoepidemiology from The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) [9].

a. Scientific Advisory Board

The Scientific Advisory Board (SAB), with consultative function, will be formed by independent experts external to the project, so that the expertise and knowledge necessary to assist the Steering Committee on scientific and technical grounds are gathered. There will be 2 fixed SAB members, who will be selectively complemented by other experts during the project's development depending on the issues to be discussed. The SAB will usually meet once per year.

b. Reporting and dissemination of results.

The work plan structure in SAFEGUARD has been carefully designed to cover all aspects requiring specific effort towards a successful completion, and divides activities into eight work packages (WP). Among them WP8 is in charge of the dissemination and communication of results. A "Communication Plan" has been set up for the dissemination of results of the studies conducted by the SAFEGUARD Consortium. Specifically for the drug utilisation study, the dissemination undertakings will entail primarily, though not exclusively, scientifically driven interactions that will include, at least:

- Publication of scientific papers in peer reviewed journals.
- Presentations at relevant events (congresses, meetings, workshops, etc)

c. Protocol amendments

Amendments to the study protocol will be generated as needed during the conduction of the study and will be properly documented in a new version of this document. The rationale of the amendment will also be documented.

5. Ethical, Data Privacy, and Legal Issues

The study protocol will be submitted to local Scientific and Ethical Advisory Boards according to local requirements. A study-specific central remote research environment (RRE) for secure access by consortium members will be used. Each participant database will process personal data collected in national/regional electronic health record databases using Jerboa © v2.9.21, a custom-built software written in Java ™. Due to data protection and ethical considerations, Jerboa© output files will contain only anonymised de-identifiable data. These output files will be shared in the RRE where consortium members will have a secure and restricted access and where data will be analysed. Details of the RRE will be given in D4.2.

REFERENCES

1 Kannel WB, McGee DL. Diabetes and cardiovascular disease. The Framingham study. JAMA 1979;241(19):2035-2038.

2 Fuller JH, Shipley MJ, Rose G, Jarrett RJ, Keen H. Mortality from coronary heart disease and stroke in relation to degree of glycaemia: the Whitehall study. British Medical Journal Clinical Research Ed 1983;287(6396):867-870.

3 Asia Pacific Cohort Studies Collaboration. The effects of diabetes on the risks of major cardiovascular diseases and death in the Asia-Pacific region. Diabetes Care 2003;26(2):360-366.

4 Stamler J, Vaccaro O, Neaton JD, Wentworth D (1993) Diabetes, other risk factors, and 12-year cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. Diabetes Care 16(2):434-444.

5 Giovannuci E, Harlan DM, Archer MC, et al. Diabetes and cancer: a consensus report. Ca Cancer J Clin 2010;60:207-221.

6 Currie CJ, Poole CD, Gale EA. The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. Diabetologia 2009;52:1766-77.

7 Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, Peters AL, Tsapas A, Wender R, Matthews DR. Management of hyperglycaemia in type 2 diabetes: a patient-centered approach. Position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2012 Jun;55(6):1577-96. Epub 2012 Apr 20.

ANNEX I: Non-insulin blood glucose lowering drugs (NIBGLD) and insulins

Substance	ATC code
Biguanides (A10BA)	
Metformin	A10BA02
Phenformin	A10BA01
Buformin	A10BA03
Sulfonamides, urea derivates (A10BB)	
Glibenclamide (Glyburide - USA)	A10BB01
Chlorpropamide	A10BB02
Tolbutamide	A10BB03
Tolazamide	A10BB05
Carbutamide	A10BB06
Gliclazide (not marketed in the USA)	A10BB09
Glimepiride	A10BB12
Glipizide	A10BB07
Acetohexamide	A10BB31
Glibornuride	A10BB04
Gliquidone	A10BB08
Metahexamide	A10BB10
Glisoxepide	A10BB11
Combinations of oral blood glucose lowering drugs (A10BD)	
Phenformin/sulfonamides	A10BD01
Metformin/sulfonamide	A10BD02
Metformin/rosiglitazone	A10BD03
Rosiglitazone/glimepiride	A10BD04
Pioglitazone/metformin	A10BD05
Pioglitazone/glimepiride	A10BD06
Sitagliptin/metformin	A10BD07
Vildagliptin/metformin	A10BD08
Pioglitazone / Alogliptin	A10BD09
Metformin / Saxagliptin	A10BD10
Metformin/Linagliptin	A10BD11
Pioglitazone/sitagliptin	A10BD12

Metformin/alogliptin	A10BD13
Alpha-glucosidase inhibitors (A10BF)	
Acarbose	A10BF01
Miglitol	A10BF02
Voglibose	A10BF03
Thizolidinediones (A10BG)	
Rosiglitazone	A10BG02
Pioglitazone	A10BG03
Troglitazone	A10BG01
Dipeptidyl peptidase 4 inhibitors (A10BH)	
Sitagliptin	A10BH01
Vildagliptin (Not in USA)	A10BH02
Saxagliptin	A10BH03
Alogliptin	A10BH04
Linagliptin	A10BH05
Sitagliptin / Simvastatin	A10BH51
Meglitinides (A10BX)	
Repaglinide	A10BX02
Nateglinide	A10BX03
GLP1 analog (A10BX)	
Exenatide	A10BX04
Liraglutide	A10BX07
Amilyn analog (A10BX)	
Pramlintide (only USA)	A10BX05
Other (A10BX)	
Benfluorex (Not in USA)	A10BX06
Mitaglinide	A10BX08
Dapagliflozin	A10BX09

Insulins and analogues for injection, fast-acting (A10AB)	
insulin (human)	A10AB01
insulin (beef)	A10AB02
insulin (pork)	A10AB03
insulin lispro	A10AB04
insulin aspart	A10AB05
insulin glulisine	A10AB06
combinations	A10AB30
Insulins and analogues for injection, intermediate-acting (A10AC)	
insulin (human)	A10AC01
insulin (beef)	A10AC02
insulin (pork)	A10AC03
insulin lispro	A10AC04
insulin aspart	A10AC30
Insulins and analogues for injection, intermediate-acting combined with fast-acting (A10AD)	
insulin (human)	A10AD01
insulin (beef)	A10AD02
insulin (pork)	A10AD03
Insulin lispro	A10AD04
insulin aspart	A10AD05
combinations	A10AD30
Insulins and analogues for injection, long-acting (A10AE)	
insulin (human)	A10AE01
insulin (beef)	A10AE02
insulin (pork)	A10AE03
insulin glargine	A10AE04
insulin detemir	A10AE05
combinations	A10AE30
Insulins and analogues for inhalation (A10AF)	
insulin (human)	A10AF01

ANNEX II: Duration and dose calculations

For abbreviations please see main document.

Duration 1 – based on regimen

Duration1 = TotUnit / UnitDay

OR (in case of missing values for TotUnit or UnitDay)

Duration 1= Duration

OR missing.

Duration 2 – assuming 1 DDD per day

Duration2 = (TotNumUnit * Strength) / DDD*

OR

Duration2 = DDDtotal

OR missing.

*This should be the DDD for that particular ATC and formulation

Duration 3 – based on the period between 2 prescription dates of the same ATC

The third duration calculation method requires 2 steps:

1. To calculate the average DDD per day between 2 Rx dates.

A = ((TotUnit * Strength) / DDD) / (RxDate2 – RxDate1)**

OR

A = (DDDtot / (RxDate2 – RxDate1))

2. To compare this to pre-specified thresholds to classify whether fulfilling criteria continuous use or as overlapping prescriptions/concomitant use.

For example: 0.7 DDD/day < A < 1.5 DDD/day
If A meets these criteria then duration3 = (RxDate2 – RxDate1)**
If A ≤ 0.7 DDD/day then not considered as continuous use, duration is then 1 DDD per day…
If A ≥1.5 DDD/day then considered concomittant use, duration is then 1 DDD per day

** In days

Dose calculations

Dose is expressed as fraction of the DDD per day.

DDD = assumed average maintenance dose per day for a drug used for its main indication in adults. It is only assigned for drugs having an ATC code and can differ between different formulations of the same ATC.

Calculation:

(Average) dose per day for single prescription= (UnitDay * Strength) / DDD

OR (slight chance the following calculation is possible)

Average dose per day for single Rx = DDDtot/Duration

OR (slight chance the following calculation is possible)

Average dose per day for single Rx = (TotUnit * Strength / DDD) / duration

ANNEX III: COVARIATES

COVARIATES
Demographics/lifestyle
Age [a]
Sex [a]
Race/ethnicity [a]
Country of origin /DB [a]
Smoking status [a]
Drug abuse [a]
Obesity/BM I [a]
Weight loss [b]
Alcohol abuse/dependence/alcohol intake [b]

Health care utilisation [a]
Number of physician visits in year prior [a]
Number of different drugs utilized in year prior (ATC-7 level) [a]
Number of Hospitalizations in year prior [a]

DM related co-variates
Year of cohort entry [a]
Duration of T2DM (from 1st Dx) [b]
Hypoglycemic events [b]
HbA1c levels [c]

Co-morbidity
Myocardial Infarction (MI) [a]
Cardiac conduction disorders (other than ventricular arrhythmia /AF) [a]
Atrial fibrilation / flutter [a]
Ventricular arrhythmia [a]
Pericardial diseases [a]
Pulmonary hypertension [a]
Cardiomyopathies [a]
Genetic arrhythmia syndromes (Long QT syndrome, Short QT syndrome, Brugada syndrome, Catecholaminergic VT) [a]
Congenital heart disease [a]
Valve disorders [a]
Ischaemic heart disease/ coronary heart disease [a]

Co-morbidity
Heart failure [a]
Peripheral artery disease [a]
Hypertension [a]
Thrombosis/embolism [a]
Coagulopaties [a]
Hypokalemia [b]
Hypomagnesemia [b]
Hypercalcaemia [b]
Hyperlipidemia [a]
Chronic kidney disease [b]
Chronic liver disease [b]
Cancer (only malignant) [a]
Severe COPD [a]
Stroke [a]
TIA [a]
Cerebral aneurysm [a]
ERCP [b]
Gallstones [b]
Kidney stones [a]
Bladder stones [a]
History of pancreatitis [a]
Metabolic and inflammatory conditions (Myocarditis, Rheumatic diseases, Endocarditis, Sarcoidosis, Amylodosis, Fabry disease, Hemocromatosis, Endocrine disorders and diabetes, End-stage renal failure, Obesity, dieting and anorexia) [a]
Gastric ulcer [a]
Head trauma [a]
Hepatits C

[a] Assessed at cohort entry [b] Assessed at index date [c] Assessed in the period 2 month prior to cohort entry till one month after cohort entry