NOVARTIS

Quantitative Safety and Epidemiology

Non-Interventional Study Protocol

SPP100A2417

A multi-database cohort study to assess the incidence rates of colorectal hyperplasia among hypertensive patients

REDACTED PROTOCOL

Authors: [redacted]

Document type:	Non-interventional study protocol
Version number:	v00
PASS Study	No
Study phase:	IV
Release date:	14-Nov-2014

Property of Novartis
Confidential
May not be used, divulged, published or otherwise disclosed
without the consent of Novartis
NI Protocol Template Version 31-Jan-2013

1 Table of contents

1 Table of contents 2
List of tables 3
List of figures 3
2 List of abbreviations 4
3 Responsible parties 5
4 Abstract 5
5 Amendments and updates 6
6 Milestones 7
7 Rationale and background 7
7.1 Purpose and rationale 7
7.2 Background 7
8 Research question and objectives 8
9 Research methods 9
9.1 Study design 9
9.2 Setting 10
9.2.1 Source population 10
9.2.2 Study population 10
9.2.3 Inclusion criteria 10
9.2.4 Exclusion criteria 11
9.2.5 Follow-up 12
9.3 Variables 12
9.3.1 Patient demographics 12
9.3.2 Information on risk factors 12
9.4 Data sources 14
9.4.1 Clinical Practice Research Datalink 16
9.4.2 Health Search/CSD Longitudinal Patient 16
9.4.3 Integrated Primary Care Information database 17
9.4.4 Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària 17
9.4.5 Data collection schedule 18
9.5 Study size 18
9.6 Data management 18
9.6.1 Identification of Unified Medical Language System® (UMLS®) concepts 19
9.6.2 Definition of data extraction algorithm 19
9.6.3 Event data extraction 20

9.6.4 Benchmarking of incidence rates of events 20
9.7 Data analysis 21
9.7.1 Primary objective 21
9.7.2 Secondary objective 22
9.8 Quality control 22
9.9 Limitations of the research methods 23
9.10 Other aspects 24
10 Protection of human subjects 25
10.1 Regulatory and ethical compliance 25
10.2 Informed consent procedures 25
11 Management and reporting of adverse events/adverse reactions 25
12 Plans of disseminating and communicating study results 25
13 References (available upon request) 25
Appendix 1 – Diagnosis codes for hypertension 29

List of tables

Table 3-1 Main responsible parties 5
Table 6-1 Milestones for study SPP100A2417 7
Table 9-1 Data collection needed for objective achievement 9
Table 9-2 Overview of databases 14
Table 9-3 Timelines for data extraction 18

List of figures

Figure 9-1 Model for data sharing and elaboration 19
Figure 9-2 Process to be followed by local data processors 20

2 List of abbreviations

BMI	Body Mass Index
CHMP	Committee for Medicinal Products for Human Use
CPRD	Clinical Practice Research Datalink
CRC	Colorectal Cancer
CRO	Contract Research Organization
DRI	Direct Renin Inhibitor
[redacted]	[redacted]
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
FAP	Familial adenomatous polyposis
FIT	Fecal Immunochemical Testing
GPP	Good Pharmacoepidemiology Practices
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HSD	Health Search/CSD Longitudinal Patient
IBS	Irritable Bowel Syndrome
ICD	International Classification of Diseases
ICPC	International Classification for Primary Care
IPCI	Integrated Primary Care Information
IR	Incidence Rate
ISPE	International Society for Pharmacoepidemiology
MedDRA	Medical Dictionary for Regulatory Activities
NIS	Non-interventional Study
PASS	Post-Authorization Safety Study
PI	Principal Investigator
RMP	Risk Management Plan
RRE	Remote Research Environment
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
T2DM	Type 2 Diabetes Mellitus
TIA	Transient Ischemic Attack
UK	United Kingdom
WHO	World Health Organization

3 Responsible parties

Table 3-1 Main responsible parties

Role	Person
Main protocol author	[redacted]
Principal investigator (PI)	[redacted]
MAH contact person	[redacted]

4 Abstract

Title

A multi-database cohort study to assess the incidence rates of colorectal hyperplasia among hypertensive patients

Version and date

Version 00 (original protocol)

29 January 2014

Name and affiliation of main author

[redacted]

Rationale and background

In experimental studies, aliskiren – an active direct renin inhibitor – approved for treatment of hypertension, showed a potential increase in risk of colorectal hyperplasia. Epidemiologic studies examining the association between hypertension and colorectal cancer show inconsistent results. Studies on the incidence rate of colorectal hyperplasia (including polyps, cysts and neoplasms) among hypertensive patients are lacking.

Research question and objectives

With this non-interventional study Novartis intends to study the frequency of colorectal polyps, cysts and neoplasms among hypertensive patients. For that reason, the objectives of the study are:

To assess the age- and sex-specific incidence rates of colorectal polyps, colorectal cysts and colorectal neoplasms among patients diagnosed with hypertension.

To assess the time from hypertension diagnosis to colorectal polyps, colorectal cysts and colorectal neoplasms among patients diagnosed with hypertension.

Study design

Multi-database dynamic cohort study, in four European primary care databases. Study period: 01 January 2000 – 31 December 2012 (or most recent version of database)

Population

Study population: All adult patients (aged 18-79 years) with an incident diagnosis of arterial hypertension and continuous enrollment in the corresponding database for at least 1 year prior to the start of follow-up.

Exclusion criteria: A prior recording of hypertension, anti-hypertensive drug use or any hypertension-induced diseases suggesting prevalent hypertension; prior history of colorectal hyperplasia or history of cancer (excluding non-melanoma skin cancer); ≥80 years of age at start of cohort entry.

Follow-up: Start of first-time recording of arterial hypertension (cohort entry) until earliest date of diagnosis of colorectal hyperplasia; end of enrollment in database; reaching age of 80 years; or end of study period (whatever comes first).

Variables

Age, sex, antihypertensive drug exposure, inflammatory bowel disease, low-dose aspirin, obesity, hereditary colon cancer syndromes, familial colon cancer syndromes, hyperlipidemia, hypertriglyceridemia, hypercholesterolemia, type 2 diabetes mellitus (T2DM), ischemic heart disease, angina pectoris, stroke/transient ischemic attack (TIA), smoking, alcohol use, body mass index (BMI), colonoscopy, irritable bowel syndrome (IBS), and cholecystectomy.

Data sources

- The Clinical Practice Research Datalink (CPRD; formerly known as General Practice Research database [GPRD]) from the United Kingdom
- The Health Search/CSD Longitudinal Patient (HSD) database from Italy
- The Integrated Primary Care Information (IPCI) database from the Netherlands
- The Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) from Spain

Study size

All eligible study participants identified from the four databases fulfilling in-/exclusion criteria

Data analysis

Incidence rates of colorectal polyps, colorectal cysts and colorectal neoplasms (benign or malignant with or without metastases; malignant and benign neoplasms will be considered separately in the analysis) will be calculated by dividing the total number of incident cases over the total number of person-years at risk within the study population for each database separately. Incidence rates will be computed with their corresponding 95% confidence intervals, and stratified by sex and age-categories (e.g. 18-44, 45-64, 65-79 years). Standardization of incidence rates will be performed using either a European reference population or by internal validation using one of the four databases as reference population to which the other three databases will be compared to.

An analysis similar to above mentioned will be carried out when including prevalent hypertension cases in the study cohort. This will allow assessing the impact of excluding or including prevalent cases on the background incidence, as patients with longer duration of hypertension are possibly more likely to develop any colorectal neoplasm.

Kaplan Meier analyses will be performed to estimate the survival time from arterial hypertension diagnosis date until one of the primary safety endpoints (colorectal polyps, cysts or neoplasms) but also for each of the endpoints separately. This analysis will be done in the study population from each database individually and also stratified by subgroups which allows exploring whether the progression of colorectal hyperplasia is different in subgroups. The one-year risk of colorectal neoplasms will also be stratified by sex and by age groups (18-44, 45-64, 65-79 years).

5 Amendments and updates

Updates to specific sections of the study protocol addressing PRAC comments are in section 9.2.4 and 9.7.1.

6 Milestones

In line with the European Union (EU) aliskiren Risk Management Plan v.11, there will be three milestones in this 15-month study. The calendar dates for each milestone will depend on the date of approval of the protocol by PRAC, and the following timelines are proposed:

Table 6-1 Milestones for study SPP100A2417

Milestone	Timelines
End of data collection	PRAC approval + 12 months
Final report of study results	PRAC approval + 15 months

There is no progress report milestone in this study, only a final report is foreseen.

Time periods considered for each milestone are based upon experience from previous studies and projects collaborating with different data sources. Before start of data-extraction, harmonization of codes and definitions is necessary and takes up to 3 months. When considering several harmonization runs of data extraction 12 months after PRAC approval data extraction is foreseen to end. Approximately 3 months are needed to analyze the data and deliver the final report.

7 Rationale and background

7.1 Purpose and rationale

In the context of the Rasilez Follow-up Measures (FUM) 025 regarding the carcinogenic potential of aliskiren (and of aliskiren in combination with hydrochlorothiazide [FUM 026]), the Committee for Medicinal Products for Human Use (CHMP) requested from Novartis to do a non-interventional study (NIS) assessing age- and sex-stratified incidence rates of colorectal hyperplasia in a hypertensive population to get background incidence rate data in this population. Based on that request, Novartis proposed to do a NIS with secondary use of data derived from various European healthcare data sources, namely the Clinical Practice Research Datalink (CPRD) from the United Kingdom (UK), the Health Search/CSD Longitudinal Patient (HSD) database from Italy, the Integrated Primary Care Information (IPCI) database from the Netherlands, and the Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) from Spain.

7.2 Background

Aliskiren (Rasilez®) is the first orally active direct renin inhibitor (DRI) approved for the treatment of hypertension. Aliskiren's inhibitory effect on angiotensin I generation, through renin blockade, is highly specific and long-lasting (24 hours) (Fisher et al 2008). Aliskiren antihypertensive effect is similar to that determined for angiotensin-converting-enzyme inhibitor, angiotensin receptor blockers and thiazides (Schmieder et al 2009, Musini et al 2011, Angeli et al 2012).

Colorectal hyperplasia is listed as a potential risk in the aliskiren risk management plan (RMP), which is based upon findings in a rodent carcinogenicity study. However, these findings have not been confirmed in a 2-year marmoset study or in targeted clinical studies including a colonoscopy study. Therefore, the increased risk of colorectal hyperplasia

observed in the rodent study may reflect high intraluminal drug concentrations in rats or could be a species-specific difference between rats and humans in response to aliskiren exposure.

Colorectal cancer (CRC) is a common type of cancer, being the third most commonly diagnosed cancer in males and the second in females. The incidence of CRC varies globally over 10-fold, mainly due to differences in dietary and environmental factors. According to data from various cancer registries worldwide, age-standardized CRC incidence rates may range from approximately 4.1 to 59.1 per 100,000 per year among males and 3.6 to 39.5 per 100,000 per year among females, respectively (Center et al 2009).

Apart from genetic susceptibility other risk factors of CRC are inflammatory bowel disease (IBD) (Eaden et al 2001, Herrinton et al 2012), alcohol use, obesity and diabetes mellitus (Ahmed et al 2006, Esposito et al 2012).

Epidemiologic studies examining a potential association of hypertension with the development of CRC have reported inconsistent findings. Various published observational studies did not find an increased risk of CRC in patients with hypertension compared to normotensive patients (Lever et al 1999, Negri et al 1999, Lindholm et al 2001, Lindgren et al 2005, Stürmer et al 2006, Kim et al 2007, Aleksandrova et al 2011, Azoulay et al 2012). However, there is also evidence from observational studies that hypertension might be associated with an increased risk of CRC (Batty et al 2003, Othman and Zin 2008, Stocks et al 2008, Pelucchi et al 2010), especially in hypertensive patients with type 2 diabetes mellitus (T2DM) and obesity (Stocks et al 2008), or with some components of the metabolic syndrome (Kim et al 2007). Additionally, there is some evidence that the mortality of CRC may be higher in hypertensive compared to normotensive patients (Batty et al 2003, Watanabe et al 2005). Again, the data on this topic are inconsistent, as a review including 10 longitudinal studies could only demonstrate an increased risk of mortality from cancer in general, but not of CRC specifically (Grossman et al 2002).

To what extent a potential increased risk of CRC during antihypertensive treatment of aliskiren may be related to the drug, is currently difficult to establish, as the underlying background risk of CRC in hypertensive patients is unknown. Specific data on the incidence rate of CRC or more general on colorectal hyperplasia in a population of hypertensive patients however are not available in the published literature. This non-interventional study is therefore planned – as requested by CHMP – to provide more details on the background incidence of colorectal hyperplasia in hypertensive patients in general. An additional non-interventional study based on US health claims data (SPP100A2418) will assess the risk of colorectal hyperplasia specifically in association with exposure to aliskiren and with other antihypertensive drugs.

8 Research question and objectives

The primary objective of this non-interventional study is to assess age- and sex-stratified incidence rates of colorectal hyperplasia among patients with diagnosed hypertension based on information from various European primary care data sources.

Secondary objective is to assess the hazard and time from hypertension diagnosis until colorectal hyperplasia. This includes estimating the one-year risk of colorectal neoplasms after hypertension diagnosis, stratified by sex.

9 Research methods

9.1 Study design

A dynamic cohort study in hypertensive patients will be performed with secondary use of data derived from multiple databases.

Primary care databases (electronic medical records) will be used allowing identification of patients with an incident diagnosis of hypertension. This is because it is likely that the first diagnosis of hypertension would occur in an outpatient primary care setting. Inpatient/hospitalization data are considered as inadequate to identify patients with incident hypertension, because it is very likely that assessment of hypertension based on e.g. a primary or secondary discharge diagnosis of hypertension would bias the patient selection to a population with either prevalent and/or more severe hypertension. The inclusion of patients with prevalent hypertension would not be reasonable in this context because the date of the start of follow-up (e.g. the date of diagnosis of hypertension) cannot be reliably assessed in these patients; inclusion of patients with prevalent hypertension, likely with more severe hypertension, may lead to an inflated colorectal hyperplasia incidence rate estimate. In addition, the follow-up time within the database should be of sufficient length to allow identifying cases with a diagnosis of colorectal hyperplasia. The databases included in the study (see Section 9.4 for details) contain information on demographics, diagnoses, and drug prescriptions. The clinical information captured by the databases is by different disease coding systems, such as the International Classification of Diseases (ICD) 9th or 10th revision, International Classification for Primary Care (ICPC) (Lamberts et al 1992), or READ (Chisholm 1990).

The primary endpoint of interest includes a diagnosis corresponding to colorectal hyperplasia defined as any of the following diagnoses and/or conditions:

- Colorectal cysts
- Colorectal polyps (with/without hyperplasia and irrespective of the type or size of polyp), and
- Colorectal neoplasms (benign or malignant with or without metastases; malignant and benign neoplasms will be considered separately in the analysis)

Part of the case definition of colorectal hyperplasia will also be: any ulcerations or colonic bleeding related or due to colorectal malignancy or cell dysplasia. Carcinoma in situ will therefore also be included.

Tumor size and tumor type (adenocarcinoma or other more rare histopathological types of cancer) will not be taken into account, neither polyp size or polyp type (such as a distinction between tubular, villous or sessile polyps).

Harmonization of corresponding disease codes of the primary outcome will be performed across coding systems (ICD-9th revision, ICD-10th revision, ICPC-, READ-coding system).

Table 9-1 Data collection needed for objective achievement

Characteristics	Objective	
	Primary	Secondary

Characteristics	Objective	
	Primary	Secondary
Demographics		
Age	Yes	Yes
Sex	Yes	Yes
Date entry database	Yes	Yes
Date exit database	Yes	Yes
Variables		
Inclusion criteria (hypertension)	Yes	Yes
Outcome (CRC, polyps, hyperplasia)	Yes	Yes
Covariates (risk factors)	Yes	Yes
Prescriptions	Yes	Yes

9.2 Setting

9.2.1 Source population

The source population consists of all subjects that are registered with general practitioners who contribute data to the participating databases between January 1st 2000 and December 31st 2012 (or the most recent version of data available).

9.2.2 Study population

The study population will be a cohort of adult patients (aged 18-80 years) in the individual databases with an incident diagnosis of arterial hypertension between January 01st 2000 and December 31st 2012 (or the most recent version of data available at the time of data extraction).

9.2.3 Inclusion criteria

Patients may enter the study cohort when they meet the following inclusion criteria:

- Continuous enrollment in the database for at least 1 year prior to the start of follow-up
- Aged 18 years or older at the start of follow-up
- Diagnosis of arterial hypertension (see Appendix 1 for codes)

Arterial hypertension will be identified by the following approaches which classify mutually exclusive hypertensive cohorts:

1. Documentation of at least 1 diagnosis code of arterial hypertension without a measurement with elevated blood pressure (recorded 4 weeks before or after diagnosis code) and without starting of antihypertensive treatment (diuretics, beta-blockers, calcium antagonists, ACE-inhibitor, angiotensin receptor blocker; on or up to four weeks after the date of the diagnosis). The date of diagnosis code recorded in the database will be used as start date.

2. Documentation of at least 1 diagnosis code of arterial hypertension **plus a measurement** with elevated blood pressure (defined as systolic blood pressure ≥ 140 mmHg and/or diastolic ≥ 90 mmHg) recorded 4 weeks before or after diagnosis code. This approach would allow capturing 'untreated' or less severe hypertensive patients. The date of diagnosis code recorded in the database will be used as start date.
3. Documentation of at least 1 diagnosis code of arterial hypertension **plus start of treatment** of any of the following antihypertensive drugs (diuretics, beta-blockers, calcium antagonists, ACE-inhibitor, angiotensin receptor blocker) on or up to four weeks after the date of the diagnosis. A four week timeframe is chosen to be more conservative and not miss any starting treatments for hypertension and take any delay in recording of the prescriptions into account. This case definition likely captures the treated hypertensive patients or the more severe hypertensive patients. The date of diagnosis code recorded in the database will be used as start date.
4. Documentation of at least 1 diagnosis code of arterial hypertension **plus a measurement** with elevated blood pressure (defined as systolic blood pressure ≥ 140 mmHg and/or diastolic ≥ 90 mmHg) recorded 4 weeks before or after diagnosis code hypertension **plus start of treatment** of any of the following antihypertensive drugs (diuretics, beta-blockers, calcium antagonists, ACE-inhibitor, angiotensin receptor blocker) on or up to four weeks after the date of the diagnosis. The date of diagnosis code recorded in the database will be used as start date.

Of note, pregnancy induced hypertension codes will not be used for definition of arterial hypertension.

9.2.4 Exclusion criteria

Patients with one or more of the following conditions will be excluded:

- A prior recording of a hypertension diagnosis or any hypertension-induced sequelae or diseases (such as hypertensive heart disease (Healey and Connolly 2003): i.e. cardiomegaly), hypertensive retinopathy, hypertensive encephalopathy or hypertensive renal disease [glomerulosclerosis or nephrosclerosis]) suggesting prevalent hypertension prior to patient eligibility or start of the study period.

Amendment: In order to assess the impact of excluding complicated hypertension persons from the cohort on the the background incidence of colorectal neoplasms, we will carry out a sensitivity analysis by including prevalent hypertension patients.

- Patients with antihypertensive drug exposure during follow-up or at cohort entry (without arterial hypertension diagnosis at or before cohort entry, e.g. use of diuretics for other indications, such as portal hypertension or pulmonary hypertension) will be excluded
- A history of colorectal malignant cancer prior to cohort entry
- A history of cancer (excluding non-melanoma skin cancer) prior to cohort entry
- ≥ 80 years at start of cohort entry

We will exclude patients at cohort entry if they are diagnosed with hypertension or any complication due to hypertension (e.g. retinopathy, encephalopathy etc.) in order to exclude prevalent hypertension cases. The sequence of hypertension to complications of hypertension may take years to decades and also due to co-morbid occurrence of other diseases, such as

obesity and type 2 diabetes mellitus (T2DM), inclusion of such patients would impact the incidence of colorectal neoplasms (overestimation). By doing so, we do not thread internal validity, but may limit the generalizability of the study.

We cannot be certain that at cohort entry patients are 'free' of any colon hyperplasia or polyps, since for this evaluation/conclusion reports from colonoscopy are required. However, since many people will not have undergone a colonoscopy we will exclude only those with documented colorectal hyperplasia, but we will falsely leave in the subjects with 'silent/unknown hyperplasia'. Subsequently, this will impact the results although the direction of the bias is unknown. Therefore, we do not exclude subjects with a history of colorectal polyps or hyperplasia.

9.2.5 Follow-up

Patients will be followed from the date of the first-time recording of arterial hypertension (cohort entry date) until the earliest date of:

- Endpoint of interest (see Section 9.1)
- A recorded diagnosis of colorectal hyperplasia or neoplasm
- The end of enrollment in the database
- Reaching the age of 80 years (see Section 9.3 for explanation)
- The end of study period (i.e. December 31, 2012)
- Last data update in the database, if more recent data is available than December 31, 2012

9.3 Variables

Data on important variables associated with the primary outcome and hypertension will be collected at **cohort entry** for the following variables (if available):

9.3.1 Patient demographics

- Age and age group (e.g. 18-44, 45-64, 65-79 years)

  Age will be censored at 80 years as those subjects are more likely to be moved to a nursing home and information on events and CRC is less likely to be captured in the primary care databases. As the underlying source population thus in this age group is uncertain, reaching age of 80 years will be considered as censoring point.
- Sex
- Antihypertensive drug exposure (use of diuretics, beta-blockers, calcium antagonists, ACE-inhibitor, angiotensin receptor blocker; classified as yes/no)

9.3.2 Information on risk factors

Information on risk factors either associated with colorectal cancer or with cardiovascular disease will be identified based on disease codes and proxy medications.

Information on the following risk factors will be identified:

- Risk factors associated with colorectal cancer:

    - Assessed in entire history before cohort entry date
        - Inflammatory bowel disease (ulcerative colitis, Crohn's disease) – risk increase
        - Hereditary colon cancer syndromes (familial adenomatous polyposis [FAP]; Lynch Syndrome/hereditary non-polyposis colorectal cancer [HNPCC]) – risk increase
        - Irritable bowel syndrome (IBS)
        - Colonoscopy
        - Familial colon cancer syndromes
        - Cholecystectomy
    - Assessed from 1 year before to 30 days after cohort entry date
        - Obesity (assessed by body mass index [BMI]; see below. Defined as BMI > 30 kg/m^2) – risk increase
        - Low dose aspirin (up to 325 mg/day of acetylsalicylic acid for cardiovascular prevention) – risk decrease
- Cardiovascular risk factors:
    - Assessed in entire history before cohort entry date
        - History of ischemic heart disease (myocardial infarction, angina pectoris, coronary artery bypass surgery) (diagnoses and drugs as proxies)
        - History of transient ischemic attack (TIA) or stroke (ischemic and hemorrhagic)
    - Assessed from 1 year before to 30 days after cohort entry date
        - Hyperlipidemia (diagnoses and drugs as proxies)
        - Hypertriglyceridemia (diagnoses and drugs as proxies)
        - Hypercholesterolemia (diagnoses and drugs as proxies)
        - T2DM (diagnoses and drugs as proxies)
- General risk factors:
    - Assessed 1 year before up to 1 year after cohort entry, most recent data will be considered
        - BMI in kg/m^2, if available. BMI will be categorized into BMI classes (<18 kg/m^2; 18- <25 kg/m^2; 25- <30 kg/m^2; 30-<35 kg/m^2 and > 35 kg/m^2)
        - Smoking, into categories (current smoking; yes/no)
        - Alcohol use (excessive/normal)

Dietary factors that are associated with an elevated or decreased risk of CRC or hypertension, such as fruit or vegetable intake, fish consumption etc. cannot be assessed from the primary care databases and will introduce substantial misclassification and therefore will not be included. Other life style factors relevant in this context, especially physical activity, cannot be assessed either in these databases.

If the disease coding system allows, hereditary causes of colorectal cancer will be identified separately and will not be excluded from the cohort. Patients diagnosed with hereditary colorectal cancer during follow-up will be analyzed as separate subgroup.

Given the high chance of Berkson bias in a hypertensive cohort, subgroup analysis within people who had a 'negative' colonoscopy will be considered. In addition, the use of antiplatelets medication may enhance the likelihood of colorectal bleeding, for which they might undergo a colonoscopy. The chance of finding any of the primary endpoints is therefore higher in a hypertensive cohort as compared to the general population.

In a population with a recorded 'negative' colonoscopy you can be certain that they were event-free at a certain point. A colonoscopy is indicated for patients with complaints of weight loss, rectal blood loss and change in bowel habits. The latter can also be present in the context of IBS, a benign disease. IBS will be considered as covariable for colonoscopy.

The cohort will be described according to the above mentioned risk factors, and additionally the use of metformin and insulin in the cohort, as a proxy for diabetes severity, will be reported.

9.4 Data sources

The following four European general practice databases will be used:

1. The Clinical Practice Research Datalink (CPRD; formerly known as General Practice Research Database [GPRD]) from the United Kingdom (Jick et al 1991, Jick et al 2003)
2. The Health Search/CSD Longitudinal Patient (HSD) database from Italy (Filippi et al 2005)
3. The Integrated Primary Care Information (IPCI) database from the Netherlands (Vlug et al 1999)
4. The Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) from Spain (Bolíbar et al 2012)

Main characteristics of the participating databases are given in Table 9-2.

Table 9-2 Overview of databases

Characteristics	**Database**			
	ICPI	CPRD	HSD	SIDIAP
Country	The Netherlands	UK	Italy	Spain
Type of database	MR	MR	MR	MR
Number of patients, **millions**	1.2	3	1.5	5.1
Date in	Yes	Yes	Yes	Yes
Date out	Yes	Yes	Yes	Yes
Date of death	Yes	Yes	Yes	Yes
Cause of death	Yes	Yes	No	No
Updates	Bi-annually	Continuously	Bi-annually: (30/06 and 31/12)	Yearly (April/May)
Prescriptions				
Outpatient Rx	Yes (specialist incomplete)	Yes (specialist incomplete)	Yes (specialist incomplete)	Yes (specialist incomplete)
Coding of drugs	ATC	BNF/Multilex code	ATC	ATC
Dosing regimen	Yes	Yes	Yes (incomplete)	Yes

Characteristics	Database			
	ICPI	CPRD	HSD	SIDIAP
Outcomes				
Hospitalizations	Yes	Yes	Yes	Yes
Outpatient diagnoses	Yes	Yes	Yes	Yes
Coding of disease	ICPC	READ	ICD-9 CM	ICD-10

ADM = Administrative; ATC = Anatomical Therapeutic Chemical; BNF = British National Formulary; ICD= International classification of disease, ICPC = International Classification of Primary Care; MR = Medical Record

More details on the individual databases are given in the subsections below. The databases have been selected based on their geographic location, the availability of population and primary care data based information plus their recognized reputation in the area of drug utilization and safety research. Multiple countries are included in order to provide international data. All of the participating databases are part of the "EU-ADR Alliance", a stable collaboration framework for running drug safety studies in a federated manner, especially when the participation of several electronic health care record databases is required (Coloma et al 2011).

All of the chosen databases comply with European Union (EU) guidelines on the use of medical data for medical research and have been validated for pharmaco-epidemiological research. Concerning the validity of a hypertension diagnosis in the participating databases; several pharmaco-epidemiological studies have been performed combining data from IPCI and HSD. This includes a study on the prevalence and treatment of hypertensive patients (Sturkenboom et al 2008). A study on incident hypertensive patients by physician diagnosis has been performed in CPRD (Burke et al 2006). For SIDIAP, there has been a recent study on the validity of cardiovascular diseases and their risk factors in SIDIAP database (Ramos et al 2012). Colorectal cancer has been studied as outcome in CPRD (García Rodríguez and Huerta-Alvarez 2001, Yang et al 2004, van Staa et al 2005, Hong et al 2013), and IPCI (van Soest et al 2008), but not yet in HSD and SIDIAP.

In some studies (García Rodríguez and Huerta-Alvarez 2001, van Soest et al 2008) extensive case validation was performed to validate the diagnostic codes for CRC. In CPRD this yielded evidence that the validity of CRC diagnoses is high: over 95% of computer-recorded incident CRC diagnoses were validated by review of original medical records or by confirmation of the diagnosis by the GP (García Rodríguez and Huerta-Alvarez 2001). In IPCI, the positive predictive value of the ICPC colorectal cancer code was 90% (van Soest et al 2008).

Concomitant or prior medications entered into the database will be coded using the World Health Organization (WHO) Drug Reference List. Medical history/current medical conditions and adverse events will be coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

Drugs are coded according to the Anatomical Chemical Therapeutic system in IPCI, HSD and SIDIAP, but according to BNF/multilex codes in CPRD.

9.4.1 Clinical Practice Research Datalink

The CPRD; from the UK collates the computerized medical records of general practitioners (GPs) in the UK who act as the gatekeepers of healthcare and maintain patients' life-long electronic health records. As such they are responsible for primary healthcare and specialist referrals, and they also store information stemming from specialist referrals, and hospitalizations. The data recorded in the CPRD include demographic information, prescription details, clinical events, preventive care, specialist referrals, hospital admissions, and major outcomes, including death. The majority of the data are coded in READ Codes (Booth 1994); however, additional text data is also available, which can improve the sensitivity and specificity of data. Validation of data with original records (specialist letters) is also available.

Importantly, CPRD operates a careful and continual quality control procedure that ensures that only practices that are "up-to-standard" (UPS) are included in the research dataset. The dataset is generalizable to the UK population based upon age, sex, socioeconomic class and national geographic coverage.

There are currently approximately 13.2 million patients (acceptable for research purposes) – of which 5.7 million are active (still alive and registered with the GP practice) – in approximately 680 practices. Data include demographics, all GP/healthcare professional consultations (phone, letter, email, in surgery, at home), diagnoses and symptoms, laboratory test results, treatments, including all prescriptions, all data referrals to other care, hospital discharge summary (date and ICD-9 code), hospital clinic summary, preventive treatment and immunizations, death (date and cause). Drug and event coding dictionaries are available upon request.

CPRD is listed under the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database. (www.encepp.eu/encepp/resourcesDatabase.jsp).

9.4.2 Health Search/CSD Longitudinal Patient

The Italian arm of the study will use the Health Search CSD Longitudinal Patient Database (HSD), a longitudinal observational database that is representative of the Italian general population. It was established in 1998 by the Italian College of General Practitioners (Filippi et al 2005). The HSD contains data from computer-based patient records from a selected group of GPs covering a total of 1.5 million patients located throughout Italy. These GPs voluntarily agreed to collect data for the database and attend specified training courses. The database includes information on the age, sex, and identification of the patient, and GP registration information, which is linked to prescription information, clinical events and diagnoses, hospital admission, and causes of death. All diagnoses are coded according to 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 (Cricelli et al 2003). The HSD complies with EU, guidelines on the use of medical data for research. The HSD has been used as data source for a number of peer-reviewed publications on the prevalence of disease conditions, drug safety and prescription patterns in Italian

primary care (Cazzola et al 2011). Approval for use of data is obtained from the Italian College of General Practitioners.

HSD is listed under the ENCePP resources database. (www.encepp.eu/encepp/resourcesDatabase.jsp).

9.4.3 Integrated Primary Care Information database

[redacted] IPCI is a longitudinal observational database that contains data from computer-based patient records of a selected group of GPs throughout The Netherlands, who voluntarily chose to supply data to the database. Collaborating practices are located throughout The Netherlands and the collaborating GPs are comparable to other GPs in the country according to age and sex. In the Netherlands, all citizens are registered with a GP practice, which forms the point of care and acts as a gatekeeper in a two-way exchange of information with secondary care. The medical records of each patient can therefore be assumed to contain all relevant medical information including medical findings and diagnosis from secondary care. The IPCI database is representative for the Dutch population regarding age and sex (Voordouw et al 2004).

The database contains information on about 1.6 million patients. This is the cumulative number of patients who have ever been part of the dynamic cohort of patients who have been registered. 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 prescribed, dosage regimens, strength and indication are entered into the computer (Vlug et al 1999). The National Database of Drugs, maintained by the Royal Dutch Association for the Advancement of Pharmacy, enables the coding of prescriptions, according to the ATC classification scheme recommended by the World Health Organization (WHO).

As this is a primary care database, information on specialist prescribing, drug dispensing and actual drug intake is missing.

IPCI is listed under the ENCePP resources database. (www.encepp.eu/encepp/resourcesDatabase.jsp).

9.4.4 Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària

GPs play an essential role in the public health care system of Spain, as they are responsible for primary health care, long-term prescriptions and specialist and hospital referrals. The Spanish public health care system covers more than 98% of the population. The SIDIAP Database comprises of electronic medical records of a representative sample of patients attended by GPs in Catalonia (North-East Spain), covering a population of more than 5.1 million patients (about 80% of the total of 7.5 million population of Catalonia) from approximately 274 primary care practices with 3,414 participating GPs. The SIDIAP data comprises the clinical and referral events registered by primary care health professionals (GPs and nurses) and administrative staff in electronic medical records, comprehensive demographic information, prescription and corresponding pharmacy invoicing data, specialist referrals, primary care laboratory test results, and hospital admissions and their major outcomes. Health professionals gather this information using ICD-10 codes, and structured forms designed for the collection

of variables relevant for primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, blood and urine test results. Only GPs who meet quality control standards can participate in the SIDIAP database. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP Database. Recent reports have shown the SIDIAP data to be useful for epidemiological research (García-Gil et al 2011).

SIDIAP is listed under the ENCePP resources database. (www.encepp.eu/encepp/resourcesDatabase.jsp).

9.4.5 Data collection schedule

This is a non-interventional study and does not impose a therapy protocol, diagnostic/therapeutic procedure, or a visit schedule. Patients are treated according to the local prescribing information, and routine medical practice in terms of visit frequency and types of assessments performed and only these data will be collected as part of the study.

Table 9-3 Timelines for data extraction

		Start	End
Start		Contract signed (CS)	
Mapping	Colorectal outcome, covariates	CS + 0-4 weeks	CS + 9 weeks
Data extraction 1	Colorectal outcome/Hypertension	CS+ 9 weeks	CS + 18 weeks
Review extractions	Colorectal cancer codes/harmonization	CS + 18 weeks	CS + 24 weeks
Data extraction 2	Covariates/outcomes	CS + 25 weeks	CS + 32 weeks
Review extractions 2	Covariates/outcomes	CS + 32 weeks	CS + 38 weeks
Data extraction 3	Covariates	CS + 39 weeks	CS + 45 weeks
Review extractions 3	Covariates	CS + 46 weeks	CS + 52 weeks
Analysis		CS + 53 weeks	CS + 62 weeks
Drafting report		CS + 62 weeks	CS + 65 weeks
Manuscript preparation		CS + 62 weeks	CS + 65 weeks

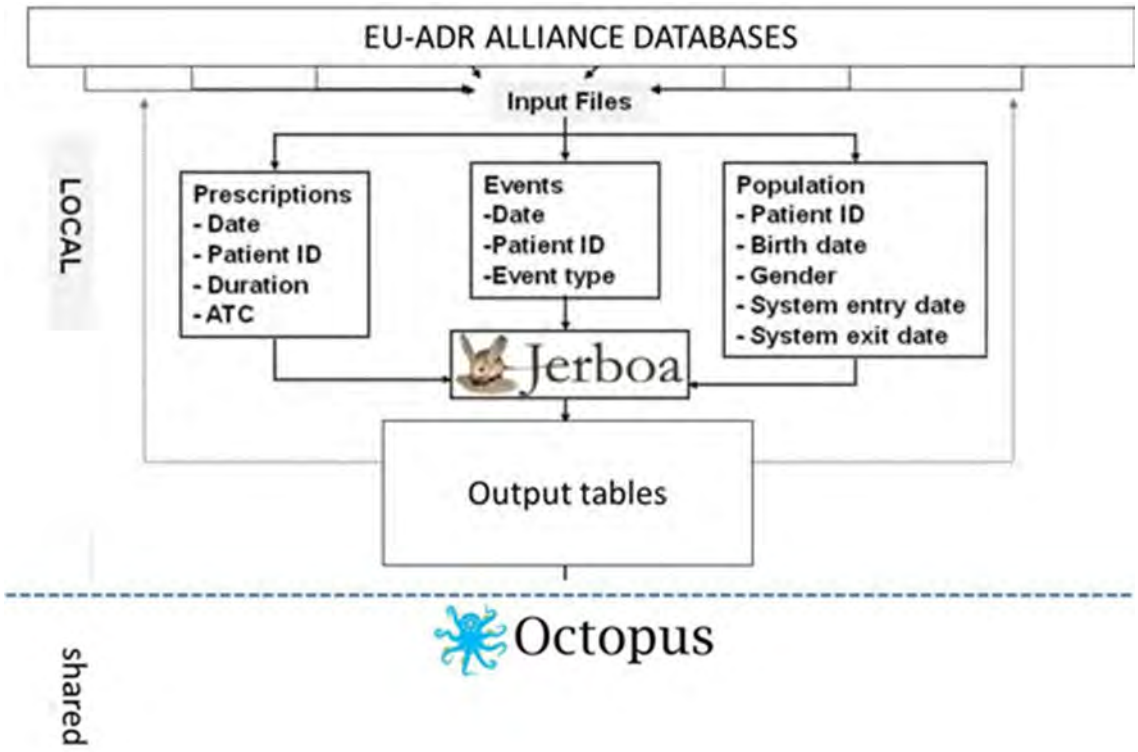
9.5 Study size

All eligible study participants identified from the four databases fulfilling in-/exclusion criteria.

9.6 Data management

The EU-ADR Alliance works in a federated manner: data extraction and elaboration is done locally and pooling of aggregated data is done on a remote research environment (see Figure 9-1 for overview). The CPRD database which is not hosted by one of the EU-ADR Alliance partners will be elaborated in a similar manner but by using Novartis access.

Figure 9-1 Model for data sharing and elaboration



Due to the different database characteristics and coding schemes it is not possible to use one single data extraction algorithm for all the databases. To reconcile differences across terminologies a shared semantic foundation will be built for the definition of events under study by selecting disease concepts from the Unified Medical Language System (UMLS, V.2008AA), and set up a multi-step and iterative process for the harmonization of event data. The sequential steps of this process are shortly described below:

9.6.1 Identification of Unified Medical Language System® (UMLS®) concepts

A UMLS concept is identified by a Concept Unique Identifier (CUI) and describes a single medical notion that can be expressed using different synonyms (terms). For each event, a medical definition will be created and, based on such definition relevant UMLS concepts are identified and projected into the database-specific terminologies. In addition, for those databases where free text is available, the labels of the codes are considered for free text search of the events.

9.6.2 Definition of data extraction algorithm

Based on the relevant diagnostic codes and key words (for free text search), a data extraction algorithm will be constructed for each event based on the consensus of the data providers. This data extraction algorithm will then be implemented by all databases. Since a cycle of harmonization of the events will be performed and thus, the algorithms will be changed and adapted accordingly, the final algorithm for data extraction cannot be provided yet. An initial list for hypertension diagnosis codes is included in Appendix 1.

9.6.3 Event data extraction

Subsequently, each database extracts data locally and transforms them into a simple common data model, i.e. standardized patient, drug and event files linkable via a patient unique identifier.

9.6.4 Benchmarking of incidence rates of events

For each endpoint and covariate database-specific incidence rates (IRs) using Jerboa© will be benchmarked, and scripts will be generated by [redacted]. The observed IRs are compared with IRs estimated from previous database studies and literature. Outliers are identified and further investigated in an iterative manner, meaning that before the final event extraction is achieved, several rounds of checking and revising the event algorithms will be performed. The purpose of this benchmarking exercise is to estimate the IRs per database separately and to understand possible differences in IR across databases. IRs will be calculated per database separately and will not be pooled per se, but if the IRs appear to be in line, pooling of data on patient-level could be considered.

This multi-step process has been used successfully in several other European multi-database projects. It maximizes the involvement of the data providers in the study by utilizing their knowledge on the characteristics and the process underlying the data collection.

After completion of harmonization, output tables for calculation and analysis of study endpoints will be created by the local data processors using the following steps (see Figure 9-2).

Figure 9-2 Process to be followed by local data processors

1. Data elaboration

A standardized Jerboa© script and instructions will be created by [redacted] to create the study specific output tables. This will be double coded in Jerboa (JAVA) and SAS (version 9.2).

2. Data sharing

   A study-specific folder on the central remote research environment (RRE) for secure access by members will be used to analyze the output provided by Jerboa©. These output files will contain only anonymized de-identifiable data that will be shared in the RRE where members will have a secure and restricted access and where data will be analyzed.

   SAS, version 9.2, will be used to analyze the time to colorectal hyperplasia by Kaplan Meier estimation.

9.7 Data analysis

All statistical analyses will be performed on the RRE (using SAS version 9.2) by [redacted], the coordinating centre of this multi-database study.

9.7.1 Primary objective

All analyses will be performed by [redacted] on output of tables provided by each partner according to the processes described above.

The crude, age, sex and year-month specific incidence rates of colorectal hyperplasia in hypertension patients will be estimated by dividing the total number of incident cases by the total number of person-years at risk within the study population. Similarly, incidence rates will be calculated specifically by each of the risk factors detailed in section 9.3.2. Aggregation of cases and person-time in age, sex and other co-variate strata will be done by Jerboa, prior to sharing data. 95% confidence intervals (95% CIs) for the incidence rates will be computed based on a Poisson distribution. Multivariate Poisson regression will be utilized to calculate age- and sex-adjusted relative risk incidence rates for the different co-variate strata for each of the databases.

Age- and sex-standardized incidence rates of colorectal cancer will be calculated by using 1) a World reference population to compare with the general population and 2) the CPRD age/sex distribution of the hypertensive cohort for internal comparisons between databases. The incidence rates obtained in the study will be applied to the age distribution (person-time) of these populations; this will yield the number of expected patients in these populations per year of which the incidence rates may be obtained if they were hypertensive. This can then be compared with the known/observed incidence rate in these countries/databases.

The analysis will be performed for the composite endpoint (using the first of the three events as incident event) but also for the three separate endpoints (colorectal cysts, polyps, neoplasms [malignant and benign neoplasms analyzed separately]), i.e. independent of metachronous events.

Amendment: A sensitivity analysis similar to the above mentioned main analysis (age and sex standardized incidence rates of colorectal cancer) will be carried out to look at the impact of excluding prevalent hypertension cases from the study cohort. Rather than excluding we will include prevalent hypertension cases in this sensitivity analysis. Patients with longer duration of hypertension are possibly more likely to develop any colorectal neoplasm and the incidence

rate may therefore be higher. This analysis is not the primary analysis because it is uncertain how long patients with complicated hypertension have hypertension and how long they have been on different types of drugs. This may also influence the subsequent development of colorectal neoplasm, which should be taken into account when interpreting the results from this analysis.

9.7.2 Secondary objective

Kaplan Meier analyses will be performed estimating the time from cohort entry until primary endpoint (colorectal polyps, colorectal neoplasms or colorectal cyst). The analysis will be done on the composite endpoint, but also on the outcomes separately. Colorectal malignant neoplasm will be considered as final endpoint (no other events can occur after malignant cancer). A Kaplan Meier survival analysis estimates the proportion of hypertensive patients being 'event-free' at each time point. A survival analysis is considered most appropriate to estimate the one-year risk of colorectal polyps/neoplasms/cysts in hypertensive patients as right-censoring (i.e. patients leaving the study cohort before they were able to develop cancer) is taken into account in the Kaplan Meier estimator. This analysis will be done in the study population from each database individually and also stratified by sex, age (in age groups being 18-44, 45-64, 65-79 years) and prior use of antihypertensive drugs and which allows exploring whether the progression of CRC differs between subgroups.

Subgroup analyses include stratification by sex, age groups (18-44, 45-64, 65-79 years), IBD and subjects having had a 'negative' colonoscopy or not.

9.8 Quality control

The use of common software (i.e. Jerboa©) makes it possible to overcome issues and problems that arise when using multiple databases. Using a common protocol and common software mitigates against differences in programming or study definitions between databases.

As part of quality control, validation of the primary endpoint (colorectal hyperplasia) will be conducted by calculating the positive predictive value (PPV) of a diagnosis code of colorectal malignant neoplasms. This activity will be carried out only in databases where additional information or specialist notes can be obtained, as will be done in IPCI and CPRD. Charts and letters will be obtained for a random sample of around 200 cases per outcome per database. The validation will be done through a common questionnaire that will be developed. Cases of colorectal neoplasms will be classified according to the validation process, because colorectal neoplasms is a specialist-confirmed diagnosis. The classification will be as follows:

- Definite case: Diagnosis of colorectal hyperplasia, polyp, cancer which includes a letter from a specialist, endoscopy report confirming the colorectal neoplasm.
- Probable case: Diagnosis of colorectal hyperplasia, polyp, cancer but without confirmation from a specialist, or endoscopy report.
- No case: No mentioning of colorectal hyperplasia, polyp, cancer.

The impact of using a validation process of the outcomes is that overestimation of colorectal neoplasms is mitigated. Secondly, in particular for IPCI clinical terms or codes may correspond to historic events; and thus to avoid inclusion of prevalent events in the study period, validation of the outcomes can assure the index date (date of start of symptoms

leading to colorectal neoplasm diagnosis) and may thus help in classifying between incident and prevalent events.

9.9 Limitations of the research methods

The limitations of this study will be mainly due to availability and level of detail of the data. Not in all databases, appropriate follow-up measurements of blood pressure according to clinical guidelines may be available. This may be related to patient's related adherence to visit the GP to measure the blood pressure regularly, but may also be related to GPs adherence to record and code every blood pressure measurement. Additionally, information on potential confounders (such as lifestyle factors or physical activity) for colorectal cancer may not be recorded in the desired detail. Although this may result in misclassification, we do include the most important other co-variates for colorectal neoplasms that contribute in larger extent to the risk of colorectal cancer (e.g. T2DM, genetic cancer syndromes). Second, lifestyle factors are indirectly corrected for by the inclusion of BMI as co-variate in the study. In addition, recording of smoking status and alcohol use is likely underreported in the databases. Alcohol (ab)use is only recorded when a subject may seek contact to the GP for problems related to alcohol use or when rehabilitation is desired by the patient. A similar explanation holds for smoking status; if subjects seek help to quit or diminish smoking this is likely to be recorded, as it will be recorded regularly for patients with COPD, asthma or other lung diseases.

Misclassification of the primary safety endpoint (colorectal hyperplasia) may be present. This is the most important reason to consider validation in CPRD, IPCI and if applicable in the other databases. By not performing case validation in the other databases, we may overestimate the incidence of colorectal neoplasms since we include false positive cases. In addition, the date of the outcome could be misclassified. We mitigate against this misclassification by considering the date of recording of first symptoms leading to a colorectal neoplasm diagnosis (e.g. melena, change in stool frequency and/or consistency, colonoscopy referral); if this information is not available/retrievable, the date of diagnosis will be considered. By that, a potential misclassification of the date of outcome will be limited.

As the risk of CRC increases with higher age, reaching age of 80 years was considered as censoring point. This is relevant as subjects 80 years and older are more likely to be moved to a nursing home and thus the underlying source population in the databases within this age group is uncertain. Generalization of the results is therefore not possible for hypertensive subjects aged 80 or older.

Presence of Berkson bias is likely in the hypertensive cohort given the fact they are routinely followed in medical care. In addition, the use of antiplatelets medication may enhance the likelihood of colorectal bleeding, for which they might undergo a colonoscopy. The chance of finding any of the primary endpoints is therefore higher in a hypertension cohort as compared to the general population. Therefore, a subgroup analysis in hypertensive subjects who had a 'negative' colonoscopy will be performed. Only in a population with a recorded 'negative' colonoscopy one can be certain that they were event-free at a certain point in time. At cohort entry it may very well occur that subjects do have polyps, but have not yet been investigated/diagnosed since they are asymptomatic. To identify true 'incident' colorectal polyps, a subject must have had a 'negative' colonoscopy at an earlier point in time. In order

to account for this issue, a subgroup analysis is planned in subjects with a 'negative' colonoscopy.

Screening and surveillance programs for colorectal cancer, for instance by means of examining occult blood in feces or by sigmoidoscopy or colonoscopy, are aimed to identify subjects at an early stage of colorectal cancer development: thus, at the stage of colon polyps. Nationwide colorectal cancer screening has been started in five European countries, including Italy, the United Kingdom and the Netherlands.

- In Italy, nationwide screening (mostly done by Fecal Immunochemical Testing [FIT]; but in some regions in Northern Italy done by sigmoidoscopy). When these programmes have been started differs per region, and to which extent screening areas covered by HSD is something to explore.
- In the United Kingdom, a phased nationwide screening programme is implemented since 2005-2006, using the guaiac based test for fecal occult blood (gFOBT).
- In the Netherlands, since January 2014 a phased nationwide screening program will be started using the Fecal occult blood (FOB) gold test.

The detection rate of polyps increases by screening programmes, however, such results will become visible after calculating the incidence rates of the endpoints separately over calendar time. A striking increase in incidence rates of polyps will likely be the result of implementation of colorectal screening strategies.

A non-hypertensive population will not be included in the study. Whether the incidence of hyperplasia, polyps and colorectal cancer is higher for hypertensive subjects as compared to the general population is difficult to answer. There is the risk that without similar diagnostic work-up in both populations, in a regular care setting, Berkson bias may occur. A comparison with age-/sex-specific incidence rates may be performed against cancer registry data. Novartis study SPP100A2418 is planned to unravel the influence of hypertensive drugs on the risk of colorectal cancer and more accurately provide answers to this question.

We cannot be certain that at cohort entry patients are 'free' of any colon hyperplasia or polyps, since for this evaluation/conclusion reports from colonoscopy are required. However, since many people will not have undergone a colonoscopy, we will exclude only those with documented colorectal hyperplasia, but we will falsely leave in the subjects with 'silent/unknown hyperplasia'. Subsequently, this will impact the results, although the direction of the bias is unknown.

None of the databases participating in the current study can identify drug dispensing, nor actual drug intake. This is a limitation of the study.

9.10 Other aspects

Not applicable.

10 Protection of human subjects

10.1 Regulatory and ethical compliance

According to the local legal/ethical requirements of each database, the study protocol will be submitted to the Review Boards / Ethics Committees if needed.

This study was designed and shall be implemented and reported in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology (ISPE 2008), the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Vandenbroucke et al 2007), and with the ethical principles laid down in the Declaration of Helsinki.

Although this study is not obligatory to be submitted to the 'European Network of Centres for Pharmacoepidemiology and Pharmacovigilance' (ENCePP), it is fulfilling the criteria of ENCePP and follows the 'ENCePP Code of Conduct' (ENCePP 2011).

10.2 Informed consent procedures

No informed consent is needed in the countries participating in the study, however, approval from each of the boards of the databases is required.

11 Management and reporting of adverse events/adverse reactions

As this is a study based on secondary data sources, safety monitoring and safety reporting on an individual case level is not applicable. In studies based on secondary data sources with a safety relevant result, only aggregated safety results, i.e. the overall association between an exposure and an outcome, should be reported and be included in the periodic aggregated regulatory reports submitted to Health Authorities.

12 Plans of disseminating and communicating study results

Upon study completion and finalization of the study report, the results of this non-interventional study may be either submitted for publication and/or posted in a publicly accessible database of results. Publications will comply with internal Novartis standards and the International Committee of Medical Journal Editors (ICMJE) guidelines.

13 References (available upon request)

Ahmed RL, Schmitz KH, Anderson KE, et al (2006) The metabolic syndrome and risk of incident colorectal cancer. Cancer; 107(1):28-36.

Aleksandrova K, Boeing H, Jenab M, et al (2011) Metabolic syndrome and risks of colon and rectal cancer: the European prospective investigation into cancer and nutrition study. Cancer Prev Res; 4(11):1873-83.

Angeli F, Reboldi G, Mazzotta G, et al (2012) Safety and efficacy of aliskiren in the treatment of hypertension and associated clinical conditions. Curr Drug Saf; 7(1):76-85.

Azoulay L, Assimes TL, Yin H, et al (2012) Long-term use of angiotensin receptor blockers and the risk of cancer. PLoS ONE; 7(12):e50893 (Internet) Available from: <http://www.plosone.org/article/fetchObject.action;jsessionid=DC6E5A21ED81AA0181DD226F292E475C?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0050893&representation=PDF> (Accessed 19 June 2013).

Batty GD, Shipley MJ, Marmot MG, et al (2003) Blood pressure and site-specific cancer mortality: evidence from the original Whitehall study. Br J Cancer; 89:1243-7.

Bolíbar B, Fina Avilés F, Morros R, et al (2012) SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research. Med Clin (Barc); 138(14):617-21.

Booth N (1994) What are the Read Codes? Health Libr Rev; 11(3):177-82.

Burke TA, Sturkenboom MC, Lu S-e, et al (2006) Discontinuation of antihypertensive drugs among newly diagnosed hypertensive patients in UK general practice. J Hypertens; 24(6):1193-200.

Cazzola M, Puxeddu E, Bettoncelli G, et al (2011) The prevalence of asthma and COPD in Italy: a practice-based study. Respir Med; 105(3):386-91.

Center MM, Jemal A, Smith RA, et al (2009) Worldwide variations in colorectal cancer. CA Cancer J Clin; 59(6):366-78.

Chisholm J (1990) The Read clinical classification. BMJ; 300(6732):1092.

Coloma PM, Schuemie MJ, Trifirò G, et al (2011) Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf; 20(1):1-11.

Cricelli C, Mazzaglia G, Samani F, et al (2003) Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. J Public Health Med; 25(3):254-7.

Eaden JA, Abrams KR, Mayberry JF (2001) The risk of colorectal cancer in ulcerative colitis: a meta-analysis. Gut; 48(4):526-35.

ENCePP (2011) The ENCePP code of conduct for scientific independence and transparency in the conduct of pharmacoepidemiological and pharmacovigilance studies. London: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP), 21 November 2011; EMA/929209/2011 (Internet) Available from: <http://www.encepp.eu/code_of_conduct/documents/CodeofConduct_Rev2.pdf> (Accessed 21 November 2013).

Esposito K, Chiodini P, Colao A, et al (2012) Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care; 35(11):2402-11.

Filippi A, Vanuzzo D, Bignamini AA, et al (2005) The database of Italian general practitioners allows a reliable determination of the prevalence of myocardial infarction. Ital Heart J; 6(4):311-4.

Fisher NDL, Danser AHJ, Nussberger J, et al (2008) Renal and hormonal responses to direct renin inhibition with aliskiren in healthy humans. Circulation; 117(25):3199-205.

García Rodríguez LA and Huerta-Alvarez C (2001) Reduced risk of colorectal cancer among long-term users of aspirin and nonaspirin nonsteroidal antiinflammatory drugs. Epidemiology; 12(1):88-93.

García-Gil MdM, Hermosilla E, Prieto-Alhambra D, et al (2011) Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). Inform Prim Care; 19(3):135-45.

Grossman E, Messerli FH, Boyko V, et al (2002) Is there an association between hypertension and cancer mortality? Am J Med; 112(6):479-86.

Healey JS and Connolly SJ (2003) Atrial fibrillation: hypertension as a causative agent, risk factor for complications, and potential therapeutic target. Am J Cardiol; 91(10A):9G-14G.

Herrinton LJ, Liu L, Levin TR, et al (2012) Incidence and mortality of colorectal adenocarcinoma in persons with inflammatory bowel disease from 1998 to 2010. Gastroenterology; 143(2):382-89.

Hong J-L, Meier CR, Sandler RS, et al (2013) Risk of colorectal cancer after initiation of orlistat: matched cohort study. BMJ; 347:f5039 (Internet) Available from: <http://www.bmj.com/highwire/filestream/659381/field_highwire_article_pdf/0/bmj.f5039> (Accessed 22 November 2013).

ISPE (2008) Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf; 17(2):200-8.

Jick H, Jick SS, Derby LE (1991) Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ; 302(6779):766-8.

Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al (2003) Validity of the general practice research database. Pharmacotherapy; 23(5):686-9.

Kim JH, Lim YJ, Kim Y-H, et al (2007) Is metabolic syndrome a risk factor for colorectal adenoma? Cancer Epidemiol Biomarkers Prev; 16(8):1543-6.

Lamberts H, Wood M, Hofmans-Okkes IM (1992) International primary care classifications: the effect of fifteen years of evolution. Fam Pract; 9(3):330-9.

Lever AF, Hole DJ, Gillis CR, et al (1999) Is cancer related to hypertension or to its treatment? Clin Exp Hypertens; 21(5-6):937-46.

Lindgren AM, Nissinen AM, Tuomilehto JO, et al (2005) Cancer pattern among hypertensive patients in North Karelia, Finland. J Hum Hypertens; 19:373-9.

Lindholm LH, Anderson H, Ekbom T, et al (2001) Relation between drug treatment and cancer in hypertensives in the Swedish Trial in Old Patients with Hypertension 2: a 5-year, prospective, randomised, controlled trial. Lancet; 358(9281):539-44.

Musini VM, Fortin PM, Bassett K, et al (2011) Blood pressure lowering efficacy of renin inhibitors for primary hypertension (review). Cochrane Database Syst Rev; (4):CD007066 (Internet) Available from: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007066.pub2/pdf> (Accessed 22 November 2013).

Negri E, Bosetti C, La Vecchia C, et al (1999) Allergy and other selected diseases and risk of colorectal cancer. Eur J Cancer; 35(13):1838-41.

Othman NH and Zin AAM (2008) Association of colorectal carcinoma with metabolic diseases; experience with 138 cases from Kelantan, Malaysia. Asian Pac J Cancer Prev; 9:747-51.

Pelucchi C, Negri E, Talamini R, et al (2010) Metabolic syndrome is associated with colorectal cancer in men. Eur J Cancer; 46(10):1866-72.

Ramos R, Balló E, Marrugat J, et al (2012) Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. Rev Esp Cardiol; 65(1):29-37.

Schmieder RE, Philipp T, Guerediaga J, et al (2009) Long-term antihypertensive efficacy and safety of the oral direct renin inhibitor aliskiren: a 12-month randomized, double-blind comparator trial with hydrochlorothiazide. Circulation; 119(3):417-25.

Stocks T, Lukanova A, Johansson M, et al (2008) Components of the metabolic syndrome and colorectal cancer risk; a prospective study. Int J Obes; 32:304-14.

Sturkenboom MCJM, Dieleman JP, Picelli G, et al (2008) Prevalence and treatment of hypertensive patients with multiple concomitant cardiovascular risk factors in The Netherlands and Italy. J Hum Hypertens; 22(10):704-13.

Stürmer T, Buring JE, Lee I-M, et al (2006) Metabolic abnormalities and risk for colorectal cancer in the Physicians' Health Study. Cancer Epidemiol Biomarkers Prev; 15(12):2391-7.

van Soest EM, von Rossum LGM, Dieleman JP, et al (2008) Proton pump inhibitors and the risk of colorectal cancer. Am J Gastroenterol; 103(4):966-73.

van Staa TP, Card T, Logan RF, et al (2005) 5-Aminosalicylate use and colorectal cancer risk in inflammatory bowel disease: a large epidemiological study. Gut; 54(11):1573-8.

Vandenbroucke JP, von Elm E, Altman DG, et al (2007) Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiology; 18(6):805-35.

Vlug AE, van der Lei J, Mosseveld BMTh, et al (1999) Postmarketing surveillance based on electronic patient records: the IPCI project. Methods Inf Med; 38(4-5):339-44.

Voordouw ACG, Sturkenboom MCJM, Dieleman JP, et al (2004) Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. JAMA; 292(17):2089-95.

Watanabe Y, Ozasa K, Ito Y, et al (2005) Medical history of circulatory diseases and colorectal cancer death in the JACC study. J Epidemiol; 15(Suppl 2):S168-72.

Yang Y-X, Hennessy S, Lewis JD (2004) Insulin therapy and colorectal cancer risk among type 2 diabetes mellitus patients. Gastroenterology; 127(4):1044-50.

Appendix 1 – Diagnosis codes for hypertension

ICPC:	K86, K87	
ICD-9:	401.x	Essential hypertension
	402.xx	Hypertensive heart disease
	403.x	Hypertensive chronic kidney disease
	404.x	Hypertensive heart and chronic kidney disease
	405.xx	Secondary hypertension
	362.11	Hypertensive retinopathy
	437.2	Hypertensive encephalopathy
ICD-10:	I10	Essential (primary) hypertension,
	I11.x	Hypertensive heart disease
	I12.x	Hypertensive renal disease
	I13.x	Hypertensive heart and renal disease
	I15.x	Secondary hypertension
	O10	Pre-existing hypertension complicating pregnancy, childbirth and the puerperium
	O11	Pre-existing hypertensive disorder with superimposed proteinuria
	I67.4	Hypertensive encephalopathy