CARING

CAncer Risk and INsulin analoGues

Common Study Protocol

Version 3.0

Date: November 2013

HEALTH.2011.4.2-2
Adverse Drug Reaction Research

Grant agreement no: 282526

Table of contents

1. List of abbreviations

The following abbreviations are used in this report:

- **AD** – Antidiabetic
- **ATC** – Anatomical Therapeutic Chemical
- **BMI** – Body Mass Index
- **CPRD** – Clinical Practice Research Datalink
- **DDD** – Defined Daily Dose
- **DPP-4** – Dipeptidyl Peptidase 4
- **DM** – Diabetes Mellitus
- **GLP-1** – Glucagon-Like Peptide-1
- **GP** – General Practitioner
- **HES** – Hospital Episode Statistics
- **HR** – Hazard Ratio
- **ICD9** – International Classification of Diseases, 9th revision
- **ICD10** – International Classification of Diseases, 10th revision
- **OAD** – Oral Antidiabetic Drug
- **T1DM** – Type 1 Diabetes Mellitus
- **T2DM** – Type 2 Diabetes Mellitus
- **TZD** – Thiazolidinedione
- **UK** – United Kingdom
- **WP** – Work Package

2. Collaborating institutes

- Utrecht University (Netherlands)
- Aarhus University Hospital (Denmark)
- Norwegian Institute of Public Health (Norway)
- Karolinska Institutet (Sweden)
- University of Tampere (Finland)
- University of Helsinki (Finland)
- Netherlands Cancer Institute (Netherlands)

3. Background

Both diabetes and cancer are prevalent diseases whose incidence is increasing globally. Worldwide, cancer is the 2nd and diabetes is the 12th leading cause of death. Cancer and diabetes are diagnosed within the same individual more frequently than would be expected by chance, even after adjusting for age. Both diseases are complex, with multiple subtypes. Diabetes is typically divided into 2 major subtypes, type 1 and type 2, along with less common types, whereas cancer is typically classified by its anatomic origin, within which there may be multiple subtypes.

Some cancers develop more commonly in patients with diabetes (predominantly type 2), whereas prostate cancer occurs less often in men with diabetes. The relative risks imparted by diabetes are greatest (approximately 2-fold or higher) for cancers of the liver, pancreas, and endometrium, and lesser (approximately 1.2-fold to 1.5- fold) for cancers of the colon/rectum, breast, and bladder. Other cancers (e.g., those of the lung) do not appear to be associated with an increased risk in diabetes, and the evidence for others (e.g., kidney and non-Hodgkin lymphoma) is inconclusive [Giovannucci, Vigneri, Nicolucci]. Only few studies have explored links between T1DM and cancer.

It remains unclear whether the association between diabetes and cancer is direct (e.g., due to hyperglycaemia), whether diabetes is a marker of underlying biologic factors that alter cancer risk (e.g., insulin resistance and hyperinsulinemia), or whether the association between cancer and diabetes is indirect and due to common risk factors such as obesity. Whether cancer risk is influenced by the duration of diabetes is a critical and complex issue and may be complicated further by the multidrug therapy often necessary for diabetes treatment.

Several studies have linked the use of insulins to the risk of cancer. Insulin is a growth factor, and it is biologically plausible that high levels of endogenous insulin and/or exposure to administered insulin could stimulate neoplastic growth. A growth-promoting effect of insulin on cancer cells has been known for more than 30 years, with these findings pre-dating the commercial availability of insulin analogues or even of recombinant human insulin. The most plausible hypothesis concerning the mechanism underlying the potential link between insulin and related peptide hormones and cancer growth is that these act through the insulin and insulin like growth factor (IGF) 1 receptors to stimulate cell growth and inhibit apoptosis [Pollak]. However, most of the studies that evaluated the risk of cancer with the use of insulins suffered from methodological drawbacks, and results have been conflicting.

4. Study objective

The primary aim of the study is to quantify the risk of cancer associated with the (long-term) use of insulin and insulin analogues[1] by studying the effects of dosage, duration

[1] Insulin and insulin analogues together are often referred to as ‘insulins’ in this protocol.

and/or intensity of insulin treatment on the likelihood of developing cancer and different types of cancer.

Secondary aims are:
● to study the hazard patterns of cancer risk associated with the use of insulin and insulin analogues over time, to test plausible biological mechanisms of initiation/promotion in clinical practice.
● to study the influence of potential confounders on the association between insulin (type, dose, duration) and cancer risk in observational studies, and to use or develop methods for external adjustment for unmeasured confounders in databases with missing information on certain confounders.
● to develop methods for screening/predicting patients at high risk of developing cancer by identification of predictors (risk factors) for developing cancer in patients with diabetes mellitus treated with insulins.
● to develop methods to combine various pharmacoepidemiological databases with long-term follow-up.

5. Study design

We will perform a series of population-based cohort studies, using the Norwegian, Swedish, Danish and Finnish National Health Registries and the Clinical Practice Research Datalink (CPRD) from the United Kingdom. Studies will be performed in all databases separately, and multi-country data will be combined using several approaches.

5.1 Data sources

Computerized medical records of all contacts to hospitals, the use of drugs in out-patients, and causes of death are kept in Denmark, Finland, Sweden and Norway for the entire population, covering a total of 25 million people. In addition, national cancer registries are available. Within each country, all registers can be linked through the use of a person specific code and the validity of the registers is high [Furu].
The CPRD comprises computerized medical records of GPs including about 12 million patients from 1987 onwards of which approximately 5.3 million (8.5% of the UK population) are currently active. GPs play a gatekeeper role in the UK health care system, as they are responsible for primary health care and specialist referrals. Patients are semi-permanently affiliated to a practice, which centralizes the medical information from the GPs, specialist referrals and hospitalisations. The data recorded in the CPRD include demographic information, prescription details, clinical events, preventive care provided, specialist referrals, laboratory results, hospital admissions and death. CPRD data will be linked to national Hospital Episode Statistics (with extended data on all hospitalisations), Death Certificates and the Cancer Registry.

The following table gives an overview of all data sources with the corresponding duration of data collection.

	Norway	**Sweden**	**Denmark**	**Finland**	**CPRD**
National Prescription Registry	2004-2011	July 2005-2011	1996-2011	1996-2011	X
National Hospital Discharge Registry	2008-2012	(1964-, national coverage 1987-, outpatients 2001-) 1987-2011	1977-2011	X	X
National Migration Registry	1961-2009	(1967-) 1987-2011	1968-2011	X	X
Cancer Registry	1953-2010	1958-2011	1977-2011	1953-2011	Years not known yet.
Cause of Death Registry	1961-2010	(1961-) July 2005-2011	1977-2011	1995-2011	1987-2012
National Registry on socioeconomic status	1960, 1970, 1980, 1990, 2001	(1990-) 2005-2011	1996-2011	X	X
Norwegian health surveys	1974-2003	X	X	X	X
Special reimbursement database of SII	X	X	X	1996-2011	X
Patient data from GPs	X	X	X	X	1987-2013
Medical data from GPs	X	X	X	X	1987-2013
Prescription data from GPs	X	X	X	X	1987-2013
Hospital Episode Statistics (England)	X	X	X	X	1997-2011

5.2 Study periods

The study periods are defined as follows, for the different data sources:

Data source	Study period	Reason for start date
Norway	2004-2010	start Prescription Registry
Sweden	2005-2011	start Prescription Registry
Denmark	1996-2011	start Prescription Registry
Finland	1996-2010	start Prescription Registry
CPRD	1987-2012	start CPRD

5.3 Cohorts of AD drug users

5.3.1 Prevalent AD drug users - Cohort A1

In all databases, the 'widest' cohort of AD drug users will consist of all prevalent AD drug users. This cohort is referred to as 'Cohort A1' in Figure 1.

Inclusion criteria

Cohort A1 will consist of all patients:

- aged 18+ during the study period
- with at least one prescription for an AD drug (OADs or insulin) during the study periods as defined in 5.2 (see Annex 1 for ATC and CPRD codes)

The start of valid data collection for a patient is defined as the start of the study period, the date that the patient turns 18 or date of immigration, whichever comes last.

From Cohort A1, different sub-cohorts will be selected. **Figure 1** displays how these sub-cohorts will be created.

5.3.2 Incident AD drug users - Cohort A2

From the main cohort of all prevalent AD drug users (A1), incident AD drug users will be selected by using a lead-in period of one year after the start of valid data collection for each patient. All patients with an AD drug prescription in this year will be excluded, resulting in Cohort A2. The index date will be defined as the first AD prescription one year after the start of valid data collection.

Exclusion criteria
All patients with a record of cancer (except non-melanoma skin cancer) before the index date will be excluded. This is done with all the history data that is available about the patients.

Cohort exit
All patients will be followed from their index date to the outcome of interest, the end of data collection, transfer out of the database, emigration or death, whichever comes first.

5.5.3 Incident insulin users - Cohort A3

From the cohort of all prevalent AD drug users (A1), incident insulin users will be selected by using a lead-in period of one year after the start of valid data collection for each patient. All patients with an insulin prescription in this year will be excluded; the use of other AD drugs does not lead to exclusion. Patients who never receive an insulin prescription are not studied in this cohort. This procedure results in Cohort A3. The index date will be defined as the first insulin prescription one year after the start of valid data collection.

Exclusion criteria
All patients with a record of cancer (except non-melanoma skin cancer) before the index date will be excluded.

Cohort exit
All patients will be followed from their index date to the outcome of interest, the end of data collection, transfer out of the database, emigration or death, whichever comes first.

Subgroups
We will study three sub-cohorts of the cohort of all incident insulin users (A3):

- patients with an OAD drug prescription in the lead-in year (Cohort A3A)
- patients without any AD drug in the lead-in year who start on an insulin (Cohort A3B)
- patients without any AD drug in the lead-in year who start on an OAD and later become insulin users (Cohort A3C)

5.4 Non-diabetic control cohort – Cohort B1

Each AD drug user from Cohort A1 will be matched to one non-diabetic control by year of birth and sex. For the CPRD, patients will additionally be matched by practice and for the Scandinavian countries patients will be matched by region (whenever possible). Non-diabetic controls are selected using incidence density sampling. This means that person-time is labeled as ‘non-diabetic control’ as long as a person is unexposed to AD drugs (OAD or insulin). If the person starts using an AD drug during follow-up, he will be censored as control and will become an exposed patient.

Controls are assigned the same index date as their matched AD drug user.

Exclusion criteria
All controls with a record of cancer (except non-melanoma skin cancer) before the index date will be excluded. This will be done with all the history data that is available about the patients.

Cohort exit
All controls will be followed from their index date to the start of AD medication, the outcome of interest, the end of data collection, transfer out of the database, emigration or death, whichever comes first.

5.5 Exposure definition

Exposure to AD drugs will be assessed in a time-dependent manner.

5.5.1 Definition of antidiabetic drug classes

We will classify AD drugs into the following sub-classes:

- insulins and analogues (ATC code starting with: A10A)
- biguanides (A10BA)
- sulfonylureas (A10BB)
- glitazones, also known as thiazolidinediones (A10BG)
- DPP-4 inhibitors (A10BH)
- glinides, also known as meglitinides (A10BX02, A10BX03,A10BX08)
- GLP-1 analogues (A10BX04, A10BX07)
- other AD drugs

Within each sub-class, individual drugs can be studied at ATC5-level. A more detailed classification of ATC codes is provided in Annex 1.

5.5.2 Definition of time windows

The total period of follow-up for each patient will be divided into fixed time periods of 90 days, starting at the index date. For non-diabetic control persons, the total period of follow-up will also be divided into periods of 90 days.

5.5.3 Definition of current exposure

At the start of each 90-day period, each patient will be classified as a current user of AD medication if they had an AD prescription on that start date or in the three months (89 days) before.

5.5.4 Definition of cumulative exposure / duration of treatment

In every database, prescription records contain the date of prescribing /dispensing, the amount that is prescribed /dispensed (in DDD), and the ATC codes (CPRD: Read codes). Based on this information, cumulative exposure will be measured at the start of every time-window. **Figure 2** graphically displays the measurement of cumulative exposure.

Cumulative exposure will be determined for any insulin (all types together), and separately for human insulin, insulin glargine, insulin detemir and other types of insulins. When studying OAD drugs, cumulative exposure is measured for any OAD, and separately for the different classes of oral agents defined in Section 5.5.1.

Based on the distribution of cumulative exposure, all time periods will be divided into categories, such as low, mid and high exposure. The cut-off points will be based on the distribution of the actual data.

At the start of each interval, the time since the first prescription will be calculated. The calculations will be carried out for any insulin (all types together), and separately for human insulin, insulin glargine, insulin detemir and other types of insulins.

Figure 1: Flowchart of cohorts

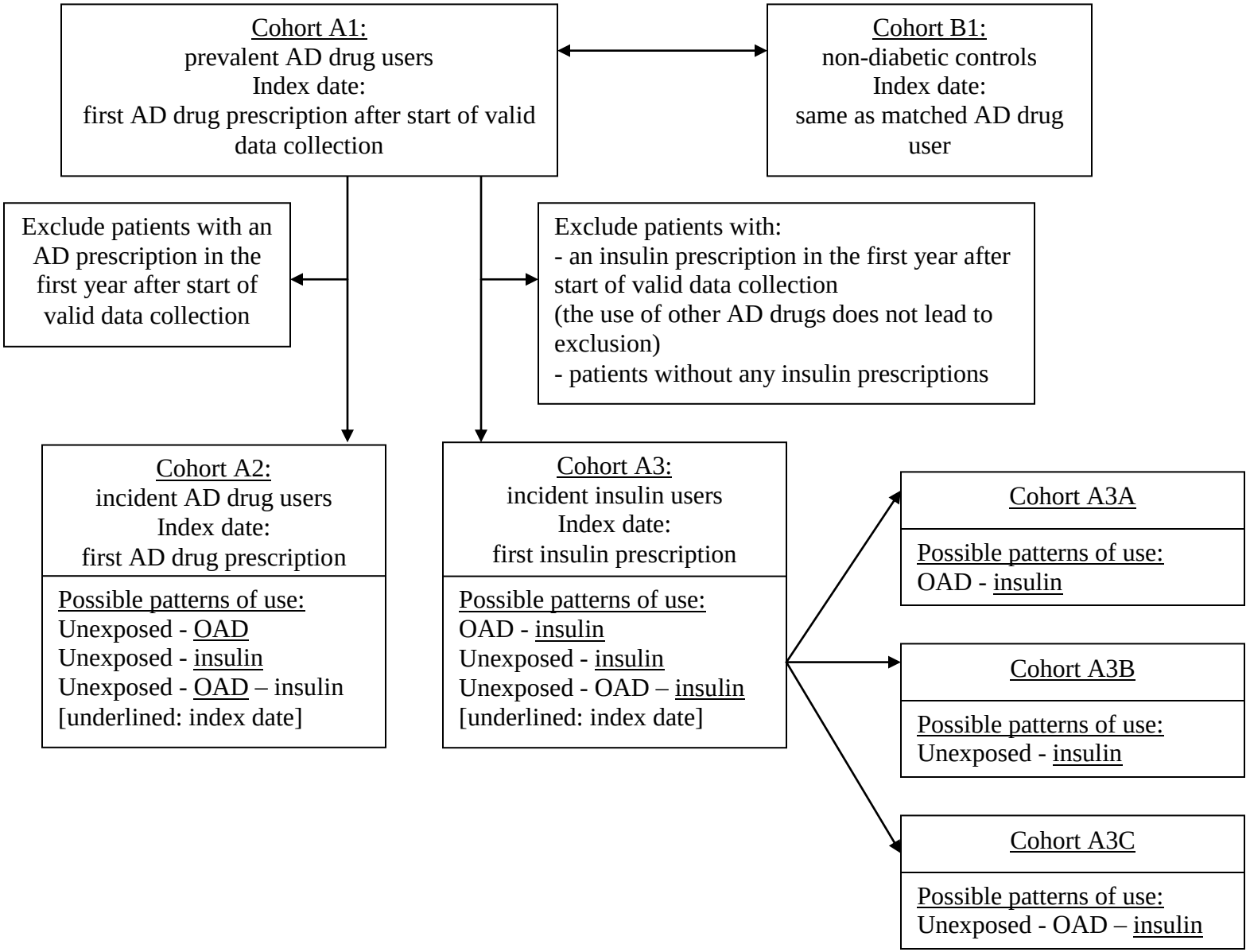
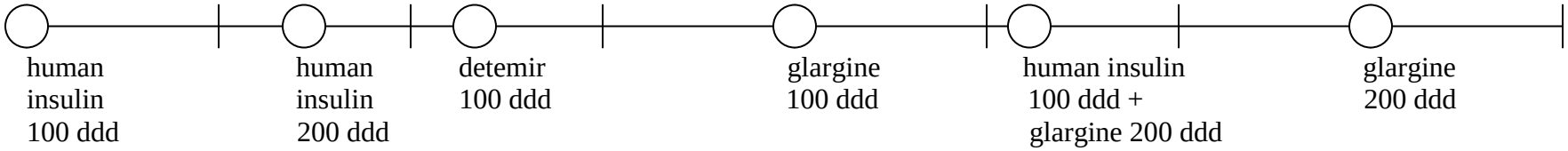


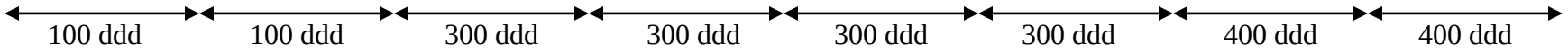
Figure 2: Cumulative exposure classification for an imaginary individual

Prescription pattern:

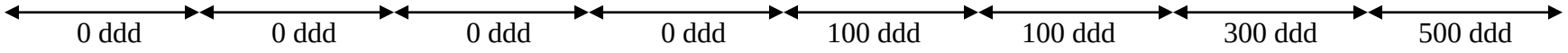


Cumulative exposure (periods of 90 days):

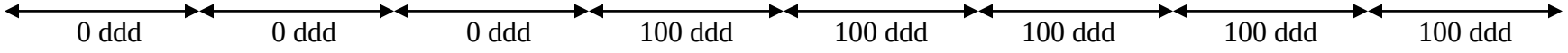
Human insulin



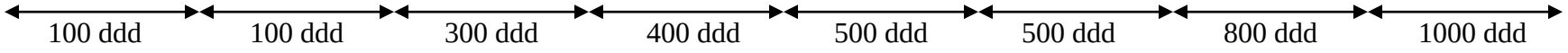
Glargine



Detemir



Any insulin



5.6 Outcome definition

1.) The primary outcome of the study is any cancer, excluding non-melanoma skin cancer. The reason for excluding non-melanoma skin cancer is incomplete registration, different treatment than other cancers (only surgery) and different impact on survival than other cancers.

2.) Secondary outcomes are the 8 most frequently occurring types of cancer in addition to pancreatic and liver cancer.

To determine the types of cancers to be studied, we relied on data on the incidence of different cancer types in the general population of the Nordic countries, as drawn from the Nordcan database. In addition, cancer of the pancreas and liver were added because of importance in diabetes patients. Non-melanoma skin cancer was excluded. A more detailed description of the procedure that was used can be found in Annex 3.

This procedure resulted in the following cancer types:

- prostate cancer
- breast cancer
- colorectal cancer
- lung cancer
- bladder cancer
- melanoma of skin
- cancer of corpus uteri
- non-Hodgkin lymphoma
- pancreatic cancer
- liver cancer

In the main analysis we will censor at the first occurring cancer.

5.7 Confounder definition

The selection of confounders has been determined in 4 steps. A detailed description of these steps can be found in Annex 4. In short, they comprised:

1.) A list of possible confounders for cancer in diabetics has been formed by conducting a literature search in several databases.
2.) Only the confounders that were shown to significantly associate to risk of cancer were selected for the list. All confounders by medical diabetes treatment were excluded from the list (e.g. metformin, sulfonylureas, insulin).
3.) Additional confounders have been added from IARC list of classifications by cancer site. From the list carcinogens with sufficient and limited evidence in humans have been selected.
4.) Confounders that were not measurable in any of the databases were excluded.

The list of all possible confounders as was found in step 3.) is shown in Annex 5.

This selection procedure resulted in the following confounders, for all of the cancer types as described in 5.7:

	Cancer type									
Potential confounder	**Bladder**	**Breast**	**Colorectal**	**Endo-metrial**	**Liver and biliary**	**Lung**	**Lymphoma**	**Melanoma**	**Pancreatic**	**Prostate**
Age	x	x	x	x	x	x	x	x	x	x
Sex	x	x	x	x	x	x	x	x	x	x
Smoking (duration/intensity)	x	x	x	x	x	x	x	x	x	x
Obesity/BMI	x	x	x	x	x	x	x	x	x	x
Education	x	x	x	x	x	x	x	x	x	x
Duration of diabetes	x	x	x	x	x	x	x	x	x	x
Alcohol intake	x	x	x	x	x	x	x	x	x	x
Glucose	x	x	x	x	x	x	x	x	x	x
Family history of cancer	x	x	x	x	x	x	x	x	x	x
Age at first child birth		x								
Hormone replacement therapy		x	x	x						
Menopausal status		x	x							
Number of child births		x								
Estrogen-progestogen contraceptives		x		x	x					
Hypercholesterolemia			x							
NSAID use			x							
Hypertension			x							
Hyperglycaemia			x							
Dislypidemia			x		x				x	x
Statin use			x		x					
Physical activity			x	x						x
Human immunodeficiency virus type 1			x				x			
Ulcerative colitis			x							
Tamoxifen use				x						
Hysterectomy				x						
Liver disease (incl cirrhosis/hepatitis)					x					
Gall bladder and bile duct disease					x					
Alcohol dependence/abuse/liver disease					x					
Dislypidemia					x					
Previous blood transfusion							x			
Azathioprine use							x			
Cyclosporine use							x			
Epstein-Barr virus							x			
Hepatitis B or C virus							x			
History of other cancer							x			

History of pancreatitis									x	
Family history of any cancer									x	
Family history of actual cancer									x	x
Cholecystectomy									x	
Finasteride use										x
Microvascular diabetic complications										x
Ischemic heart disease										x

6. Primary Statistical Analyses

The primary statistical analyses will be carried out in each of the databases.

6.1 Comparison groups

In the primary analysis, the risk of cancer will be estimated in relation to cumulative exposure to different types of insulins. The analyses will be carried out in Cohort A3 (Figure 1).

6.2 Methods

A.) Descriptive tables will be provided for each of the databases:

1. Baseline tables will be generated containing information on shared confounders (section 5.7), as well as relevant country-specific information (for example lifestyle information from CPRD and information from health surveys in Norway).

2. Tables with exposure information will be generated containing numbers of patients who are exposed to the different types of insulins (during follow-up).

3. Life tables will be constructed by aggregating the number of cancer cases and the numbers of person-years for different exposure groups:

- Cumulative exposure to any insulin
- Cumulative exposure to human insulin
- Cumulative exposure to insulin glargine
- Cumulative exposure to insulin detemir
- Cumulative exposure to other insulin
- Time since first insulin prescription
- Time since last insulin prescription
- Time since start of any AD drug
- Time since start of follow-up

These life-tables will be created for any cancer (primary outcome) and separately for the 10 most frequent cancers (secondary outcomes). They will further be stratified into different sex and age categories.

For descriptive purposes, the numbers will be aggregated into broad categories, such as low / mid / high cumulative exposure or short / mid / long use. The cut-off points will be based on the distribution of the actual data.

Exact templates for these life tables will be provided in the Statistical Analysis Plan.

B.) Life tables will also be used for analysis, i.e. in **Poisson regression models**. In that case, more categories will be used (for example by splitting cumulative exposure into 10 different categories).

Using the Poisson models, different groups will be compared:

Model 1a: insulin glargine versus human insulin

Model 1b: insulin detemir versus human insulin

Model 1c: insulin glargine versus insulin detemir

Model 2: insulin glargine, insulin detemir and other insulins versus human insulin (all different exposure groups together in 1 model)

Comparison between types of insulin will be carried out by adding an interaction term between duration of insulin and type of insulin in the Poisson regression. If this interaction term is significant (measured by maximum likelihood test), the shapes of effects of duration will be checked by plotting spline function of effects with 95% confidence intervals (see Annex 2).

The association between duration of insulin use and risk of cancer will be modeled in principle as described in [Carstensen]. However, CARING data will include more detailed information of drug use that makes it possible to make comparisons, such as between types of insulin. Utilization of spline function in modeling makes it possible to describe the association between duration of exposure and risk of cancer. An example of such spline modeling is found in Annex 2.

A minimal confounder set will be used, which contains confounders that can be measured in every database. Methods will be described by WP4.

Interactions between cumulative exposure to different types of insulins and other factors, such as sex, will be tested. For interactions where there is no prior expectation of an

association a Bonferroni correction will be applied. The Statistical Analysis Plan will further specify this.

A more detailed explanation of the Poisson modeling will also be provided in the Statistical Analysis Plan.

6.3 Combination of data sources

All data will be combined in a multi-country study, by adding up the numbers from the life tables (from every data source) and by running analyses on these combined life tables.

6.4 Sensitivity analyses around main analysis

- Sensitivity analyses will be conducted with respect to the modelling. For example, Cox proportional hazards models will be used and results will be compared with the Poisson models.

- Results will be stratified to type 1 and type 2 diabetes - because type 1 and type 2 diabetes are very different diseases, with different risk factors, prognosis, treatment, and patient characteristics. Type 1 and type 2 diabetes will be classified based on prescription records.

7. Secondary Statistical Analyses

7.1 Hazard patterns

The goal is to study the hazard patterns of cancer risk associated with the use of insulin and insulin analogues over time, to test plausible biological mechanisms of initiation/promotion in clinical practice. Methods will be described by WP6.

7.2 Influence of potential confounders

The goal is to study the influence of potential confounders on the association between insulin (type, dose, duration) and cancer risk in observational studies, and to use or develop methods for external adjustment for unmeasured confounders in databases with missing information on certain confounders. Methods will be described by WP4.

7.3 Prediction models

The goal is to develop methods for screening/predicting patients at high risk of developing cancer by identification of predictors (risk factors) for developing cancer in patients with diabetes mellitus treated with insulins. Methods will be described by WP6.

7.4 Explore methods for combining data

The goal is to develop methods to combine various pharmacoepidemiological databases with long-term follow-up, in order to increase power to study the cancer risk associated with insulin and insulin analogues in detail.

Objectives are to:
a) use minimal data sets for one large database;
b) explore the need for mapping of coding terms, in light of other on-going projects;
c) review and evaluate statistical techniques for analysis of multi-database studies;
d) analyse combined datasets using different approaches for confounder control and levels of aggregation;
e) perform simulations to evaluate statistical techniques accounting for inter-database variability;
f) adoption/finalisation in consensus meeting with all consortium members.

Methods will be described by WP3.

8. Tertiary Statistical Analyses

The tertiary statistical analyses need not necessarily to be carried out in each of the databases, but can be performed based on the researcher's interest and on the suitability of the specific data source to address the question.

8.1 Comparison groups

In a tertiary analysis, the risk of cancer will be estimated for cumulative exposure to insulin versus other AD drugs. The analyses will be carried out in Cohort A2 (Figure 1).

<u>Model 1: any insulin versus any OAD</u>

Cumulative exposure to any insulin will be divided over different groups (for example by splitting cumulative exposure into 10 different categories) and will be compared to no insulin use (only OAD).

<u>Model 2: any insulin versus specific OAD groups</u>

Cumulative exposure to any insulin will be compared with cumulative exposure to biguanides, sulfonylureas and other OAD drugs.

8.2 Methods

Methods will be developed by those partners interested to perform these analyses.

9. Quaternary Statistical Analyses

The quaternary statistical analyses need not necessarily to be carried out in each of the databases, but can be performed based on the researcher's interest and on the suitability of the specific data source to address the question.

9.1 Comparison groups

In a quaternary analysis, the risk of cancer will be estimated for cumulative exposure to insulin versus non-diabetic controls. The analyses will be carried out in Cohort A3 and B1 (Figure 1). Non-diabetic controls will be assigned the same index date as their matched AD drug user.

<u>Model: any insulin versus non-diabetics</u>

Cumulative exposure to any insulin will be divided over different groups (for example by splitting cumulative exposure into 10 different categories) and will be compared to non-diabetic controls.

9.2 Methods

Methods will be developed by those partners interested to perform these analyses.

10. Limitations

- If diabetic patients are on diet control this person-time is measured as ‘non-diabetic controls’.

- Data on patients aged 85+ may not be reliable in Sweden.

- Drug use in nursing home patients is not recorded in the Norwegian Prescription Database.

11. References

- Carstensen B, Witte D, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. Diabetologia 2012;55:948–58
- Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. Basic Clin Pharmacol Toxicol. 2010;106(2):86-94.
- Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. CA Cancer J Clin 2010 Jul;60(4):207-21.
- Nicolucci A. Epidemiological aspects of neoplasms in diabetes. Acta Diabetol 2010 Jun;47(2):87-95.
- Pollak M, Russell-Jones D. Insulin analogues and cancer risk: cause for concern or cause celebre? Int J Clin Pract 2010 Apr;64(5):628-36.
- Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R. Diabetes and cancer. Endocr Relat Cancer 2009;16(4):1103-23.

Annexes

Annex 1. Definition of diabetic prescriptions

ATC codes

DRUGS USED IN DIABETES	Insulin	Biguanides	Sulfonylurea	DPP4_blockers	Glinides	Glitazones	GLP1	Other
A10A INSULINS AND ANALOGUES *	1							
A10B BLOOD GLUCOSE LOWERING DRUGS, EXCL. INSULINS								
A10BA Biguanides		1						
A10BB Sulfonamides, urea derivatives			1					
A10BD Combinations of oral blood glucose lowering drugs								
A10BD01 phenformin and sulfonamides		1	1					
A10BD02 metformin and sulfonamides		1	1					
A10BD03 metformin and rosiglitazone		1				1		
A10BD04 glimepiride and rosiglitazone			1			1		
A10BD05 metformin and pioglitazone		1				1		
A10BD06 glimepiride and pioglitazone			1			1		
A10BD07 metformin and sitagliptin		1		1				
A10BD08 metformin and vildagliptin		1		1				
A10BD09 pioglitazone and alogliptin				1		1		
A10BF Alpha glucosidase inhibitors								1
A10BG Thiazolidinediones						1		

A10BH Dipeptidyl peptidase 4 (DPP-4) inhibitors				1				
A10BX Other blood glucose lowering drugs, excl. insulins								
A10BX02 repaglinide					1			
A10BX03 nateglinide					1			
A10BX04 exenatide							1	
A10BX05 pramlintide								1
A10BX06 benfluorex								1
A10BX07 liraglutide							1	
A10BX08 mitiglinide					1			

* INSULINS AND ANALOGUES:

A10AB Insulins and analogues for injection, fast-acting

A10AB01 insulin (human)
A10AB02 insulin (beef)
A10AB03 insulin (pork)
A10AB04 insulin lispro
A10AB05 insulin aspart
A10AB06 insulin glulisine
A10AB30 combinations

A10AC Insulins and analogues for injection, intermediate-acting

A10AC01 insulin (human)
A10AC02 insulin (beef)
A10AC03 insulin (pork)
A10AC04 insulin lispro
A10AC30 combinations

A10AD Insulins and analogues for injection, intermediate-acting combined with fast-acting

A10AD01	insulin (human)
A10AD02	insulin (beef)
A10AD03	insulin (pork)
A10AD04	insulin lispro
A10AD05	insulin aspart
A10AD30	combinations

A10AE Insulins and analogues for injection, long-acting

A10AE01	insulin (human)
A10AE02	insulin (beef)
A10AE03	insulin (pork)
A10AE04	insulin glargine
A10AE05	insulin detemir
A10AE30	combinations

A10AF Insulins and analogues for inhalation

A10AF01	insulin (human)

Annex 2. Spline modeling

An example from in Carstensen et al. 2012 paper is presented in Figure 1. The usage of duration instead of cumulative DDD in modeling is that there is large variation of dosing of insulin between subjects and thus duration would better describe the long term exposure.

Poisson regression model will be used in analyses. R language package "Epi" (Bendix Carstensen, Martyn Plummer, Esa Laara, Michael Hills (2012). Epi: A Package for Statistical Analysis in Epidemiology. R package version 1.1.34. URL http://CRAN.R-project.org/package=Epi) will be utilized. This package includes also tools for calculation of drug exposures based from prescription data.

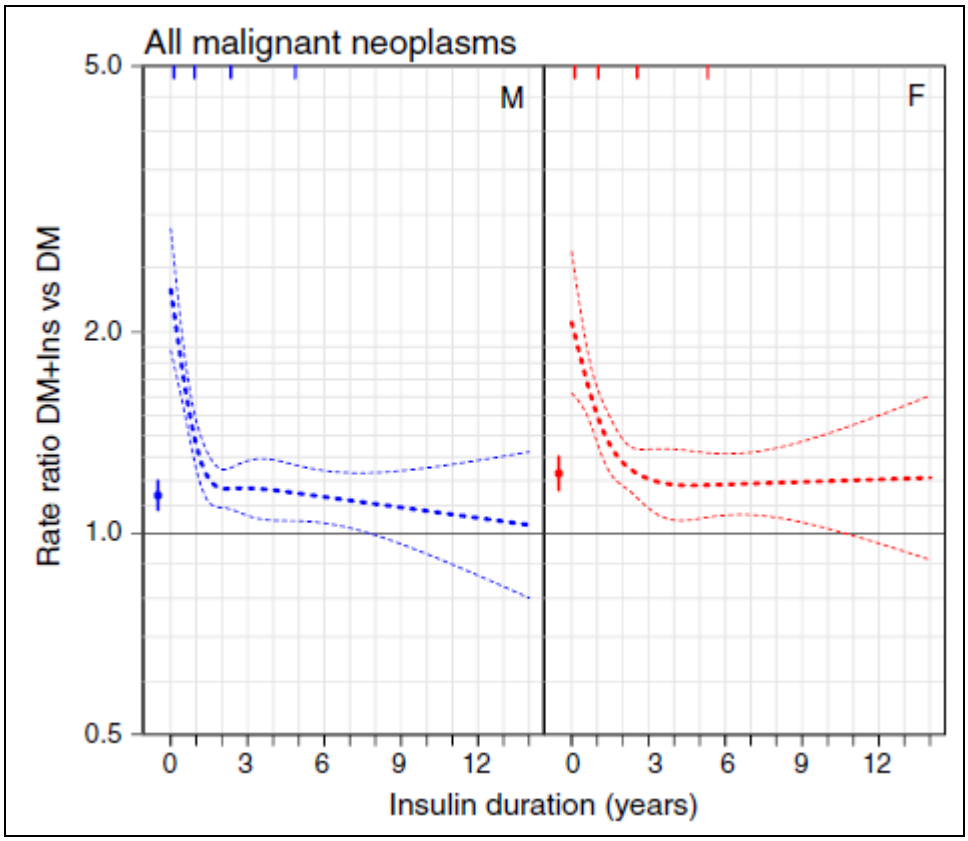


Figure 1: Cancer incidence rate ratio for insulin users vs non-users for all malignant neoplasms. Spline models with five (diabetes duration) and four (insulin duration) parameters. Blue, men; red, women. Thin lines indicate 95% CIs. The bars on the left are estimates from the model ignoring duration effects, and the coloured ticks inside indicate the location of the knots for the spline functions. F, female; M, male 2012 (Carstensen B, Witte D, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. Diabetologia 2012;55:948–58., Fig 3)

Annex 3. Definition of cancer types

The most common cancer types will be studied in CARING. Data on incidence of different cancer types in the general population of the Nordic countries is drawn from the Nordcan database. In addition, cancer of the pancreas and liver will be studied because of importance in diabetes patients. Non-melanoma skin cancer is excluded.

Description of the Nordcan webpage/database:
Nordcan is a joint activity project of the Association of Nordic Cancer Registries (ANCR), the International Agency for Research on Cancer (IARC), with financial support from the Nordic Cancer Union (NCU). Data for NORDCAN are delivered from the national cancer registries. The Association of the Nordic Cancer Registries (ANCR) has proprietary rights to the materials on the Website.
The database includes 41 common cancer types from the Nordic countries (Sweden, Denmark, Norway, Finland, Iceland, Faroe Islands).
The national incidence data were, in cooperation with each of the registries, first recoded to ICD-O-3 and then to ICD-10 using the international rules in IARCcrgTools.
Coding of cancer types by ICD-10 codes or ICD-7 is given in "The cancer dictionary"
http://www-dep.iarc.fr/NORDCAN/English/frame.asp

Data extracted for CARING project:
Online analysis => Incidence/mortality data => Tables => Cancers

Selection: Incidence, Nordic countries combined, year 2009, all age groups, separate rate by gender, sorted by crude rate.

Cancer types/sites were sorted by the highest crude incidence rate (new cases per 100,000 inhabitants) observed in either men or women combined for the Nordic countries.
The 10 cancer types/sites with highest incidence rate were identified.
Pancreas cancer and liver cancer were not among the 10, but will be studied because of high incidence in diabetic patients. They replace number 9 and 10 in the list.
The most frequent types are shown in the following table:

Table: Cancer types/sites sorted by the highest crude incidence rate observed in males or females in 2009, Nordic countries combined.
Source: Nordcan webpage/database.

	Cancer type/site	ICD-7	ICD-10	Incidence rate (per 100,000)		Remarks
				Males	Females	
1	Prostate	177	C61	192.0	-	
2	Breast	170	C50	0.7	151.1	
3	Colorectal	153-4	C18-21	69.2	62.5	
4	Lung	162.0,1,8	C33-34	57.5	44.7	
5	Colon	153	C18	41.5	43.0	Part of Colorectal cancer (no. 3)
6	Bladder etc.	181	C65-68+D09.0+D41.4	38.2	13.5	
7	Skin, non-melanoma	191	C44+C46.0	32.8	27.9	Exclude according to protocol
8	Corpus uteri	172	C54	-	29.1	
9	Melanoma of skin	190	C43	27.9	28.1	
10	Rectum and anus	154	C19-21	27.6	19.5	Part of Colorectal cancer (no. 3)
11	Non-Hodgkin lymphoma	200+202	C82-85,C96	20.3	16.1	
12	Brain, central nervous system	193	C70-72+D32-33+D42-43	15.0	18.7	
13	Ovary etc.	175	C56,C57.0-4	-	18.2	
14	Kidney	180	C64	14.8	9.5	
15	Lip, oral cavity and pharynx	140-8	C00-14	14.7	8.0	
16	Pancreas	157	C25	14.2	14.6	Include because of importance
17	Leukaemia	204	C91-95	12.8	9.5	
18	Stomach	151	C16	12.6	8.1	
19	Cervix uteri	171	C53	-	10.2	
20	Testis	178	C62	8.4	-	
21	Other leukaemia	204.0-2,8	C91-95\C9X.0	8.1	5.6	Part of Leukaemia (no. 17)
22	Oesophagus	150	C15	7.9	2.9	
23	Liver	155	C22	7.5	4.2	Include because of importance

Annex 4. Selection of confounders

The selection of confounders has been determined in 4 steps:

1.) A list of possible confounders for cancer in diabetics has been formed by conducting a literature search in the databases Pubmed, Embase, Cinahl, Web of science, Cochrane library, Svemed + and Bibliotek.dk on the subjects diabetes, neoplasms/cancer and prospective, cross-sectional and case-control studies. Also limits and subject headings were used in the search. The records were screened by title, abstract and full text by following eligibility criteria:
- The study had to collate diabetes and cancer.
- The study had to be of a specific design (prospective, cross-sectional and case-control).
- The study had to have a non diabetes reference.
- The outcome in the study must be risk of cancer.

2.) Only the confounders, which were shown to significantly associate to risk of cancer were selected for the list. All confounders by medical diabetes treatment were excluded from the list (eg. metformin, sulfonylureas, insulin).

3.) Additional confounders for the top ten incident cancers based on the NORDCAN (1) database have been added from IARC list of classifications by cancer site (2). From the list carcinogens with sufficient and limited evidence in humans have been selected.
Definition of sufficient and limited evidence by the IARC
Sufficient evidence of carcinogenicity: The Working Group considers that a causal relationship has been established between exposure to the agent and human cancer. That is, a positive relationship has been observed between the exposure and cancer in studies in which chance, bias and confounding could be ruled out with reasonable confidence. A statement that there is *sufficient evidence* is followed by a separate sentence that identifies the target organ(s) or tissue(s) where an increased risk of cancer was observed in humans. Identification of a specific target organ or tissue does not preclude the possibility that the agent may cause cancer at other sites.
Limited evidence of carcinogenicity: A positive association has been observed between exposure to the agent and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence.

4.) Confounders that were not measurable in any of the databases were excluded.

Annex 5. All possible confounders

General confounders

- Age
- Duration of diabetes
- Gender
- Obesity/BMI
- Smoking
- Education

From the IARC list with sufficient evidence

- 2,3,7,8-Tetrachlorodibenzo-para-dioxin

Specific confounders by cancer site

Bladder cancer

From the IARC list with sufficient evidence

- Aluminum production
- 4-Aminobiphenyl
- Arsenic and inorganic arsenic compounds
- Auramine production
- Benzidine
- Chlornaphazine
- Cyclophosphamide
- Magenta production
- 2-Naphthylamine
- Painting
- Rubber production industry
- Schistosoma haematobium
- Tobacco smoking
- ortho-Toluidine
- X-radiation, gamma-radiation

From the IARC list with limited evidence

- 4-Chloro-ortho-toluidine
- Coal-tar pitch
- Coffee
- Dry cleaning
- Engine exhaust, diesel
- Hairdressers and barbers (occupational exposure)
- Printing processes
- Soot
- Textile manufacturing

Breast cancer

- Number of living children
- Age at first childbirth /pregnancy
- Hormonal replacement therapy
- Family history of breast cancer
- Number of induced abortions
- Age at menarche
- Menopausal status
- Number of child births /pregnancies
- Glucose level

From the IARC list with sufficient evidence

- Alcoholic beverages
- Diethylstilbestrol
- Estrogen-progestogen contraceptives
- Estrogen-progestogen menopausal therapy
- X-radiation, gamma-radiation

From the IARC list with limited evidence

- Estrogen menopausal therapy
- Ethylene oxide
- Shiftwork that involves circadian disruption
- Tobacco smoking

Brain and central nervous system cancer

From the IARC list with sufficient evidence

- X-radiation, gamma-radiation

From the IARC list with limited evidence

- Radiofrequency electromagnetic fields (including from wireless phones)

Colorectal cancer

- Colonoscopies
- Total Energy intake (kcal/day)
- Alcohol intake (g/day or drinks/day)
- Physical activity
- Overweight at age
- Hypercholesterolemia
- Number of components of metabolic syndrome (hyperglycaemia, obesity, hypertension and dyslipidaemia)
- Fiber intake (g/day)
- Aspirin
- Menopausal status
- Hormone replacement therapy use

- Calcium intake (mg/day)
- Vitamin E intake (mg/day)
- Blood glucose level
- Eating cereal
- Milk intake
- HbA1C level
- Hypertension

From the IARC list with sufficient evidence
- Alcoholic beverages
- Tobacco smoking
- X-radiation, gamma-radiation
- Human immunodeficiency virus type 1
- Human papillomavirus type 16

From the IARC list with limited evidence (including anal cancers)
- Asbestos (all forms)
- Schistosoma japonicum
- Human papillomavirus types

Endometrial Cancer
- Physical activity
- Glucose level

From the IARC list with sufficient evidence
- Estrogen menopausal therapy
- Estrogen-progestogen menopausal therapy
- Tamoxifen

From the IARC list with limited evidence
- Diethylstilbestrol

Liver and biliary cancers
- Hepatitis
- Cirrhosis
- Alcohol dependence
- Alcoholic liver disease
- Cholelithiasis
- Jaundice
- Hemochromatosis
- HBsAg (hepatitis B)
- Anti-HCV (hepatitis C)
- Triglycerides
- Total cholesterol

- LDL cholesterol
- Cholangitis
- Cholecystitis
- Cholelithiasis
- Choledocholithiasis
- Biliary cirrhosis
- Alcohol

From the IARC list with sufficient evidence

- Aflatoxins
- Alcoholic beverages
- Clonorchis sinensis
- Estrogen-progestogen contraceptives
- Hepatitis B virus
- Hepatitis C virus
- Opisthorchis viverrini
- Plutonium
- Thorium-232 and its decay products
- Tobacco smoking (in smokers and in smokers' children)
- Vinyl chloride
- Thorium-232 and its decay products

From the IARC list with limited evidence

- Androgenic (anabolic) steroids
- Arsenic and inorganic arsenic compounds
- Betel quid without tobacco
- Human immunodeficiency virus type 1
- Polychlorinated biphenyls
- Schistosoma japonicum
- Trichloroethylene
- X-radiation, gamma-radiation

Lung cancer

From the IARC list with sufficient evidence

- Aluminum production
- Arsenic and inorganic arsenic compounds
- Asbestos (all forms)
- Beryllium and beryllium compounds
- Bis(chloromethyl)ether; chloromethyl methyl ether (technical grade)
- Cadmium and cadmium compounds
- Chromium(VI) compounds
- Coal, indoor emissions from household combustion
- Coal gasification
- Coal-tar pitch

- Coke production
- Engine exhaust, diesel
- Hematite mining (underground)
- Iron and steel founding
- MOPP (vincristine-prednisone-nitrogen mustard-procarbazine mixture)
- Nickel compounds
- Painting
- Plutonium
- Radon-222 and its decay products
- Rubber production industry
- Silica dust, crystalline
- Soot
- Sulfur mustard
- Tobacco smoke, secondhand
- Tobacco smoking
- X-radiation, gamma-radiation

From the IARC list with limited evidence

- Acid mists, strong inorganic
- Art glass, glass containers and pressed ware (manufacture of)
- Biomass fuel (primarily wood), indoor emissions from household combustion of
- Bitumens, occupational exposure to oxidized bitumens and their emissions during roofing
- Bitumens, occupational exposure to hard bitumens and their emissions during mastic asphalt work
- Carbon electrode manufacture
- alpha-Chlorinated toluenes and benzoyl chloride (combined exposures)
- Cobalt metal with tungsten carbide
- Creosotes
- Engine exhaust, diesel
- Frying, emissions from hightemperature
- Insecticides, non-arsenical (occupational exposures in spraying and application)
- Printing processes
- 2,3,7,8-Tetrachlorodibenzo para-dioxin
- Welding fumes

Lymphoma

- Previous blood transfusion
- History of cancer

From the IARC list with sufficient evidence

- Azathioprine
- Benzene
- Busulfan

- 1,3-Butadiene
- Chlorambucil
- Cyclophosphamide
- Cyclosporine
- Epstein-Barr virus
- Etoposide with cisplatin and bleomycin
- Fission products, including Strontium-90
- Formaldehyde
- Helicobacter pylori
- Hepatitis C virus
- Human immunodeficiency virus type 1
- Human T-cell lymphotropic virus type 1
- Kaposi sarcoma herpes virus
- Melphalan
- MOPP (vincristine-prednisone-nitrogen mustard-procarbazine mixture)
- Phosphorus-32
- Rubber production industry
- Semustine (methyl-CCNU)
- Thiotepa
- Thorium-232 and its decay products
- Tobacco smoking
- Treosulfan
- X-radiation, gamma-radiation

From the IARC list with limited evidence

- Azathioprine
- Benzene
- Busulfan
- 1,3-Butadiene
- Chlorambucil
- Cyclophosphamide
- Cyclosporine
- Epstein-Barr virus
- Etoposide with cisplatin and bleomycin
- Fission products, including Strontium-90
- Formaldehyde
- Helicobacter pylori
- Hepatitis C virus
- Human immunodeficiency virus type 1
- Human T-cell lymphotropic virus type 1
- Kaposi sarcoma herpes virus
- Melphalan
- MOPP (vincristine-prednisone-nitrogen mustard-procarbazine mixture)
- Phosphorus-32

- Rubber production industry
- Semustine (methyl-CCNU)
- Thiotepa
- Thorium-232 and its decay products
- Tobacco smoking
- Treosulfan
- X-radiation, gamma-radiation
- Bischloroethyl nitrosourea (BCNU)
- Chloramphenicol
- Ethylene oxide
- Etoposide
- Hepatitis B virus
- Magnetic fields, extremely low frequency (childhood leukemia)
- Mitoxantrone
- Nitrogen mustard
- Painting (childhood leukemia from maternal exposure)
- Petroleum refining (occupational exposures)
- Polychlorophenols or their sodium salts (combined exposures)
- Radioiodines, including Iodine-131
- Radon-222 and its decay products
- Styrene
- Teniposide
- Tetrachloroethylene
- Trichloroethylene
- 2,3,7,8-Tetrachlorodibenzopara-dioxin
- Tobacco smoking (childhood leukemia in smokers' children)
- Malaria (caused by infection with Plasmodium falciparum in holoendemic areas)

Melanoma of the skin

From the IARC list with sufficient evidence

- Solar radiation
- Ultraviolet-emitting tanning devices

Pancreatic cancer

- History of pancreatitis
- Cholecystectomy
- Number of components of metabolic syndrome (hyperglycaemia, obesity, hypertension and dyslipidaemia)
- Family history of pancreatic cancer
- Family history of any cancer
- Alcohol consumption (mL ethanol/day)

From the IARC list with sufficient evidence

- Tobacco, smokeless

- Tobacco smoking

From the IARC list with limited evidence
- Alcoholic beverages
- Thorium-232 and its decay products
- X-radiation, gamma-radiation

Prostate cancer
- Family history of prostate cancer in first-degree relative
- Height
- Finasteride
- Physical activity
- Microvascular diabetic complications
- Nephropathy
- Dyslipidemia
- Ischemic heart disease
- History of PSA test
- PSA positive before cancer diagnosis

From the IARC list with limited evidence
- Androgenic (anabolic) steroids
- Arsenic and inorganic arsenic compounds
- Cadmium and cadmium compounds
- Rubber production industry
- Thorium-232 and its decay products
- X-radiation, gamma-radiation