

A DATA-RICH WORLD

Population-based registers
in healthcare research

Ann-Britt Wiréhn



Linköping University
FACULTY OF ARTS AND SCIENCES

Linköping Studies in Arts and Science No. 404
Dissertations on Health and Society No. 10

Linköping University, Department of Medicine and Health Sciences
Linköping 2007

Linköping studies in Arts and Science • No. 404

At the Faculty of Arts and Science at Linköpings universitet, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Science. This thesis comes from tema Health and Society at the Department of Medicine and Health Sciences.

Distributed by:
Department of Medicine and Health Sciences
Linköping University
SE-581 83 Linköping

Ann-Britt Wiréhn
A data-rich world.
Population-based registers in healthcare research

©Ann-Britt Wiréhn
Department of Medicine and Health Sciences 2007

Published papers have been reprinted with the permission of the copyright holders.

Printed in Sweden by LiU-tryck, Linköping, Sweden, 2007
Cover design: Elinor Jacobsson

ISBN: 978-91-85895-96-0
ISSN: 0282-9800
ISSN: 1651-1646

~Sometimes one pays most for those things one gets for nothing~

Albert Einstein

ABSTRACT

Advances and integration of information and communication technologies into healthcare systems offer new opportunities to improve public health worldwide. In Sweden, there are already unique possibilities for epidemiological research from registers because of a long tradition of centralized data collection into population-based registers and their allowance for linkage. The growing efficiency of automated digital storage provides growing volumes of archived data that increases the potential of analyses further.

The purpose of this thesis can be divided into two parallel themes: illustrations and discussions of the use and usefulness of population-based registers on the one hand, and specific research questions in epidemiology and healthcare research on the other. The research questions are addressed in separate papers.

From the Swedish Cancer Registry, 25 years of incidence data on testicular cancer was extracted for a large cohort. Record linkage to survey data on serum cholesterol showed a highly significant positive association, suggesting that elevated serum cholesterol concentration is a risk factor for testicular cancer. Since the finding is the first of its kind and because of wide confidence intervals, further studies are needed to confirm the association.

Östergötland County council's administrative database (the Care Data Warehouse in Östergötland (CDWÖ)) provided data for prevalence estimations of four common chronic diseases.

The prevalence rate agreed very well with previous estimates for diabetes and fairly well with those for asthma. For hypertension and chronic obstructive pulmonary disease, the observed rates were lower than previous prevalence estimates. Data on several consecutive years covering all healthcare levels are needed to achieve valid prevalence estimates.

CDWÖ data was also used to analyse the impact of diabetes on the prevalence of ischemic heart disease. Women had higher diabetes/non-diabetes prevalence rate ratios across all ages. The relative gender difference remained up to the age of 65 years and thereafter decreased considerably.

The age-specific direct healthcare cost of diabetes was explored using data from the CDWÖ, the county council's Cost Per Patient database and the Swedish Prescribed Drug Register. The cost per patient and the relative magnitude of different cost components varied considerably by age, which is important to consider in the future planning of diabetes management.

The Cancer Registry was established mainly as a basis for epidemiological surveillance and research, exemplified in this thesis by a study on testicular cancer. In contrast, the newly established and planned healthcare databases in different Swedish counties are mainly for managerial purposes. As is shown in this thesis, these new databases may also be used to address problems in epidemiology and healthcare research.

CONTENTS

LIST OF PAPERS	9
ABBREVIATIONS	11
INTRODUCTION.....	13
BACKGROUND	15
National databases	15
<i>The Total Population Register.....</i>	<i>15</i>
<i>The Cause of Death Register.....</i>	<i>16</i>
<i>The Cancer Registry</i>	<i>16</i>
<i>The Swedish Prescribed Drug Register.....</i>	<i>17</i>
Administrative healthcare registers.....	17
<i>The Care Data Warehouse in Östergötland</i>	<i>18</i>
<i>The Cost Per Patient Database</i>	<i>20</i>
The Värmland–Hofors survey.....	21
Research topics and available population-based register data	22
<i>Testicular cancer.....</i>	<i>22</i>
<i>Diabetes.....</i>	<i>22</i>
<i>Hypertension.....</i>	<i>23</i>
<i>Asthma.....</i>	<i>24</i>
<i>Chronic obstructive pulmonary disease.....</i>	<i>24</i>
<i>Ischemic heart disease</i>	<i>25</i>
<i>Cholesterol.....</i>	<i>25</i>
Epidemiological concepts and methods.....	26
<i>Prevalence</i>	<i>27</i>
<i>Incidence</i>	<i>28</i>
<i>Rate ratio.....</i>	<i>28</i>
<i>Cox regression.....</i>	<i>29</i>
Health economics.....	29
AIMS.....	33

MATERIALS AND METHODS..... 35

- Paper I 36
- Paper II..... 36
- Paper III 37
- Paper IV 38
- Ethical considerations 39

RESULTS..... 41

- Paper I – Serum cholesterol and testicular cancer incidence in 45 000 men followed for 25 years..... 41
- Paper II – Estimating disease prevalence using a population-based administrative healthcare database 41
- Paper III – Age and gender differences in the impact of diabetes on the prevalence of ischemic heart disease. A population-based study 44
- Paper IV – Age-specific direct healthcare costs attributable to diabetes in a Swedish population: a register-based analysis 48

DISCUSSION 55

- Paper I 55
- Paper II and III 56
 - Completeness of individuals* 56
 - Accuracy and completeness of data* 57
 - Registration period* 59
- Paper IV 60
 - Accuracy and completeness of data* 61
- Implications and future improvements 62

CONCLUSIONS 63

SUMMARY IN SWEDISH..... 64

ACKNOWLEDGEMENTS 65

REFERENCES 67

LIST OF PAPERS

This thesis is based on following studies, referred to by Roman numerals (Papers I-IV).

Paper I

Wiréhn AB, Törnberg S, Carstensen J. Serum cholesterol and testicular cancer incidence in 45 000 men followed for 25 years. *British Journal of Cancer* (2005) 92, 1785-1786.

Paper II

Wiréhn AB, H Karlsson M, Carstensen JM. Estimating disease prevalence using a population-based administrative healthcare database. *Scandinavian Journal of Public Health* (2007) 35, 424-431.

Paper III

Wiréhn AB, Östgren CJ, Carstensen J. Age and gender differences in the diabetes impact on the prevalence of ischemic heart disease. A population-based study. *Diabetes Research and Clinical Practice*. (Accepted)

Paper IV

Wiréhn AB, Andersson A, Östgren CJ, Carstensen J. Age-specific direct health care costs attributable to diabetes in a Swedish population: a register-based analysis. (Submitted)

ABBREVIATIONS

ATC	Anatomical Therapeutic Chemical classification system
CDWÖ	Care Data Warehouse in Östergötland (<i>Vårddatalagret</i>)
COPD	Chronic obstructive pulmonary disease
CPP	Cost per patient
CVD	Cardiovascular disease
DDD	Defined Daily Doses
GP	General practitioner
HDL	High density lipoprotein
IHD	Ischemic heart disease
IPR	Ischemic heart disease prevalence rate ratio
LAH	Hospital-based homecare (<i>Lasarettansluten hemsjukvård</i>)
LDL	Low density lipoprotein
LiO	Östergötland County council (<i>Landstinget i Östergötland</i>)
MI	Myocardial infarction
PHC	Primary healthcare
SCB	Statistics Sweden (<i>Statistiska centralbyrån</i>)
SKL	Local Authorities and Regions (<i>Sveriges kommuner och landsting</i>)

INTRODUCTION

This section is a brief introduction to population-based register research in healthcare. It will also outline limitations of this method in studies of common chronic diseases often treated in primary healthcare.

Advances and integration of information and communication technologies in healthcare systems offer extraordinary opportunities to improve public health worldwide [1]. Further, a growing efficiency of automated digital storage of collected information provides exponentially growing volumes of archived data [2]. However, with these large amounts of information, which are sometimes recorded for purposes other than assessment or research, it is important to understand both the possibilities and the limitations involved.

The main purposes of epidemiological research are to estimate the prevalence of diseases and to find their causes. Additionally, an important function of epidemiological work is to identify and eliminate various sources of errors in studies, for example missing subpopulations, a problem that may affect the validity of the research results. Studies using population-based approaches, i.e. when the framework for the study population includes a well-defined population [3], from a sample or a total, are therefore highly desirable. A population-based approach is not ideal in all

situations. However, in this thesis, all analyses are population-based and here the concept has a geographical/geopolitical definition. This is the most frequently used inclusion criterion for a defined population in the sense of being population-based [4].

In Sweden, population-based data on deaths, cancers, births, congenital malformations and hospital admissions have been registered at national level in the Cause of Death Register and in various health databases for several decades. For the last two years, national prescription data have also been available in the new Swedish Prescribed Drug Register [5, 6]. As a result of the long tradition of population-based death and health data registration, there are exceptional potentials for register research in Sweden. Thus, these health databases are frequently used in epidemiological studies. Besides methodological advantages such as large populations and absence of recall bias in the use of register data, the low cost of the studies is a further major plus point.

In addition, from cradle to grave, information on almost all healthcare contacts is documented,

including administrative, health status, demographic, pharmaceutical and clinical details. Data from the most recent decade are stored in diagnosis-related administrative registers developed in the different Swedish counties. These registers are primarily set up for the billing of services, and so far their use for research in Sweden is limited, obviously due to their original purpose but probably also due to uncertainties regarding their usefulness.

Since information on outpatient care is not yet included in any national register, population-based register studies on the country as a whole are not yet feasible for chronic diseases that are usually treated in primary healthcare (PHC). Thus, so far, such studies are only possible for conditions treated in inpatient care. However, with diagnoses for each visit to a doctor

recorded in the county council administrative databases, population-based data also ought to be available for these chronic diseases.

In Sweden, the integrated healthcare system covers almost all inhabitants, and utilization is very high. Hence, all registered information in the national databases as well as in the administrative registers have made the Swedish healthcare system very rich in population-based data. Furthermore, the unique personal code system for each inhabitant makes it possible to link registers to each other or to other data sources.

This thesis focuses on the use and usefulness of existing population-based register data in healthcare and examines different types of research questions in epidemiology as well as providing cost calculations.

BACKGROUND

The aim of this section is to describe the different data materials, the chosen research topics, and the relevant concepts and methods in the thesis.

National databases

The national health data registers are frequently used for research [7]. The registers are conducted within the law relating to health data registers under the authority of the National Board of Health and Welfare, Stockholm, Sweden. This law permits use of the health data registers for the following purposes only: research, compilation of statistics, quality assurance and assessment of healthcare. The registers, including the Cause of Death Register, all have national coverage. In order of year of establishment, the registers are: the Cancer Registry (1958), the Cause of Death Register (1961), the Medical Birth Register including congenital malformation surveillance (1973), the Register of Hospital Discharges including surgical procedures requiring hospital admission (1987) and the Swedish Pre-

scribed Drug Register (2005) [6]. All the registers, except the Swedish Prescribed Drug Register, are diagnosis-related. Information in the registers is based on either observations at patient contacts with health services, hospitalization, or death. The data collection procedure is consecutive and, although the transfer date varies in the five national registers, they are updated at least annually.

In addition, large quantities of official national statistics other than the above-mentioned registers are provided by several other authorities. One such authority is Statistics Sweden (in Swedish *SCB*), which compiles data in numerous fields, e.g. on the population (the Total Population Register), migration, employment, income, education and so on.

The Total Population Register

The Swedish Tax Agency keeps registers on all residents in every county/region. Since 1968, Statistics Sweden has transferred data from these registers into a register called the Total Population Register. This

register is mainly used as a base register for preparation of statistics regarding the size and composition of the population, stratified according to sex, age, marital status, etc. in the counties and municipalities. Sta-

tistics on the population are published every month on Statistics Sweden's website [8] and also in the annual publication, *The Official Sta-*

The Cause of Death Register

The Cause of Death Register [9] comprises data from 1961 on all deaths of individuals registered as Swedish citizens at the time of death, irrespective of whether they died in Sweden or abroad. For each death, the cause should be deter-

The Cancer Registry

A directive to report newly detected cases of cancer was established in 1957, and the Cancer Registry [10] was set up the following year. In Sweden, there are six regional oncology centres where cancer reports are first coded and from where information on cancer is further transferred into the national register once a year. The duty to report cases of cancer concerns both regional and local healthcare providers, i.e. county councils, municipalities and private clinics [11]. Information on deaths from cancer is obtained from the Cause of Death Register. Migration data are ob-

tistics of Sweden. Data from the Total Population Register are fundamental to demographic research as well as in research in medicine.

mined from death certificates issued by a doctor, who is also responsible for its consignment to the local Swedish county tax agency. Besides personal identification, data include underlying and multiple causes of death and date of death.

tained from the Total Population Register to verify whether a cancer patient is still registered in Sweden, but also to view the migration among these patients within the country.

The Cancer Registry has been frequently used in research, for example in studies of associations between various factors and different types of cancer. Examples include the relationship between radon and lung cancer [12] or radon and lymphatic leukaemia [13], but also heredity studies on breast and ovarian cancer [14].

The Swedish Prescribed Drug Register

In 2005, a new national register was established – the Swedish Prescribed Drug Register [15]. This new national healthcare register contains all dispensed drug prescriptions and covers the whole Swedish population. All drugs are classified according to the Anatomical Therapeutic Chemical (ATC) classification system. Measurement units of utilization are prescriptions, Defined Daily Doses (DDDs) and expenditure. The register contains data on the following: 1) drugs

(dispensed amount per item for each patient); 2) the patient (unique identifier, age, sex, place of residence (county, municipality and parish)); 3) date (prescribing and dispensing); 4) practice (code of the primary healthcare centre or hospital clinic issuing the item); and 5) prescriber's profession (e.g. general practitioner (GP), paediatrician). Data from the Swedish Prescribed Drug Register are available from July 2005 and provide new opportunities for research.

Administrative healthcare registers

Registers that are products of the routinely collected claim and/or discharge data in administration of healthcare delivery are termed administrative healthcare registers [16-18]. From a research perspective, these registers contain already existing data and are sometimes therefore also labelled secondary data. There are 21 county councils/regions in Sweden. All have some sort of administrative register with information on inpatient care and day surgery. Some of the register data are transferred annually into the Register of Hospital Discharges [19] at the National Board of Health and Welfare, Stockholm. Several of the county administrative databases contain data on all public

healthcare in the county, i.e. including all outpatient care. Östergötland County Council (LiO) has had an administrative database of this type since 1999 – the Care Data Warehouse in Östergötland (CDWÖ). Until national registration of outpatient care is implemented, the CDWÖ and similar administrative databases run by other county councils are the only population-based registers on diagnoses for public healthcare as a whole.

The coverage of an administrative register depends partly on the availability of healthcare to all inhabitants, which is a political issue, varying between countries. In the Swedish system with a relatively low proportion of private medicine,

public healthcare covers almost all inhabitants. Hence, public healthcare users comprise a population similar to that of the country as a whole. Despite this, administrative registers are as yet rarely used for research in Sweden, while in for example Ontario, Canada, administrative databases have been used in numerous studies on various topics over the last decade [20-25]. However, in the municipality of Tierp in

Sweden, a comprehensive research database has been generated from different healthcare utilization registers. Wigertz and Westerling (2001) [26] have analysed the usefulness of these registers, concluding that a central register is needed with information on patient diagnoses across all types of healthcare to make reasonable prevalence estimates.

The Care Data Warehouse in Östergötland

The purpose of a data warehouse is to archive historical data as raw material. It is often used as a management decision support system in an organization. This methodology was developed in the late 1980s [27] and is a type of relational database, i.e. it is possible to relate all variables to each other. Usually, a database is called a data warehouse if it is *subject-oriented*, meaning that all variables are linked together with an identifier; if it is *time-variant*, i.e. data are consecutively added; if it is *non-volatile*, i.e. data are never overwritten or deleted; and if it is *integrated*, i.e. it includes most or all of an organization's operational applications. Requested raw data from the database can be aggregated and extracted to data sets that have suitable structures for various analyses. An unaddressed search for unknown patterns and/or relationships in data warehouses can be

performed using different techniques. This is called data mining.

In healthcare, administrative databases and also other more disease-specific registers are nowadays often constructed as data warehouses [28-31]. In 1997, the IT council at LiO initiated a project to set up a healthcare database. Subsequently, in the three-year budget from 1997 to 1999, the county council decided to develop a data warehouse – the CDWÖ – in which data on all healthcare provision in the county council would be consecutively registered. Registration of information from the three hospitals in the county started in 1998, while data from the 42 primary healthcare (PHC) centres began in 1999. In an agreement with LiO's finance department, all healthcare production units (PUs) (n~15) were included and were responsible for data transfer to CDWÖ from all their health

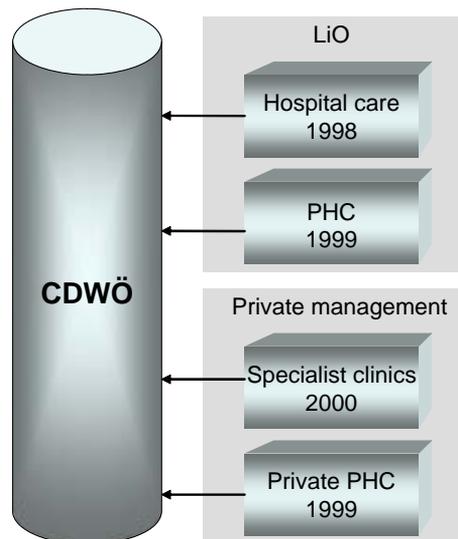
care providing subsets, i.e. the base units (BUs) ($n \sim 100$). In 2000, two private specialist clinics were added to the database. Hospital-based homecare in Östergötland (in Swedish *LAH*) is run by the municipalities in collaboration with the LiO. Data from these units were included in the CDWÖ from 2004. The types of units and the first year of data transfer into the CDWÖ are given in *Figure 1*.

The information compiled in the CDWÖ covers aspects such as administrative data on the patient and on the visit or hospitalization. For all visits to a doctor and all hospitalizations, it is possible to record

the main diagnosis and up to 10 secondary diagnoses for hospital care, while in the PHC up to 10 un-ranked diagnoses are possible. Diagnoses are recorded according to the International Classification of Diseases, 10th version (ICD-10) [32]. Data are transferred once a month from all the BUs and private clinics. However, and importantly, in the CDWÖ data are sometimes overwritten to correct errors and to fill in information as it becomes known.

The CDWÖ is Östergötland County Council's provider of information to the Register of Hospital Discharges.

Figure 1. Type of units routinely transferring information on a monthly basis to the CDWÖ database, plus the first year of transfer



The Cost Per Patient Database

This database is not an ordinary administrative register. In practice, it is often linked to CDWÖ data and for this reason is described in this section.

In 1999–2002, the Swedish Association of Local Authorities and Regions (in Swedish *SKL*) initiated a national cost per patient (CPP) project to improve efficiency in healthcare. The aim of the project was to develop a useful tool to improve the foundation upon which allocation of resources to healthcare is based. The basis of the calculations is a four-stage process: 1) identification of the relevant healthcare cost, 2) identification and distribution of the costs for joint activities, 3) calculation and description of the healthcare services, 4) linking consumption of different services to separate healthcare contacts. The main task of the project was to create national principles and models for CPP accounts in all sectors of the health service by developing a cost account based on individual data. The essence of CPP calculations is to determine prices for different activities and resources carried out or used in different types of clinics, in

order to describe costs as specifically as possible. The CPP project produced cost account proposals for somatic care, psychiatric care and PHC [33].

Over the last few years, Östergötland County Council's finance department has made efforts to follow and further develop the national CPP principles, with the result that a CPP database is now available. This includes costs for each healthcare contact or each patient that has contacted health services in LiO from 2005. Standard costs have been calculated for all healthcare services, e.g. a visit to a doctor or a laboratory test, based on unique information for each clinic. Thus, it is for example possible to summarize the CPP for healthcare in different clinics or for each individual, over a certain period of time. Furthermore, other costs, i.e. costs not attributed to specific healthcare contacts, are distributed across the individual CPPs.

The Värmland–Hofors survey

The revolutionary development of clinical-chemical laboratory technology in the 1950s made it possible to carry out blood analyses with a high-degree of automation. Hence, blood analyses could be made more quickly and at much lower costs than previously. Therefore, in 1961, the Swedish National Board of Health proposed a mass screening health survey of all residents in the county of Värmland. All Värmland residents over 25 years of age were to be included in the survey. The total number of residents invited to participate was about 117 000, with about 90 000 actually attending. The primary reason why the Board of Health suggested Värmland was the county council's procurement of a mobile X-ray unit and well-equipped general hospitals. Non-medical staff could perform the

survey and this was seen as a great advantage because of the shortage of available doctors. The aim of the survey was to detect early-stage diseases by health checks based on chemical analyses of the blood before any subjective symptoms had appeared, and these blood tests were to be combined with other measurements. The health check included a number of urine and blood analyses including serum cholesterol. Other measurements in the study were height, weight and blood pressure plus a chest X-ray. A questionnaire was also used to take histories of previous disorders. Because of the large amount of material, an additional envisaged gain of the survey was to evaluate normal values and ranges of distribution of chemical tests [34].

Research topics and available population-based register data

Six diagnoses and one measured laboratory test are included in this thesis and are briefly presented below. They were chosen in order to evaluate the usefulness of the various registers for research questions that have not previously been studied in detail using population-based register data in Sweden. The availability of population-based register data covering all health care levels is described for each topic.

Testicular cancer

The most common cancer in men aged 15–44 years is testicular [35]. Although still relatively low (6.8 per 100 000 in 2004), the incidence of testicular cancer has increased in recent decades [36]. There are two main types of testicular cancer, seminoma and non-seminoma. These two types grow differently and are treated in different ways. However, after treatment, the prognosis is favourable for both types. It is unclear what causes testicular cancer, but it has been suggested that predisposition to testicular cancer is present from an early age,

probably *in utero* [37], and well-established associated factors are non-descended testes at birth and diet, e.g. a high intake of fat [35, 38, 39]. Furthermore, the mechanism causing a deteriorating trend in male reproductive health is also suggested to involve testicular cancer [40].

Almost half a century of population-based national register data are available on testicular cancer incidence and principally all epidemiological studies on cancer are through data from the Cancer Registry.

Diabetes

Diabetes develops when the insulin hormone does not adequately regulate the levels of blood glucose. When the disturbance is due to destruction of the insulin-producing cells in the pancreas, the condition is commonly referred to as type 1 diabetes. However, the predominant form, type 2 diabetes, is caused

by a combination of inadequate insulin production and an acquired inability to effectively use the produced insulin, referred to as insulin resistance.

Type 2 diabetes usually affects middle-aged and elderly people. The prevalence of known diabetes is about 5% in developed countries

[41, 42], and is in fact one of the fastest growing public health problems. Based only on demographic changes, i.e. increasing life expectancy, and assuming that age-specific diabetes prevalence remains constant, the prevalence of diabetes is expected to approximately double within the next two decades [43]. The total healthcare resources used by patients with diabetes is substantial, and the estimated medical cost of type 2 diabetes in eight European countries was €29 billion (1999 values), giving an average yearly cost per patient in the diabetes population of €2834 [44].

Hypertension

Hypertension is defined as chronically elevated blood pressure, measured as systolic/diastolic blood pressure. The systolic pressure is measured when the heart contracts to pump out the blood, and the diastolic when the heart relaxes and fills with blood. In LiO, hypertension is defined as blood pressure \geq 140/90 mmHg, following WHO guidelines [47]. Obesity, smoking and salt intake are examples of common factors that affect blood pressure but the disease also has high heritability.

In a screening program for hypertension, the hypertension prevalence was estimated to be 25% according to the criterion above or

No national population-based register data are available on diabetes, and thus epidemiological measures are often analysed from screening studies or questionnaires. There is a national healthcare quality register for diabetes, the National Diabetes Register (NDR), which in 2006 included about 130 000 registered diabetes patients [45]. However, some more complete regional registers are available, suitable for epidemiological analyses, for example including one for the municipality of Tierp [26] and another for the county of Skaraborg, called the Skaraborg Diabetes Registry [46].

current use of antihypertensive medication [48]. Another study on the screened hypertension prevalence, using the same criteria, included six European countries (Germany, Finland, Sweden, England, Spain, Italy), Canada and the US. The hypertension prevalence for persons aged 35 to 64 years was on average 44% in the European countries and 28% in North America [49].

At present, no population-based register data on blood pressure or hypertension are available.

Asthma

Asthma is a chronic inflammatory disorder in the bronchial tubes, resulting in obstruction of the airways which is most often reversible, spontaneously or after treatment. The condition is characterized by dyspnoea, cough and wheeze and can normally be effectively controlled and treated. Asthma attacks (or exacerbations) are episodic, but the airway inflammation is chronically present. Asthma is principally caused by allergy. Patients with suspected asthma are examined by spirometry, a test of lung function.

Worldwide data show that the prevalence of asthma has increased

over the past decades [50]. Among children, asthma is the most common chronic disease, and the prevalence of physician diagnosis of asthma ever is estimated to be about 9% in Sweden in the age group 11-12 years, significantly higher among boys [51]. To develop a network and to incorporate results of scientific investigations into asthma care, a Global Initiative for Asthma (GINA) was implemented [52].

At present, no population-based register data on asthma are available.

Chronic obstructive pulmonary disease

Patients with chronic obstructive pulmonary disease (COPD) have damaged, inflexible lungs resulting in chronic obstruction of the airways. The symptoms are similar to those of asthma, e.g. dyspnoea, cough and wheeze. Spirometry is the evaluation method also for COPD, complemented by a test of reversibility [53]. COPD is often classified according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [54] for dif-

ferent grades of severity. Despite the similarities between the diseases, the cause of COPD differs from that of asthma: COPD is principally caused by smoking or in some cases other environmental exposures. It has been suggested that about 50% of all smokers will develop COPD [55].

At present, no population-based register data on COPD are available.

Ischemic heart disease

A lack of oxygen in the heart muscle because of reduced blood supply is designated as ischemic heart disease (IHD) caused by narrowed or blocked coronary arteries due to atherosclerosis. IHD includes myocardial infarction (MI) and angina pectoris. The diagnostic criteria for MI are based on the consensus document of the European Society of Cardiology and the American College of Cardiology [56]. IHD mortality rates have decreased in most industrialized countries in recent decades [57]. Nevertheless, it remains the most common cause of death in these countries in both men and women.

Women are generally at much lower risk of IHD than men. However, the relative risk of IHD in people with diabetes, compared to subjects without diabetes, is higher in women than in men [58-61]. In a meta-analysis, it was found that the relative risk of fatal coronary heart disease associated with diabetes

was about 50% higher in women than in men [62], and in another meta-analysis diabetes conferred an equivalent IHD risk of ageing 15 years [63].

In 1996, a record linkage was set up between the Hospital Discharge Register and the Cause of Death Register, giving the Swedish statistics of Acute Myocardial Infarction [64]. National data on patients with acute MI cared for as inpatients is thereby available. These linked statistics have no data on diabetes.

However, the healthcare quality register, the Register of Information and Knowledge about Swedish Heart Intensive care Admissions (RIKS-HIA) [65], covers almost all patients. Assuming that *all* patients with MI are treated in cardiac intensive care, population-based data on MI is available here. For these patients, diabetes data are also available. For angina pectoris, however, there are no national population-based data.

Cholesterol

Cholesterol is a fatty lipid found in the body tissues and blood plasma. It comes either from the body's own production, mainly in the liver, or by food intake. Cholesterol is an important building block of the body's cell membranes. It travels in the blood plasma to cells, carried by

the proatherogenic low density lipoprotein (LDL), and back to the liver for secretion by antiatherogenic high density lipoprotein (HDL). Elevated proatherogenic lipoproteins in the blood is defined as hyperlipidaemia, which over time can predispose to atherosclero-

sis, i.e. accumulated fat in the walls of the arteries. Hyperlipidaemia is a strong risk factor for atherosclerotic cardiovascular disease (CVD). To prevent CVD, generally a serum cholesterol level of < 5.0 mmol/L is

recommended by the Swedish Medical Products Agency [66]. There are no national register data on hyperlipidaemia or blood cholesterol.

Epidemiological concepts and methods

The use of epidemiological concepts sometimes differs between epidemiologists. The intention in this section is both to describe the use of the epidemiological concepts in this thesis in particular and to describe epidemiological methods used generally.

In analyses of binary outcomes, i.e. when a variable has two alternatives for each individual (e.g. diseased or not diseased), proportion (p), rate and ratio are the mathematical applications often used. Differences in the meanings of the concepts, from an epidemiological point of view, depend on the substance of the quantities. A proportion always includes the numerator in the denominator and is thus the fraction in the decimal range $0.0 \leq p \leq 1.0$. A rate has several uses in epidemiology but is commonly restricted to the frequency with which an event occurs in a defined population, i.e. the number of events in a specified period divided by the average population in the same period. However, in this usage, prevalence rate would not be a true rate since it is a synonym of

proportion. Ratio is usually an expression of the relationship between two distinct quantities, neither being included in the other. There are, however, exceptions and sometimes ratios are expressed as percentages, e.g. standardized mortality ratio. Rates and ratios are possible for all values ≥ 0.0 [67, 68].

The outcomes diseased or not diseased can be related to a categorical exposure variable and cross-tabulated in a contingency table of size $2 \times c$, where c describes the number of categories in the exposure variable. Thus, when the exposure variable is also binary, the construction is a 2×2 table. Provided that all individuals are independent and the probability of being diseased is constant within each group, this is a simple but useful comparison (Table 1) [67, 69, 70].

Table 1. Example of a contingency table of size 2×2

Exposure	Outcome		
	Diseased	Not diseased	Total
Exposed	d_1	h_1	n_1
Unexposed	d_0	h_0	n_0
Total	d^*	h	n^*

d^* = in prevalence studies – all diseased subjects, in incidence studies – newly diseased subjects

n^* = in prevalence studies – the total population, in incidence studies – initial population at risk or sum of person-time at risk

The proportion of diseased subjects in Table 1 is $d/n = p$ and the sampling distribution for p is binominal. For large sample sizes binomially distributed variables become normally distributed. This gives an option to make calculations of confidence intervals (CIs) and hypothesis testing sufficiently well approximated to the normal distribution. This is very useful since using the binominal

distribution to derive CIs is complicated [67]. However, this approximation may be insufficiently sophisticated for very small proportions in large populations, generating lower CI boundaries with negative values. To avoid this, as an alternative, approximation of binomial tails via the F-distribution, is a more exact method to use [70].

Prevalence

Cross-sectional study design is a sufficient approach when deriving prevalence. Prevalence refers to all subjects with a disease in a defined population (d in Table 1) either at a point in time (point prevalence) or over a particular period (period prevalence), divided by the population at risk of having the disease (n in Table 1). When calculating the period prevalence, it may be difficult to define the most appropriate denominator. However, point prevalence is the most used and the

prevalence term without qualification usually refers to point prevalence [4]. It is also common to describe the prevalence as the number per 100 000 persons for example.

Knowledge of the case definition and the population definition is essential in interpretations of the estimated prevalence. Therefore, before comparing different estimations, it is important to find out these definitions and the research method. Surveys with self-reported prevalence give the survey partici-

pants' opinions of the prevalence and clinical examinations give the doctors' opinions of the prevalence, while screening studies give the prevalence of both known and unknown cases. Although the concept of prevalence is principally a description of the morbidity in a population, it is possible that deaths

also come under the inclusion criterion for prevalence. This may be the case when a subject dies in the first incidence of the condition. Death from other diseases and migration also affect the prevalence and should be considered when interpreting estimates.

Incidence

Incidence is a measure of the newly detected subjects with a disease or death from a disease (d , in Table 1) over a given period, in relation either to the sum of person-time (often person-years) at risk for the disease or the initial population at risk (n , in Table 1) for the same period. The incidence rate is given by the formula d/n , where n is the sum of person-years at risk, i.e. the total sum of every study member's contributed year in the study, or as the number per 100 000 person-years

for example. Incidence rates can also be obtained by survival analyses, a method of predicting an outcome at any point of time for individuals with a given condition. The outcomes from these analyses are mostly termed hazard rates. When n is the total population rather than sum of person-years, the outcome may be called average risk, cumulative incidence, incidence proportion or attack rate, depending on the basis and assumptions. [71].

Rate ratio

A ratio of two rates is called a rate ratio; it gives a measure of the relative difference between the rates (prevalence or incidence). Relative ratio, relative risk, odds ratio and hazard ratio are examples of rate ratios. In longitudinal incidence studies, the rate ratio is called a relative risk and, with a survival analysis applied, it is termed a hazard ratio (HR). The same calculation

in a cross-sectional study describes the relative difference between two prevalences and can thus be termed prevalence rate ratio [67, 71].

As for CIs for single proportions, described above, the calculation of 95% CI for rate ratios may also generate negative values. This occurs when the standard error is large and the rate ratio is close to zero. To overcome this problem, the

logarithm of the ratio and its standard error can be used in the so-

called delta method in CI calculations [67].

Cox regression

Survival analyses are used in longitudinal studies aimed at estimating the risk for a disease or death at any point in time for individuals with a given condition. The Cox proportional hazard regression model is a popular method in multivariate analyses of survival data and is used to explain the effect of the survival time among several possible explaining variables [72]. Data collected over a certain time period sometimes become incomplete since people disappear from a study for

reasons other than the observed one. The strength of the Cox model is that the disappeared individuals contribute with information as long as they are present in the data material.

The Cox regression model is designated as half-parametric since there is no claim for a particular probability distribution on the survival time. The outcome measure of the analysis is often expressed as a variable's effect on the relative hazard or the HR.

Health economics

This section describes the general content of a cost-of-illness study, followed by a focus on the direct medical costs attributable to a disease.

In developing strategies of resource allocation, prioritization and prevention policies in the healthcare sector, it is essential for decision makers to have accurate, research-based information on for example cost-effectiveness of treatments and the burden of diseases [73-75].

Several chronic diseases show increasing prevalences, to some extent because of changes in the stipulated norms for being diseased, but mostly due to today's widespread lifestyle-related health problems

and increasing life expectancy in the population [43]. The economic burden due to poor health consists of both individual suffering and financial strain on healthcare systems as well as costs due to changes in productivity.

Health economic studies can be conducted from different perspectives; e.g. the patient's, the health care provider's or the society. The chosen perspective decides cost data of primary interest [76].

Estimates of the costs of diseases have a descriptive approach and are termed cost-of-illness (COI) studies. A COI study includes all the costs involved in a disease, i.e. direct medical costs (related to healthcare), direct non-medical costs (unrelated to healthcare, e.g. unpaid assistance and care by relatives), indirect costs (productivity losses because of illness, e.g. sick leave), and intangible costs (consequences of the illness, e.g. psychosocial suffering) [77]. Since COI studies have no built-in comparative approach, they are not intended to be used as guidance for improving healthcare efficiency [78]. However, with several repeated studies it is possible to compare the cost development of a disease for different periods. Furthermore, stratified by different cost items and subpopulations, COI studies may provide important information identifying those items/subpopulations that are exceptionally cost driving.

A study of direct costs address the quantities of resources used to treat a disease. COI studies and studies of direct costs can be distinguished in several ways. The study design can be either incidence-based or prevalence-based. In incidence-based cost studies, lifetime costs of the disease can be provided for newly diagnosed patients and interventions can be assessed by calculating the economic benefits of

reducing new cases [79]. The most common type is, however, prevalence-based. This approach examines the cost in a given year associated with all those with a disease.

In addition, studies can be divided by the costs of care for people with a disease or by the cost attributable to a disease. The attributable (or additional) cost of a disease is the cost for the diseased patients exceeding the level that would be expected if this population did not have the disease. For the costs attributable to a disease, data on both the diseased and non-diseased are necessary.

Another distinguishing factor is whether the cost study is disease-specific (including all costs for a certain disease) or general (examining all relevant costs for all disease categories) [80]. The estimation procedure may also vary in that it may have either a top-down or a bottom-up perspective [79, 81, 82]. The top-down approach is based on aggregated national cost data, stratified by disease. This approach can be perceived as conservative since only the costs related to the main diagnosis usually are available on this level. In the bottom-up approach, costs are calculated on an individual basis and usually all costs related to the patient are included.

The cost study in this thesis (Paper IV) includes estimations on the direct medical costs attributable

to diabetes, thus, a disease-specific type of study. The design is based on prevalence and the procedure is bottom-up. All estimations are adjusted to the age and gender distribution of the diseased population. The additional cost per person was derived by *Formula 1* and the additional total cost by *Formula 2* (mone-

tary values). The percentage share of the costs related to the total healthcare cost was calculated by *Formula 3*. Health cost ratios describing the relative difference between diseased and not diseased were calculated according to *Formula 4*.

Additional cost per person =

mean cost in the diseased population – mean cost in the non-diseased population

Formula 1

Additional total cost =

number of diseased x additional cost per person

Formula 2

Cost share in percent of the total healthcare cost =

100 x additional total cost/ total healthcare costs

Formula 3

Cost ratio =

mean cost in the diseased population / mean cost in the non-diseased population

Formula 4

AIMS

The purpose of this thesis can be divided into two parallel themes: illustrations and discussions of the use and usefulness of population-based registers on the one hand, and specific research questions in epidemiology and healthcare research on the other. The specific research questions include estimates of the diagnosis-specific disease burden as prevalence (Paper II) and the economic burden (Paper IV). They also include analyses of associations between diseases and potential risk factors (Papers I and III).

The specific aims were:

- To describe the relationship between serum cholesterol and testicular cancer (Paper I)
- To estimate prevalences of a number of common chronic diseases often treated in primary healthcare: diabetes, hypertension, asthma and COPD (Paper II)
- To estimate age and gender differences in diabetes and IHD comorbidity (Paper III)
- To explore the age-specific direct medical costs attributable to diabetes (Paper IV)

MATERIALS AND METHODS

The study-specific materials and methods are presented below.

Although the studies in this thesis have different focuses, they all concern the use of large existing population-based databases in healthcare research. Record linkage to other registers or other data sources was necessary in some form in all four studies to constitute suitable data materials (Table 2). However, ag-

gregated population data from the Total Population Register were obtained from the Statistics Sweden website [8].

The methods used in the four studies were survival analyses (Paper I), prevalence estimations (Papers II and III) and cost estimations (Paper IV).

Table 2. Study overview

Paper	Data source (period of data collection)	Studied outcome	Study population n Sex ^a Age
I	Cancer Registry (1958–87) Cause of Death Register (1963–87) Värmland–Hofors Survey (1963–65)	Incidence of testicular cancer	44 864 M 17-74
II	CDWÖ^b (1999–2003) Cause of Death Register (1999–2003) Total Population Register (Dec 2003)	Prevalence of diabetes, hypertension, asthma, COPD ^c	70 766 M & F All ages
III	CDWÖ (1999-03) Cause of Death Register (1999–2003) Total Population Register (Dec 2003)	Prevalence of diabetes & IHD ^d	141 400 M & F 45-74
IV	CPP^e database (2005) Swedish Prescribed Drug Register (Jul–Dec 2005) CDWÖ (1999–2005) Cause of Death Register (1999–2004) Total Population Register (Dec 2004)	Direct health-care costs of diabetes	415 990 M & F All ages

^a: M=male, F=female, ^b: Care Data Warehouse in Östergötland, ^c: chronic obstructive pulmonary disease, ^d: ischemic heart disease. ^e: cost per patient. **Bold text** = outcome data

Paper I

The Värmland-Hofors survey data from 1963-1965 were matched with death and cancer data from 1958 to 1987 from the Cause of Death Register and the Cancer Registry giving a 25-year period of follow-up.

The obtained data from the cohort was the measure of serum cholesterol for all men aged 17-74 years at risk of testicular cancer. Subjects with reported cancer (at any site) before they were examined within the survey were excluded from the study population. The Cox proportional hazard model was used for

statistical analysis [72]. The follow-up time variable was the month of the serum cholesterol test to the testicular cancer event. In the analysis, serum cholesterol was classified into three categories: <5.7 , $5.7-6.9$ and ≥ 7.0 mmol l⁻¹. The cholesterol categories were treated as an indicator variable with the lowest category as a reference group. The regression model was adjusted by age. A χ^2 -test for trend was also included in the analysis [67].

Paper II

A case-finding algorithm searched retrospectively in the CDWÖ for diagnoses in a five-year period starting on 31 December 2003. The algorithm captured the cases (one case = one patient) regardless of whether the disorders of interest constituted the main or secondary diagnosis; it also specified the healthcare level at which the patient was diagnosed, i.e. PHC, outpatient hospital care, and/or inpatient hospital care.

The extracted variables from the CDWÖ are presented in *Table 3*. The following case definition was applied: at least one contact with healthcare services with a relevant diagnosis during the period 1 Janu-

ary 1999 to 31 December 2003. The ICD-10 codes for the selected disorders were E10-14 (diabetes), I10-I13 and I15 (hypertension), J45 (asthma), and J44 (COPD). Dates of deaths were obtained from the Cause of Death Register.

Information on the number of residents in Östergötland County was obtained from the Total Population Register. The numerator of a prevalence rate was the number of cases (excluding deaths) identified in the five-year study period; the denominator was the population of the county on 1 January 2004. The prevalence rates are presented as proportions with 95% CIs [70]. Gender-specific rates were given as

totals and for the following age groups: 0–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, and over 85 years. Cumulative saturation of case findings was calculated as the proportion of the captured number of cases in different time frames (i.e. one, two, three, and four years) relative to the number of cases during the entire five-year period. The case

finding per healthcare level was given as the proportion of patients diagnosed with a particular disorder on the level in question, irrespective of registrations on other healthcare levels. The two-tailed z-test was used to analyse differences in proportions at a significance level of 5% [83].

Table 3. Variables in the Care Data Warehouse in Östergötland (CDWÖ) used for the analyses in this thesis

Item	Item declaration
Entry type	visit ^a /stay ^b
Personal code number	yyyyymmddxxxx
Gender	
Diagnosis	ICD code/s (main and secondary ^c /up to 10 un-ranked ^d)
Domicile	county, municipality, parish
Healthcare organization	performing healthcare level, clinic, unit, section
Healthcare staff category ^a	physician/others ^e
Date and time	visit ^a /admission to hospital ^b , discharge from hospital ^b

^a = outpatient only, ^b = inpatient only, ^c = in- and outpatient at hospital, ^d = primary healthcare, ^e = assistant nurse, audiometrics, nurse, occupational therapist, ophthalmologist, psychologist, speech therapist, physiotherapist or welfare officer

Paper III

The extracted diabetes data from the CDWÖ for this study were similar to those in Paper II, i.e. five-year data capturing all subjects with diagnosed diabetes alive on 31 of December 2003. However, in this study a search was also performed for IHD (ICD-10 codes I20-I25). Prevalence rates for diabetes and IHD were estimated separately, as were the IHD prevalences for sub-

jects with and without diabetes. The rates were estimated both as totals and by gender in the age groups: 45-54, 55-64, and 65-74 years. IHD prevalence rate ratios (IPRs) were calculated as the prevalence of IHD in diabetic versus non-diabetic subjects, and the ratios of female IPRs to male IPRs were also computed. The IPRs and the gender ratios of the IPRs were given with 95% CIs

[67]. P-values for the trend over age in the IPRs and the gender ratios of the IPRs were derived from a log

linear regression model in which the age groups were scored 1, 2 and 3 respectively [67].

Paper IV

This population-based register study used a cross-sectional design with estimates of the diabetes prevalence on 31 December 2004 and the direct medical costs of diabetes (type 1 and type 2) in Östergötland County in 2005. The cost estimates were based on the prevalence and had a bottom-up approach. The study included all inhabitants alive on 1 January 2005.

The diabetic patients were obtained by using the case-finding algorithm to search a six-year period in the CDWÖ database retrospectively from 31 December 2004 to 1 January 1999. Deaths before 1 January 2005 were identified from the Cause of Death Register and omitted from the study. Thus the remaining patients, including diabetic patients diagnosed in 2005, comprised the diabetes population in the cost study. The non-diabetes population was obtained by subtraction of the diabetes population from the total population. Information on the number of residents in Östergötland County was obtained from the Total Population Registry.

Data on costs for healthcare resources consumed in 2005 were ob-

tained from the CPP database for the total population as well as for each diabetes patient. The resources were specified in hospitalizations and different outpatient contacts, for example visit to a GP, another specialist, or contacts with other staff.

Data on total drugs sold during the period July–December 2005 were obtained from the Swedish Prescribed Drug Register, both for the total population and for each diabetes patient. Drug data were specified in ATC groups A10A (insulin), A10B (oral antidiabetic agents), and in test strips for self-monitoring glycaemic control. All cost estimates are given as €2005.

The diabetes prevalence was estimated as the ratio of the number of identified patients alive on 31 December 2004 to the total population at the same point in time. The prevalence was given as a total and in five-year age groups.

The additional cost due to diabetes was defined as the difference in cost between patients with diabetes and the non-diabetic population. The additional cost per patient was the difference in average costs per

year between the populations (*Formula 1*), while the additional total cost was the additional cost per patient multiplied by the number of patients with diabetes (*Formula 2*). Cost estimates per person for subjects with and without diabetes and for the additional costs were given for all ages (age- and sex-standardized rates to five-year age groups of the diabetes population in Östergötland on 31 December 2004) and specifically for the following age groups: 0-14, 15-44, 45-54, 55-64, 65-74, and over 75 years. Percentage shares of the total healthcare cost

(*Formula 3*) as well as a health cost ratio (*Formula 4*) (defined as the diabetes versus non-diabetes ratio) were given for all ages and for each age group. CIs of 95% were computed via the delta method for the health cost ratios [67] and for the cost per person and year in the diabetes population.

Since cost data represented only six months, the drug cost estimates were doubled. Estimates for the non-diabetes population were calculated by subtracting the diabetes data from the total population data.

Ethical considerations

As the research in this thesis was conducted using several linked data sources, consent by ethics committee was required for all the studies. Therefore, each study (Papers I-IV) was approved by The Local Ethics Committee at Linköping

University: Paper I: Dnr 03 582 and The Regional Ethical Review Board of Linköping University: Paper II: Dnr M53-04, Paper III: Dnr M53-04 amendment 30-05, and Paper IV: Dnr M53-04 amendment 8-06.

RESULTS

This section presents the study-specific results of the thesis.

Paper I – Serum cholesterol and testicular cancer incidence in 45 000 men followed for 25 years

Twenty-four cases of testicular cancer were found during the follow-up period among the 44 867 men aged 17-74 years at risk in the cohort. The cases that occurred within two years of the start of the period were excluded to avoid effects of inverse causality (n=3). Thus, the remaining number of cases was 21. A positive correlation between serum cholesterol level and testicular

cancer incidence was found and the estimated HRs for the middle and highest serum cholesterol categories compared to the lowest were 1.3 (95% CI: 0.3 to 5.1) and 4.5 (95% CI: 1.3 to 16.2) respectively. There was a significant trend for increased testicular cancer incidence for increased serum cholesterol categories ($p = 0.005$) (Table 4).

Table 4. Testicular cancer incidence in relation to serum cholesterol in a 25-year follow-up study of the Värmland cohort (n = 44 864)

Serum cholesterol categories mmol/L	No. of cases	Hazard Ratio	95% Confidence Interval
< 5.7 ^a	3	1.0	–
5.7 – 6.9	7	1.3	0.3 – 5.1
≥ 7.0	11	4.5	1.3 – 16.2

P-value for the trend = 0.005

^a: Reference category

Paper II – Estimating disease prevalence using a population-based administrative healthcare database

The study included 70 766 patients in total and the overall prevalence rates were 4.4%, 10.3%, 4.5% and 1.2% for diabetes, hypertension, asthma and COPD respectively (gender-specific prevalences are

presented in Table 5). The prevalences were higher for men than for women for diabetes, but higher for women for hypertension, asthma and COPD.

Table 5. Gender-specific prevalence rates and 95% confidence intervals (CIs) of four diagnoses in Östergötland County, Sweden, on 31 December 2003

	Men			Women		
	No. of cases	Rate	95% CI	No. of cases	Rate	95% CI
Diabetes	9593	4.6	(4.5 – 4.7)	8541	4.1	(4.0 – 4.2)
Hypertension	18 814	9.1	(9.0 – 9.2)	23982	11.5	(11.4 – 11.6)
Asthma	8580	4.1	(4.1 – 4.2)	9871	4.8	(4.7– 4.9)
COPD	2277	1.1	(1.05 – 1.14)	2535	1.22	(1.17– 1.27)

The prevalence of diabetes increased up to ages 75–84 years for both sexes, and the rates were higher among men in the age group 35–84 years. The prevalence of hypertension also peaked in subjects 75–84 years of age, and was higher among women in the age groups 25–34, 35–44, 65–74, 75–84, and > 85 years. No gender differences in hypertension were detected in those 45–54 years of age, whereas men had a higher prevalence in the age group 55–64 years. The age curve for the prevalence of asthma was bimodal for both sexes, and there were peaks among children and those aged 65–74 years. Asthma was considerably more common in boys than in girls, although the re-

verse was the case for ages 15–24 years and older, and this dominance in females remained until the age of 84 years. The COPD prevalence rates were higher for women up to 64 years, whereas in older age groups there was a marked increase among men, while rates for women stabilized and decreased.

The healthcare level-specific capture of cases was highest for PHC: 75% for diabetes, 86% for hypertension, 66% for asthma and 67% for COPD. Further, examining the respective diagnoses captured solely from PHC data gave proportions of 23%, 68%, 53% and 48% (these gender-specific proportions are presented in *Figure 2*).

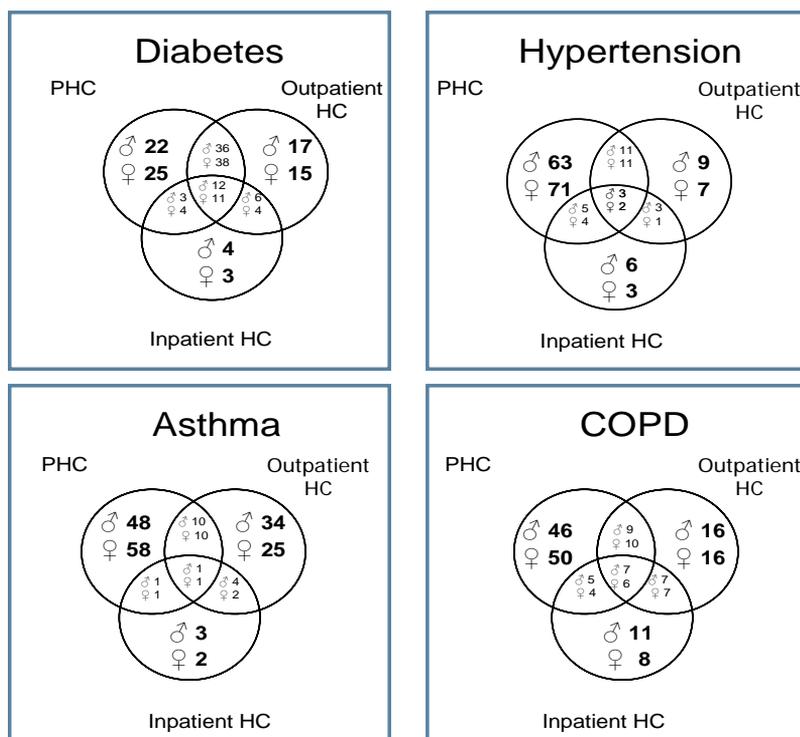


Figure 2. Proportions of captured cases by gender (♂= men and ♀=women) and healthcare level: primary healthcare (PHC), outpatient hospital care (HC) and inpatient hospital care in a five-year period for the residents of Östergötland County, Sweden, on 31 December 2003 (diabetes, $n = 18\,134$; hypertension, $n = 42\,796$; asthma, $n = 18\,451$; chronic obstructive pulmonary disease (COPD), $n = 4812$).

In addition, the proportions of cases retrieved solely from PHC data were larger for women than for men, whereas the opposite was true (i.e. larger proportions for men) for cases captured solely from inpatient hospital care and outpatient hospital care data, with the exception of COPD in outpatient hospital care and asthma in inpatient hospital care. Furthermore, considering data originating solely from either PHC or inpatient hospital care, a larger

proportion of cases was captured for the population ≥ 65 years than for those < 65 years.

Using a *one-year* case-finding procedure (i.e. applying the algorithm to search data from 2003 only) to give the rate of case capture in relation to the total five-year period resulted in values of 71%, 50%, 38% and 58% for diabetes, hypertension, asthma and COPD respectively. To achieve 95% saturation of found cases, the case-finding algorithm

had to be used to search a period of three years for diabetes and four years for hypertension and COPD. For asthma, the algorithm captured 89% of the total cases over four

years. There were only slight age and gender differences in the cumulative case-finding saturation (*Table 6*).

Table 6. Proportions of annual cumulative case findings in the CDWÖ database over the period 1999–2003 among inhabitants in Östergötland on 31 December 2003

Year	Cumulative case findings, %							
	Diabetes		Hypertension		Asthma		COPD	
	M ^a n=9593	F n=8541	M n=18 814	F n=23 982	M n=8580	F n=9871	M n=2277	F n=2535
2003	73	70	51	49	38	38	57	59
02-03	90	88	75	73	59	59	78	79
01-03	95	95	88	87	74	75	89	90
00-03	98	98	95	94	88	89	95	96
99-03	100	100	100	100	100	100	100	100

^a: M = male, F = female

Paper III – Age and gender differences in the impact of diabetes on the prevalence of ischemic heart disease. A population-based study

In the population aged 45-74 years in Östergötland County on 31 December 2003, 6044 (8.6%) men and 4320 (6.1%) women were diagnosed with diabetes mellitus, and there were 7504 (10.7%) and 3807 (5.3%) prevalent cases of IHD among men and women respectively (*Table 7*).

Among all identified diabetic patients, 77% were registered in PHC; the corresponding figure for all identified IHD patients was 55%.

When considering PHC solely, these proportions were 24% and 26% for diabetes and IHD respectively. In each age group, men had a higher prevalence of IHD than women, both among diabetic and non-diabetic subjects. However, the IHD prevalence rate was higher among diabetic women than among non-diabetic men in each age group (*Table 8*).

Table 7. Prevalence of diabetes and ischemic heart disease (IHD) by age and sex in the county of Östergötland, Sweden, on 31 December 2003

Age	Number of persons in the population	Diabetes		IHD		
		No. of cases	Rate, %	No. of cases	Rate, %	
Men						
45–54	27 058	1123	4.2	940	3.5	
55–64	26 642	2517	9.4	2884	10.8	
65–74	16 429	2404	14.6	3680	22.4	
45–74	70 129	6044	8.6	7504	10.7	
Women						
45–54	26 549	754	2.8	403	1.5	
55–64	26 429	1506	5.7	1252	4.7	
65–74	18 293	2060	11.3	2152	11.8	
45–74	71 271	4320	6.1	3807	5.3	

Table 8. Prevalence of ischemic heart disease (IHD) among subjects with and without diabetes by age and sex, in Östergötland County, Sweden, on 31 December 2003

Age	IHD in subjects with diabetes			IHD in subjects without diabetes		
	No. of cases	Rate, % and (CI)*		No. of cases	Rate, % and (CI)	
Men						
45–54	138	12.3	(10.4–14.4)	802	3.1	(2.9–3.3)
55–64	548	21.8	(20.2–23.4)	2336	9.7	(9.3–10.1)
65–74	816	33.9	(32.1–35.9)	2864	20.4	(19.8–21.1)
45–74	1502	24.9	(23.8–26.0)	6002	9.4	(9.1–9.6)
Women						
45–54	68	9.0	(7.1–11.3)	335	1.3	(1.2–1.4)
55–64	236	15.7	(13.9–17.6)	1016	4.1	(3.8–4.3)
65–74	472	22.9	(21.1–24.8)	1680	10.3	(9.9–10.8)
45–74	776	18.0	(16.9–19.1)	3031	4.5	(4.4–4.7)

* CI = confidence interval

The overall IPRs were 3.97 (95% CI: 3.69 to 4.27) in women and 2.65 (2.52 to 2.79) in men. Stratified by age group, these values decreased (p-value for the trends <0.001) with increasing age for both sexes. The change in IPRs was greater in women, declining from 6.94 (5.41 to 8.92) in the age group 45-54 to 2.21 (2.02 to 2.43) among those aged 64-75 (Figure 3A); the corresponding decrease in men was from 3.97 (3.35

to 4.71) to 1.66 (1.56 to 1.77) (Figure 3B). The gender ratios of the IPRs were 1.75 (1.38 to 2.21) in the age group 45-54 and 1.71 (1.59 to 1.84) in age group 55-64, whereas it was 1.33 (1.26 to 1.40) in the age group 65-74 (Figure 4). Thus, the gender difference decreased with increasing age (p-value for the trend = 0.018): by 56% from the age group 45-54 years to 65-75 years.

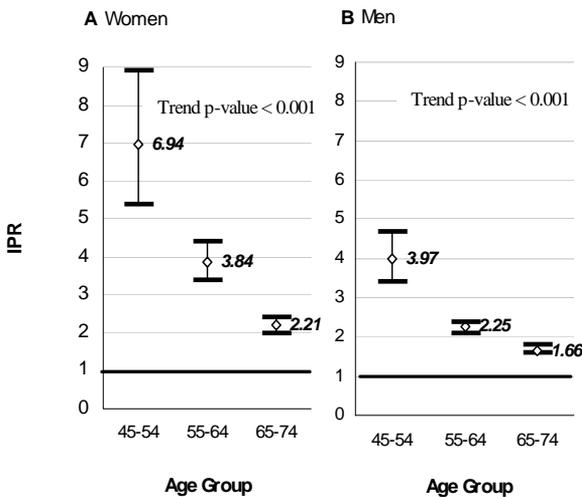


Figure 3. Diabetes/non-diabetes IHD prevalence rate ratios (IPRs) for different age groups with 95% confidence intervals (CIs). The calculations were based on 71 271 women (3A) and 70 129 men (3B) in Östergötland County, Sweden, on 31 December 2003.

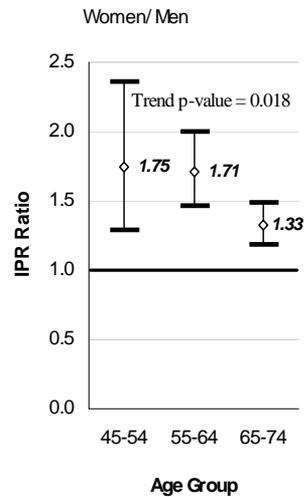


Figure 4. Ratios of the diabetes/non-diabetes IHD prevalence rate ratios (IPRs) for women to men for different age groups with 95% confidence intervals (CIs). The calculations were based on 71 271 women and 70 129 men in Östergötland County, Sweden, on 31 December 2003.

To address the possibilities of bias from the broad inclusion of IHD diagnoses (i.e. the inclusion of angina pectoris) and from the inclusion of IHD in PHC, additional data were analysed. However, the results based on analyses with and

without angina cases (*Figure 5*) as well as with and without IHD cases captured in PHC (*Figure 6*) were virtually the same, indicating that our findings are robust in these respects.

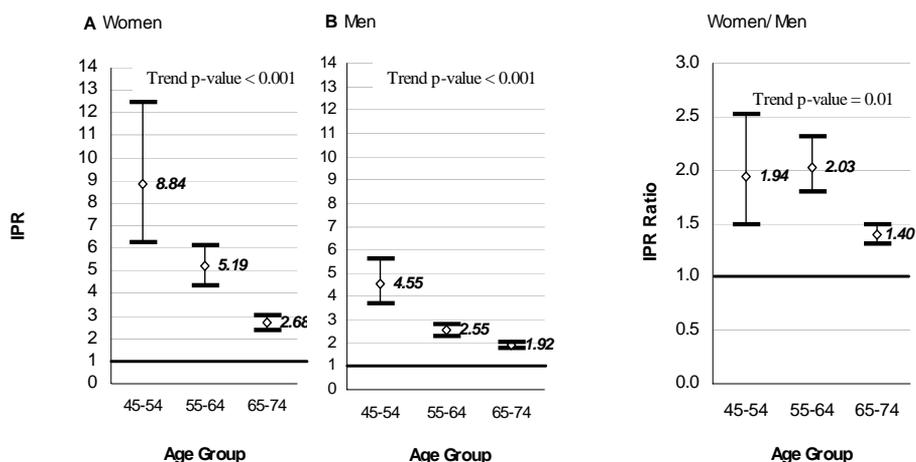


Figure 5. Diabetes/non-diabetes IHD prevalence ratios (IPR) and the female/male ratio of the IPRs (analogously calculated as in Figures 3 and 4) without angina pectoris in the IHD data

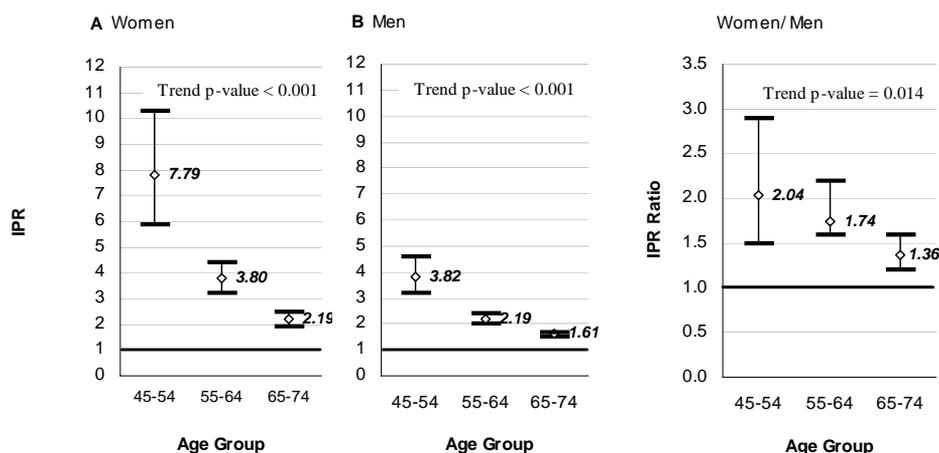


Figure 6. Diabetes/non-diabetes IHD prevalence ratios (IPR) and the female/male ratio of the IPRs (analogously calculated as in Figures 3 and 4) without IHD in PHC data

We found a substantial relative gender difference in IHD associated with diabetes in the younger middle ages, which remained essentially the same up to about 65 years but thereafter decreased considerably. For women, this translates into

a diabetes-caused reduction of the protective effect conferred by female gender by 54% and 52% in the age groups 45-54 years and 55-64 years respectively, compared to 34% in the age group 65-74 years.

Paper IV – Age-specific direct healthcare costs attributable to diabetes in a Swedish population: a register-based analysis

In the population of Östergötland County on 31 December 2004, 19 226 patients (4.6%) had a diagnosis of diabetes mellitus (type 1 and type 2) (Table 9). In all five-year age groups up to 40 years, the preva-

lence was below 2% but increased considerably with age (Figure 7); in the population > 40 years about 10% of the population had a diagnosis of diabetes.

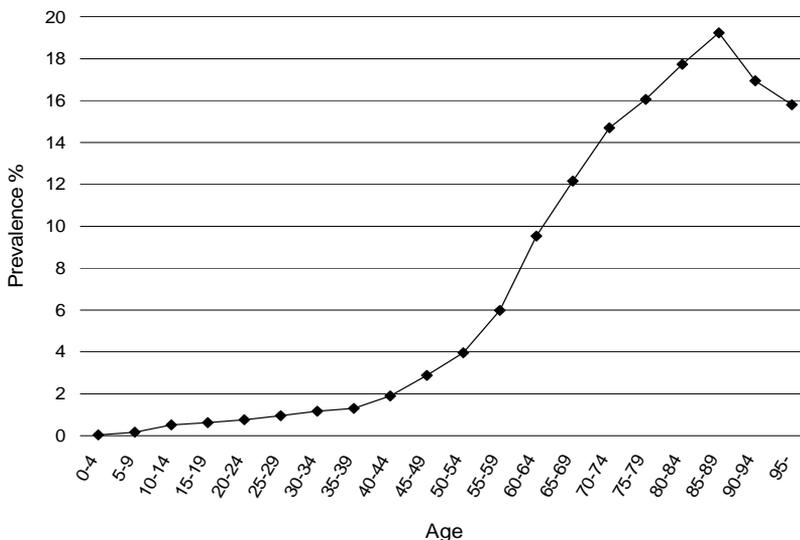


Figure 7. Diabetes prevalence in Östergötland County, Sweden, on 31 December 2004 (n= 19 226)

In 2005, a further 1650 patients were diagnosed with diabetes, and thus the diabetes population in this cost study comprised 20 876 patients

(Table 9) with a mean age of 64.6 years (SD=16.7). In the non-diabetes population, 396 764 individuals were included.

Table 9. Number of subjects with diabetes in the county of Östergötland, Sweden, in total and divided into known cases on 31 December 2004 and new^a cases in 2005

Age	No. of known subjects with diabetes alive on 31 Dec 2004	No. of new ^a subjects with diabetes in 2005	No. of subjects with diabetes in the study population
0-14	195	30	225
15-44	1825	153	1978
45-54	1810	232	2042
55-64	4133	420	4553
65-74	4689	357	5046
75-	6574	458	7032
Total	19 226	1650	20 876

^a: new subjects are those that were not previously identified in CDWÖ. They may have moved to the county from other counties or are previously missed cases and are thus not necessarily incident cases.

The cost per person and year was €4474 (95% CI: €4315–€4633) for the diabetes population versus €2504 (unadjusted: €1330) for the non-diabetes population. The cost per person for in- and outpatient care increased by age in both groups, while the drug cost decreased by age in the diabetic population and

increased in the non-diabetic population. The health cost ratio was 1.8 (95% CI: 1.7-1.9) (unadjusted: 3.4; 3.3-3.5) (Table 10). This ratio decreased with increasing age: from 7.7 (95% CI: 2.0-29.8) in the 0-14 year age group declining to 1.4 (95% CI 1.3-.5) in the oldest group.

Table 10. Annual costs (€) per person for the diabetic (DM) and non-diabetic populations (non-DM) and the additional cost due to diabetes in Östergötland County, Sweden, in 2005

Age	Diabetes population	Cost per person (€)							Additional cost of diabetes		
		Out- and inpatient care ^b		Pharmacy drug sale ^c		All direct healthcare			Per person	Total	Share of total health care cost (%)
	n ^a	DM	non-DM	DM	non-DM	DM	non-DM	Health cost ratio (95% CI)	(€)	(€1000)	(%)
0-14	225	2828	529	1690	60	4518	588	7.7 (2.0-29.8)	3930	884	2.0
15-44	1978	2409	789	1311	130	3720	918	4.1 (1.5-11.2)	2802	5542	3.6
45-54	2042	2844	917	1153	217	3996	1134	3.5 (1.8-7.0)	2862	5845	8.9
55-64	4553	2842	1326	1036	343	3878	1669	2.3 (2.0-2.8)	2209	10 056	10.0
65-74	5046	3484	2056	1020	449	4504	2505	1.8 (1.7-2.1)	1999	10 088	10.3
75-	7032	4303	3213	886	609	5189	3822	1.4 (1.3-1.5)	1367	9610	6.2
All ages, ^d	20 876	3448	2076	1026	427	4474	2504	1.8 (1.7-1.9)	1971	41 137	6.6

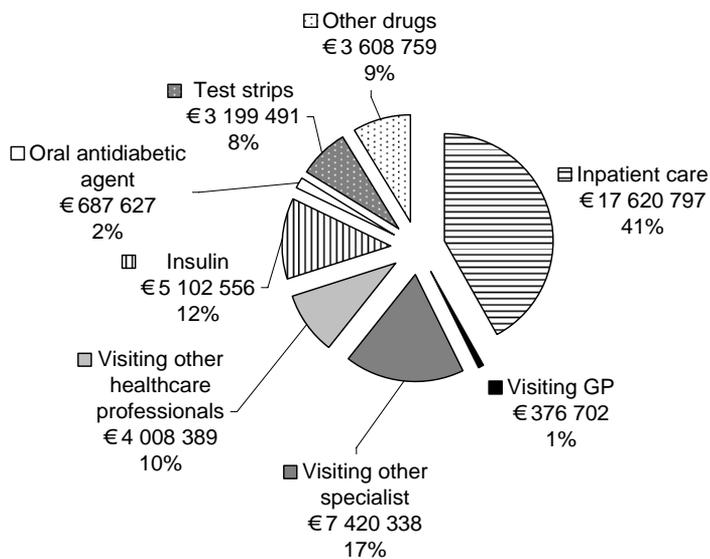
^a number of patients in the diabetes population, ^b including unspecified drugs dispensed at visits to health services or during hospitalization, ^c over-the-counter drug sale including test strips, ^d adjusted to the demographic distribution in the diabetes population

The resource component with the largest additional total costs across all ages was inpatient care followed by medication and visit to a specialist (Figure 8a). Age-specifically, as a percentage share of the total costs (of all utilized resources), inpatient care increased from 25% in children to 44% in subjects over 75 years. Inpatient care was the largest cost component in each age group except in children, for whom visit to a specialist was the largest at 32% of the total costs. The percentage share

of costs for visits to a specialist, for test strips and other drugs decreased while visits to other staff increased with increasing age (Figure 8b).

The proportion of diabetes-related additional costs relative to the total direct healthcare cost for all ages was 6.6% (unadjusted: 9.8%), varying by age in the range 2.0-10.3%, showing a fivefold increase between children and those aged 55-74 years (Table 10).

8a)



8b)

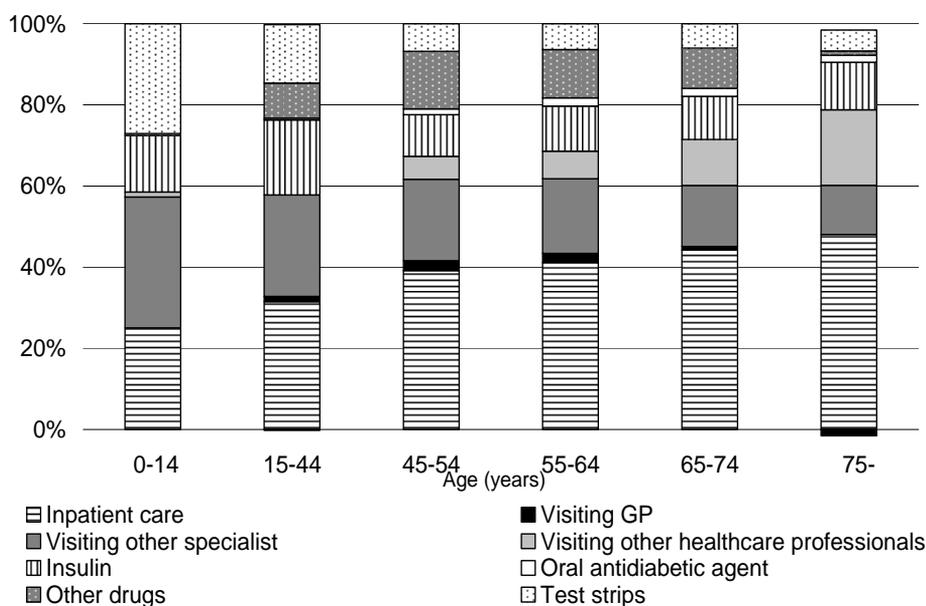


Figure 8. Additional cost due to diabetes: overall age- and sex-adjusted total costs for different cost components (8a) and their relative costs in different age groups (8b), Östergötland County, Sweden 2005

The additional cost per patient across all ages was €1971 (unadjusted: €3145). In contrast to the increase in the total cost with age, the additional cost per patient decreased with age from €3930 in children to €1367 in the oldest age

group. Thus, the directions of the age-specific additional cost per patient and additional total cost diverged (Figure 9a). The distribution of cost components for each age group is presented in Figure 9b.

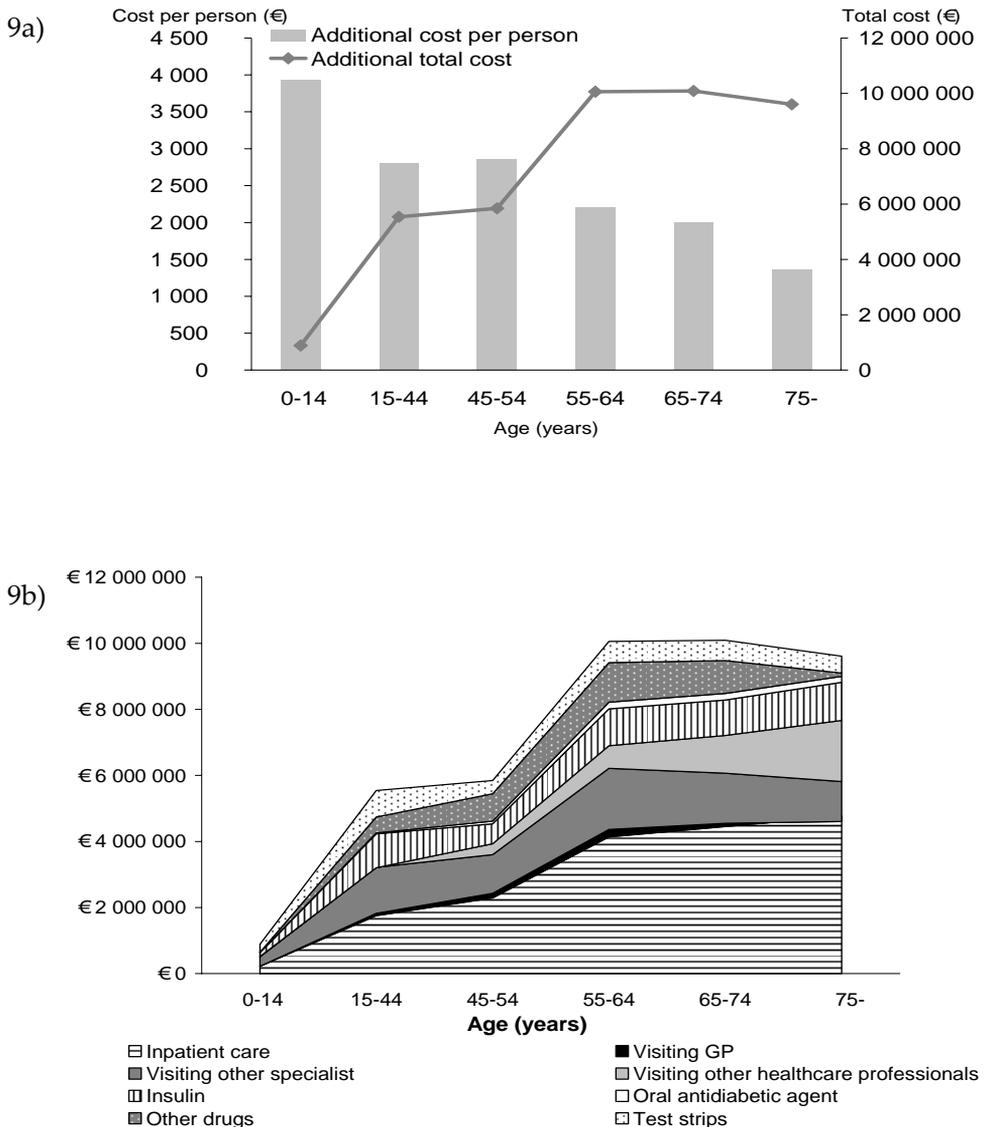


Figure 9. Additional costs due to diabetes: cost per person and total cost (9a) and total cost for different cost components (9b) in Östergötland County, Sweden, 2005

For the subgroups and cost components in this study, the largest additional cost per person was for visits to a specialist by children (€1267), followed by inpatient care in the 45-54 year age group (Table 11). The additional cost per person for visits to a GP was very low for all ages

and in the oldest ages lower even than in the non-diabetes population. Hence, the interpretation of - €22 for a visit to a GP in that age group is an additional cost for those *not* having diabetes. Total additional costs are shown in Table 12.

Table 11. Annual additional cost per person (€) due to diabetes for different cost components in Östergötland County, Sweden, 2005

Age (years)	Cost per person (€)							
	Inpatient care ^a	Outpatient care ^b			Pharmacy sale ^c			
		Visit to GP	Visit to specialist at hospital clinic	Visit to other profes- sionals	Insulin	Oral anti- diabetic agent	Test strips	Other drugs
0-14	984	- 1	1267	49	547	0	1064	20
15-44	887	36	704	- 7	520	13	406	243
45-54	1120	70	574	163	294	41	196	405
55-64	909	50	406	151	245	45	141	263
65-74	883	18	301	226	213	39	120	198
75-	678	- 22	171	263	165	25	72	14
<i>All ages</i> ^d	813	20	351	188	244	33	153	168

Table 12. Annual additional total costs (€1000) due to diabetes for different cost components in Östergötland County, Sweden, 2005

Age (years)	Additional total cost (€1000)							
	Inpatient care ^a	Outpatient care ^b			Pharmacy sale ^c			
		Visit to GP	Visit to specialist at hospi- tal clinic	Visit to other profes- sionals	Insulin	Oral anti- diabetic agent	Test strips	Other drugs
0-14	221	0	285	11	123	0	239	5
15-44	1754	72	1393	13	1028	26	802	480
45-54	2287	143	1171	333	600	83	400	827
55-64	4137	227	1848	689	1114	204	641	1197
65-74	4456	92	1519	1141	1077	198	607	998
75-	4765	-157	1205	1848	1160	176	510	102
<i>All ages</i> ^d	16 972	408	7321	3932	5103	688	3199	3513

^a including unspecified drugs dispensed during hospitalization, ^b including unspecified drugs dispensed at outpatient visits, ^c over-the-counter sale, ^d adjusted to the demographic distribution in the diabetes population (notes apply to both Table 11 and 12)

DISCUSSION

The main focus of this discussion is methodological issues. The author refers to the specific papers for critical considerations of the conclusions in the individual studies.

The results of the various issues raised in the different papers of this thesis have shown the importance of collecting and storing population-based incidence data, as well as the applicability of diagnosis-related administrative registers in healthcare research.

Associations between testicular cancer incidence and serum cholesterol were found in a follow-up study based on cancer data (Paper I). The estimated diabetes prevalence from administrative healthcare data agreed very well with previous estimates (Paper II) and also demonstrated that the diabetes impact on IHD differed by age and gender (Paper III). The direct healthcare costs attributable to diabetes appeared to vary by age in a study using CPP data and cost data on prescribed drugs (Paper IV).

The relevance of using registers in research depends not only on whether it is methodologically feasible, but also on whether problems that affect the results can be identi-

fied. Important factors for being methodologically possible are that suitable data exist in a format that can be handled, that data can be linked to other data, and further that the data are accessible. These factors may appear self-evident, but they are nevertheless important. Sørensen et al. (1996) [84] developed a framework for evaluation of register studies by identifying common, important factors that may affect the results. Besides methodological possibility above, the authors identified factors such as: completeness of individuals, accuracy and completeness of registered data, the registration period, and costs. These factors illustrate various types of misclassifications and other problems that may affect the results and will be discussed further in this section. The factors relating to methodological possibility will not be dealt with here since they were straightforward in these studies.

Paper I

The testicular cancer incidence and serum cholesterol association in Paper I was very strong even though

few cancer cases were found ($n=21$). It is not likely that missed cases differ between serum cholesterol cate-

gories. Therefore, the association found may even have been underestimated assuming that a number of cancer cases were missing. Since the testicular cancer diagnosis was not specified further, i.e. seminoma or non-seminoma, misclassifications due to specification of the wrong cancer type is not likely either, and for this reason the accuracy can be considered high.

To assert an association to a condition with this very low incidence (6.8 per 100 000 men in 2004) [36], a large study population and a long study period were needed. Since all data already existed, this study was low cost.

The highly significant positive association between serum chole-

sterol and testicular cancer suggests that an elevated concentration of serum cholesterol is a risk factor for testicular cancer. However, changes in serum cholesterol concentration in the population do not explain the increasing incidence in recent decades, since available data suggests that these levels rather decreased [85]. Interpretation of the association may instead be complicated by other influencing and confounding factors. One such factor could be undernutrition, the essence of the “foetal origins” hypothesis, suggesting that several adult diseases may be caused by undernutrition *in utero* [86].

Paper II and III

Completeness of individuals

In Paper I and Paper II, the data extraction procedure was the same, i.e. a case-finding algorithm that searched for physician-diagnosed diabetes, hypertension, asthma, COPD and IHD in a five-year period retrospectively from 31 December 2003 to 1 January 1999. The algorithm captured the cases regardless of whether there was a main or secondary diagnosis in both the out- and inpatient health-care contacts registered in the CDWÖ. This means that regardless

of whether a patient contacted health services once or 50 times in this five-year period, and the diagnosis in question was registered, he/she still represents only one case.

A number of factors may have affected the completeness of individuals in Papers II and III. First, visits solely to occupational care and private practices may affect the prevalences for some diagnoses. No data are available for occupational care in Östergötland. However, in a similar prevalence study, albeit us-

ing only one case-finding year, though including occupational care, Wigertz & Westerling (2001) found that 2.7% of all consultations/hospitalizations were from occupational care. We do not know the ranking of diagnosis type in occupational care, but probably the percentage decreases for every year added for this patient group as whole. For private non-primary healthcare, we know that the cost was 3% of total healthcare expenditures. Of these, 70% were for musculoskeletal, gynaecological and psychiatric disorders. Another factor that may have affected completeness is that inter-county migration was not considered in the analyses; however, the migration rate was low, e.g. 0.5% in both directions in the age group < 45 years [87]. Therefore, the lack of data from occupational care and private practices as well as migration probably had little effect on the prevalence estimates for these specific diagnoses.

Macrovascular complications because of diabetes, e.g. IHD, are attributed primarily to type 2 diabetes. Thus, in Paper III, it would have preferable to study only type 2 dia-

betes and its impact on IHD. This was not possible, since the diagnostic coding according to the ICD10 system does not differentiate between type 1 and 2 diabetes. Thus, this is also the case in CDWÖ data.

To explore whether IHD data derived from PHC affected the results, calculations and ratios of the IPRs were also performed for hospital clinics only (out- and inpatients combined). Although IHD data in this analysis covered 74% of the total captured IHD data and the overall IHD prevalence rate as a consequence was 26% lower, the IPR age trend patterns proved to be similar to the result with all data included.

Furthermore, to ascertain whether inclusion of angina pectoris (ICD-10 code I20) in the definition of IHD would influence the estimates, we performed the calculations both with and without the angina diagnosis. We found that although the prevalence rates were indeed lower when angina was not included, the relationships between the different groups (i.e. considering the IPRs and the ratios of the IPRs) were virtually unchanged, as were significant trends.

Accuracy and completeness of data

Accuracy of the registered data includes problems of shortcomings in the diagnosing/coding/registering procedure. Although the right to

diagnose is not regulated by law, in common practice diagnoses are made by a doctor [88], including in the LiO. For this reason, in Papers II

and III, only data on visits to a doctor or hospitalizations were included. In Paper II, the prevalence of diabetes was identical to the self-reported prevalence found in the 2004 follow-up conducted as part of the project the northern Sweden World Health Organization Monitoring of Trends and Determinants in Cardiovascular disease (WHO MONICA) [41]. The prevalences of hypertension, asthma and COPD were lower than previous estimates from surveys. This may be an effect of the development in recent years towards an increased use of specialised nursing practices and these diagnoses are all examples of such practices. As is their purpose, this leads to fewer visits to doctors' surgeries, but it presumably also reduces the case-finding rate, at least in the short term.

The accuracy of diagnosing can only be assessed using a defined "true instrument" as a gold standard. This may be some sort of standardized procedure with thor-

ough assessment of patients' medical charts. So far, this procedure has not been applied in any study using CDWÖ data. Validation of other administrative data, for instance in studies in the United States and Canada [89-91], has shown high specificity in registers covering all types of care but lower, though varying, sensitivity in registers of PHC and inpatient hospital care. It is probable that the algorithm used on CDWÖ data also show lower sensitivity, although this will improve with every year added. Although the accuracy of the diagnoses can only be speculated upon, the completeness of registered diagnoses is available information. In all visits to a doctor and all hospitalizations for the diabetic population in Paper III, about 90% of visits to a doctor recorded at least one diagnosis, versus 83% in outpatient hospital care (*Table 13*). Virtually all hospitalizations recorded at least one diagnosis.

Table 13. Registered diagnoses, first and secondary, at visits to a doctor (GP and other) and hospitalizations in the period 1999-2003, among 10 364 subjects with diabetes aged 45-74 years in Östergötland County on 31 December 2003

	No. of contacts 1999-2003	At least one diagnosis		At least two diagnoses	
		n	%	n	%
PHC	82 683	74 193	89.7	55 589	32.8
Outpatient hospital care	99 315	82 262	82.8	27 915	28.1
Inpatient hospital care	14 600	14 451	99.0	9345	64.0

Gender differences in the risk of IHD in diabetic and non-diabetic patients have been the focus of numerous studies. Although reported rates vary somewhat, they all demonstrate higher relative risks of IHD in women than men, and our data show similar results. The female/male ratio of the relative risks derived from two meta-analyses, Lee et al. (2000) [92] and Huxley et al. (2006) [62], were 1.39 and 1.46 respectively. Our corresponding ratio, 1.50 (95% CI: 1.39 to 1.61), agreed well and is some sort of validation of the accuracy of the diagnoses. However, assuming different misclassifications in the data, it is possible to speculate on the direction in which the ratios are moving. First, when using register data to identify comorbid conditions, it is likely that subjects will be detected in an earlier phase of morbidity in a population that is already affected by the disease, since these patients will be under closer scrutiny – a phenomenon known as Berkson’s bias [93]. In our investigation, such bias may have led to overestimation of the IPRs. However, this would not have affected the detected trends unless there were differences

Registration period

One reason why diabetes is estimated more in accordance with previous results compared with the other diagnoses in Paper II is that

in the bias with respect to gender. Second, it is recognized that a considerable number of people have undiagnosed diabetes. Eliasson et al. (2002) [94] found that the prevalence estimate including both known and unknown diabetes was doubled compared to known diabetes only. This also seems to be true in an infarct population, which has been shown to have an estimated known diabetes prevalence of 20% and an unknown diabetes prevalence of 25%. Furthermore, patients with previously unknown diabetes are more likely to have an acute coronary syndrome than patients with diabetes and those without diabetes, whereas a cardiovascular history is seen less often among these patients [95]. Considering the true diabetes prevalence as a summary of both known and unknown diabetes, we have underestimated the IPRs in our study. The underestimation is slightly greater among women and also among those in the oldest age group, assuming that the gender and age distribution of unknown diabetes prevalence in our population corresponds to that found by Eliasson et al. (2002).

LiO has passed a resolution stating that health services must advise all diabetic patients to visit their doctor once a year. The estimated satura-

tion rate suggests, however, that 71% of all subjects with diabetes in Östergötland visited a doctor in 2003 (Table 14). The differences in

saturation rate indicate that the years needed to capture the whole diseased population in the CDWÖ varies between the diagnoses.

Table 14. The proportions of annual cumulative case findings in the Care Data Warehouse in the CDWÖ database in the period 1999–2003 among inhabitants in Östergötland on 31 December 2003

Year	Cumulative case findings, %			
	Diabetes (n = 18 134)	Hypertension (n = 42 796)	Asthma (n = 18 451)	COPD (n = 4 812)
2003	71	50	38	58
2002–2003	89	74	59	78
2001–2003	95	87	75	89
2000–2003	98	95	89	96
1999–2003	100	100	100	100

Paper IV

The analyses in paper IV reflect the organization of the Swedish health service, and may not necessarily apply to other healthcare systems.

Although good register data does exist in Sweden, the area of register-based studies in health economics needs to be further explored [96]. This is the first study in Sweden using administrative healthcare data together with the new Swedish Prescribed Drug Register for analyses of direct healthcare costs of a disease.

The prevalence-based study design with a bottom-up approach

was an appropriate method for this study since the cost for a certain time interval was acquired for all healthcare services used. A further strength of the study was that estimates are for the total population rather than extrapolating from a sample or subpopulation.

The identifying process of individuals followed the same procedure in this cost study as in the epidemiological studies in Paper II and III. Thus, the discussion of completeness of individuals with diabetes follows that discussion.

Accuracy and completeness of data

The proportion of known diabetes agrees with previous studies of diabetes prevalence in Sweden [41]. As mentioned above, it is well known that there is a large number of people with unknown diabetes. However, in this study, we have used the prevalence of known diabetes. This lowers the true sensitivity. Since the diabetes data are subtracted from the total data to obtain the non-diabetes population and their cost, a lowered sensitivity underestimates the cost estimates in this study.

The main focus in the CPP calculations, performed by the CPP group in the LiO finance department, has been the accuracy of the service-specific estimates. There is also a continuous refining of the estimates. For example, there are ongoing discussions with clinic representatives on whether there are relevant differences between first-time and repeat visits in the same type of contact with the same clinic.

Some costs are not included in the CPP database. One such post that might have affected the results in this study is the cost for specialist care at hospitals in other counties (e.g. transplants). Examples of other known posts missing from the database are the costs in one of the

three county regional hospital-based homecare units and some personnel costs in emergency clinics. These amount to about 1% of total healthcare costs and probably do not affect this study. Of course, the aim is to achieve completeness in the CPP database and this work is indeed progressing. However, the view of the CPP group is that at present there is underestimation rather than overestimation of the CPP values compared to the true costs [97].

A strength regarding the reliability of the cost estimates in Paper IV is that data extractions and estimates were performed similarly in the diabetes and non-diabetes populations, giving similar possible systematic errors in both groups. Hence, the relative estimates probably have high accuracy. Furthermore, although it is difficult to compare cost estimates from different studies and especially between countries, this may be one way of strengthening their accuracy. The medical costs in 2001 estimated by Ferber et al. (2007) [98] were €4447 per diabetic patient, and the excess cost was €1619 for complications due to diabetes. These agree very well with those in paper IV: €4474 and €1971 respectively.

Implications and future improvements

Data from a laboratory test measured several decades ago in a defined population were linked to 25 years of incidence data from the Cancer Registry. The combined data made it possible to suggest a previously unknown association. The finding demonstrates the usefulness of storing population-based recorded test values for future unplanned analyses. Further, it strengthens the already known usefulness of archived incidence data in the Cancer Registry.

Moreover, outcome data from the CDWÖ were used in three different analyses. Thus, this thesis can be seen as an initiation of the use of administrative registers in epidemiological research at all healthcare levels in Sweden. The data composition in the CDWÖ is useful for cross-sectional analyses. Conversely, for analyses of risk, incidence and to some extent also mortality, i.e. longitudinal studies, the current CDWÖ structure does not provide suitable data. There is, however, one relatively small but necessary improvement that would increase the possibilities for future longitudinal studies from administrative register data: a date registering the first time a diagnosis of interest is recorded for each patient. Such incidence data would also be useful in cost studies, making it possible to

estimate the economic benefits of reducing the number of new cases. Further improvements in the CDWÖ, from a research perspective, would be to incorporate both death and migration data into the register to simplify the data processing.

Due to political opinion in Sweden today, it is probable that private actors (and perhaps others) will increase in the Swedish healthcare sector. From a research perspective, it is important that contracts between the county councils and those actors are formulated to maintain (or even improve) the coverage in the public databases.

A national register similar to the CDWÖ, i.e. covering all healthcare levels, would be useful in healthcare research. Together with the improvements discussed above, such a register would provide national population-based data that could be used in longitudinal studies, i.e. similar to the Cancer Registry but for several diagnoses or associations.

Importantly, in the archiving and use of register data, continuous discussions about ethical issues are necessary for a number of reasons, not least in order to maintain the well-developed security we have in Sweden today regarding personal information.

CONCLUSIONS

Storing of incidence data over a long period is crucial in epidemiological studies of disorders with low incidence, and especially for diseases with low lethality. This is exemplified in this thesis by a study based on data from the useful Cancer Registry, a national health database established in 1958. The new administrative healthcare databases in various Swedish counties are mainly created for managerial purposes. As shown in this thesis, these new databases are useful in that they can provide valuable information on healthcare in a defined population in terms of prevalence. Furthermore, with the new opportunities, i.e. available CPP data, cost estimates are also feasible.

The specific conclusions of this thesis are:

- High levels of serum cholesterol may be a risk factor for testicular cancer (Paper I).
- Several consecutive years of data for both hospital care and primary healthcare are needed when estimating the prevalence of chronic diseases from an administrative database (Paper II).
- The diabetes-caused reduction of the protective effect against IHD conferred by female gender decreases with age (Paper III).
- The cost per patient and the relative magnitude of different cost components varied considerably by age, which is important to consider in future planning of diabetes management (Paper IV).

SUMMARY IN SWEDISH

I Sverige finns unika möjligheter att bedriva registerbaserad forskning, dels tack vare en lång tradition av datainsamling till populationsbaserade register, men också på grund av systemet med personnummer som gör det lätt att länka information från olika register. De senaste decenniernas framsteg inom informationsteknologin har bland annat medfört att man inom hälso- och sjukvården kan arkivera allt större mängder data. Detta borde innebära en ytterligare potential för hälsorelaterad forskning.

Det övergripande syftet med avhandlingen är att beskriva och diskutera användningen och användbarheten av populationsbaserade register för hälsorelaterad forskning. Detta görs med hjälp av specifika frågeställningar inom epidemiologi och hälsoekonomi, redovisade i fyra delarbeten.

I det första delarbetet hämtades incidensdata för testikelcancer från Cancerregistret. Dessa cancerdata länkades samman med uppgifter om serumkolesterol från en stor screeningstudie, och visade att en hög koncentration av serumkolesterol var kopplad till ökad risk för testikelcancer. Efter som upptäckten är den första av sitt slag, och på grund av vida konfidensintervall, behövs ytterligare studier för att fastställa sambandet.

Data från Östergötlands läns landstings administrativa databas (Vårddatalagret (VDL)) användes för att beräkna prevalensen av fyra vanliga kroniska sjukdomar. Prevalensen stämde mycket väl överens med tidi-

gare estimat för diabetes och relativt väl för astma. Prevalensen för hypertoni och kronisk obstruktiv lungsjukdom var lägre än tidigare presenterade beräkningar. Studien visade att data för flera konsekutiva år för både primärvård och sjukhusvård krävs för att fall inte ska missas.

Information från VDL användes även för att studera hur diabetes påverkar förekomsten av ischemisk hjärtsjukdom (IHD). I alla åldrar var kvoten mellan diabetikers och ickeddiabetikers IHD-prevalens högre hos kvinnor. Den relativa könsskillnaden hade samma storlek upp till 65-årsåldern men minskade därefter markant.

Via data från VDL, landstingets Kostnad-Per-Patient-databas och Läkemedelsregistret, studerades åldersspecifika direkta hälso- och sjukvårdskostnader för diabetes. Kostnaden per patient och den relativa storleken för olika kostnadskomponenter varierade avsevärt mellan olika åldrar, vilket är viktigt att ta hänsyn till i den framtida planeringen av diabetesvården.

Cancerregistret etablerades huvudsakligen för epidemiologisk bevakning och forskning. I den här avhandlingen exemplifieras dess användbarhet genom en studie om testikelcancer. De nyetablerade svenska administrativa landstingsdatabaserna är i första hand skapade för att ge underlag för ekonomisk styrning. Här visas att de även kan användas för att besvara frågeställningar inom hälsorelaterad forskning.

ACKNOWLEDGEMENTS

In my work on this thesis, I have had the privilege of being supervised and taught principally by *John Carstensen*. I wish to express my sincere gratitude for excellent guidance and for the many pleasant chats during this time, about everything and nothing.

I also wish to thank the following people:

Carl Johan Östgren, my co-supervisor, for valuable guidance in diabetes research, and for constructive commentaries and stimulating discussions.

Agneta Andersson, my co-author and friend, for her support in health economic issues and valuable comments on the thesis in general and, of course, for lots of laughter.

My other co-authors, *Mikael Karlsson* for guidance in the CDWÖ database, and *Sven Törnberg*, for sharing his expertise in cancer research.

Bengt Grip, *Lars Svensson* and *Rolf Wiklund* in the CPP group, for their great helpfulness, and for patiently answering my questions.

Anette Wickström, my co-PhD-student and friend, for our second breakfast chats, and for feeding me, in her theatrical manner, with crazy stories. It has been good fun!

All other friends and colleagues at tema Health and Society for interesting discussions from different perspectives.

All dear friends and colleagues at the Local Healthcare Research and Development Unit, for cheer and support.

Ewa Grodzinsky, *Malou Lindberg* and *Susan Wilhelmsson* for friendship and culinary evenings.

Carin Carelind, for constant friendship throughout the years.

Ingrid Svensson, my wonderful mother, for her support of me and my family.

Finally, I would like to thank my husband, *Tony*, for his love, pep talks and patient listening, and our fantastic daughters, *Lotten* and *Amanda*, not only for being just the way they are, but also for their love, support and a numerous variety of home-made fruit pies and cakes.

Studies in this thesis have been supported by the Research Council of South-Eastern Sweden (FORSS).

REFERENCES

1. World Health Organization. *Building foundations for eHealth. Progress of Member States. Report of the WHO Global Observatory for eHealth*. 2006. World Health Organization, 2006.
2. Muggleton, S. H. *2020 computing: exceeding human limits*. *Nature*, 2006; 440: 409-10.
3. Szklo, M. *Population-based cohort studies*. *Epidemiol Rev*, 1998; 20: 81-90.
4. Last, J. M., *A dictionary of epidemiology*. 1988, New York: Oxford University Press.
5. Wettermark, B., Hammar, N., Michaelfored, C., Leimanis, A., Otterblad Olausson, P., Bergman, U., et al. *The new Swedish Prescribed Drug Register-Opportunities for pharmacoepidemiological research and experience from the first six months*. *Pharmacoepidemiol Drug Saf*, 2007; 7: 726-35.
6. Centre of Epidemiology, National Board of Health and Welfare. *Information om EpCs register och statistik*. [Information about EpCs registers and statistics (In Swedish)] Available from: http://www.socialstyrelsen.se/Om_Sos/organisation/Epidemiologiskt_Centrum/Enheter/R/Information_om_EpCs_register_och_statistik.htm. Accessed March 15 2007.
7. Centre of Epidemiology, National Board of Health and Welfare. *Hälsodatabaser räddar liv och förbättrar livskvalitet [Health databases may save lives and improve quality of life (In Swedish)]*. Stockholm, Sweden: National Board of Health and Welfare, 2005.
8. Statistics Sweden. *Befolkningsstatistik [The Total Population Register]*. Available from: http://www.scb.se/templates/tableOrChart_25896.asp. Accessed March 15 2007.
9. Centre of Epidemiology, National Board of Health and Welfare. *Dödsorsaksregistret [The Cause of Death Registry]*. Available from: http://www.socialstyrelsen.se/Statistik/statistik_amne/dodsorsaker/Dodsorsaksregistret.htm. Accessed March 15 2007.

10. Centre of Epidemiology, National Board of Health and Welfare. *Cancerregistret [The Cancer Registry]*. Available from: http://www.socialstyrelsen.se/Statistik/statistik_amne/Cancer/cancerregistret.htm Accessed March 15 2007.
11. National Board of Health and Welfare. *Ändring i föreskrifterna och allmänna råden (SOSFS 2006:13) om uppgiftsskyldighet till cancerregistret vid Socialstyrelsen. [Change of the directions and the general advices about obligations to report to the Cancer Registry at the National Board of Health and Welfare (In Swedish)]*. Report: SOSFS 2006:15 (M). Stockholm, Sweden: National Board of Health and Welfare, 2006. Available from: http://www.sos.se/sosfs/2006_15/2006_15.pdf Accessed March 15 2007.
12. Pershagen, G., Akerblom, G., Axelson, O., Clavensjo, B., Damber, L., Desai, G., et al. *Residential radon exposure and lung cancer in Sweden*. N Engl J Med, 1994; 330: 159-64.
13. Kohli, S., Noorlind Brage, H., Lofman, O. *Childhood leukaemia in areas with different radon levels: a spatial and temporal analysis using GIS*. J Epidemiol Community Health, 2000; 54: 822-6.
14. Anderson, H., Bladstrom, A., Olsson, H., Moller, T. R. *Familial breast and ovarian cancer: a Swedish population-based register study*. Am J Epidemiol, 2000; 152: 1154-63.
15. Centre of Epidemiology, National Board of Health and Welfare. *Läkemedelsregistret. [The Swedish Prescribed Drug Register]*. Available from: http://www.socialstyrelsen.se/Statistik/statistik_amne/lakemedel/Lakemedelsregistret.htm. Accessed March 15 2007.
16. Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., Hearst, N., Newman, T. B., *Designing Clinical Research. An epidemiological approach*. Second ed. 2001, Philadelphia: Williams & Wilkins.
17. Iezzoni, L. I. *Assessing quality using administrative data*. Ann Intern Med, 1997; 127: 666-74.
18. Schwartz, R. M., Gagnon, D. E., Muri, J. H., Zhao, Q. R., Kellogg, R. *Administrative data for quality improvement*. Pediatrics, 1999; 103: 291-301.
19. Centre of Epidemiology, National Board of Health and Welfare. *Patientregistret. [The Register of Hospital Discharges]*. Available from: http://www.socialstyrelsen.se/Statistik/statistik_amne/sluten_vard/Patientregistret.htm Accessed March 15 2007.

20. Scales, D. C., Guan, J., Martin, C. M., Redelmeier, D. A. *Administrative data accurately identified intensive care unit admissions in Ontario*. J Clin Epidemiol, 2006; 59: 802-7.
21. McAlister, F. A., Campbell, N. R., Duong-Hua, M., Chen, Z., Tu, K. *Antihypertensive medication prescribing in 27,822 elderly Canadians with diabetes over the past decade*. Diabetes Care, 2006; 29: 836-41.
22. Tai, D., Dick, P., To, T., Wright, J. G. *Development of pediatric comorbidity prediction model*. Arch Pediatr Adolesc Med, 2006; 160: 293-9.
23. Campbell, N. R., Tu, K., Brant, R., Duong-Hua, M., McAlister, F. A. *The impact of the Canadian Hypertension Education Program on antihypertensive prescribing trends*. Hypertension, 2006; 47: 22-8.
24. Hux, J. E., Ivis, F., Flintoft, V., Bica, A. *Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm*. Diabetes Care, 2002; 25: 512-6.
25. Hux, J. E., Lipscombe, L. L. *Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study*. Lancet, 2007; 369: 750-756.
26. Wigertz, A., Westerling, R. *Measures of prevalence: which healthcare registers are applicable?* Scand J Public Health, 2001; 29: 55-62.
27. Inmon, W. H., *Building the data warehouse*. 1992, New York: John Wiley & sons.
28. Hristovski, D., Rogac, M., Markota, M. *Using data warehousing and OLAP in public health care*. Proc AMIA Symp, 2000; 369-73.
29. Breault, J. L., Goodall, C. R., Fos, P. J. *Data mining a diabetic data warehouse*. Artif Intell Med, 2002; 26: 37-54.
30. Grant, A., Moshyk, A., Diab, H., Caron, P., de Lorenzi, F., Bisson, G., et al. *Integrating feedback from a clinical data warehouse into practice organisation*. Int J Med Inform, 2006; 75: 232-9.
31. Junttila, K., Meretoja, R., Seppala, A., Tolppanen, E. M., Ala-Nikkola, T., Silvennoinen, L. *Data warehouse approach to nursing management*. J Nurs Manag, 2007; 15: 155-61.

32. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems 10th Revision for 2007*. Available from: <http://www.who.int/classifications/apps/icd/icd10online/> Accessed March 15 2007.
33. Swedish Association of Local Authorities. *Kostnad Per Patient (KPP) för bättre ledningsinformation i sjukvården. [Cost per Patient (CPP) for improving management information in health care (In Swedish)]*. Stockholm, Sweden: Association of Local Authorities, 2002. Available from: http://brs.skl.se/brsbibl/kata_documents/doc30415_1.pdf Accessed March 15 2007.
34. National Board of Health and Welfare, *Socialstyrelsen redovisar: The Värmland Survey. Report No 23*. 1971, National Board of Health and Welfare: AB Allmänna förlaget. Stockholm.
35. Scottenfield, D., Fraumeni, J., *Cancer Epidemiology and Prevention. Testicular Cancer*. 1996, Oxford: Oxford University Press.
36. Centre of Epidemiology, National Board of Health and Welfare. *Incidence matrices, The Cancer Registry*. Available from: http://www.socialstyrelsen.se/Sos/ExternalWeb/Templates/Common.aspx?NRMODE=Published&NRNODEGUID=%7b6028F89C-81AA-4B41-A1FA-5792D9EE2219%7d&NRORIGINALURL=%2fStatistik%2fstatistik_amne%2fCancer%2fcancerregistret%2htm&NRCACHEHINT=Guest#Matriser. Accessed March 15 2004.
37. Skakkebaek, N. E., Berthelsen, J. G., Giwercman, A., Muller, J. *Carcinoma-in-situ of the testis: possible origin from gonocytes and precursor of all types of germ cell tumours except spermatocytoma*. *Int J Androl*, 1987; 10: 19-28.
38. Sigurdson, A. J., Chang, S., Annegers, J. F., Duphorne, C. M., Pillow, P. C., Amato, R. J., et al. *A case-control study of diet and testicular carcinoma*. *Nutr Cancer*, 1999; 34: 20-6.
39. Armstrong, B., Doll, R. *Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices*. *Int J Cancer*, 1975; 15: 617-31.

40. Kavlock, R. J., Daston, G. P., DeRosa, C., Fenner-Crisp, P., Gray, L. E., Kaattari, S., et al. *Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop*. Environ Health Perspect, 1996; 104 Suppl 4: 715-40.
41. Eliasson, M., Bostrom, G. *Chapter 5.2: major public health problems - diabetes*. Scand J Public Health Suppl, 2006; 67: 59-68.
42. Cowie, C. C., Rust, K. F., Byrd-Holt, D. D., Eberhardt, M. S., Flegal, K. M., Engelgau, M. M., et al. *Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health And Nutrition Examination Survey 1999-2002*. Diabetes Care, 2006; 29: 1263-8.
43. King, H., Aubert, R. E., Herman, W. H. *Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections*. Diabetes Care, 1998; 21: 1414-31.
44. Jonsson, B. *Revealing the cost of Type II diabetes in Europe*. Diabetologia, 2002; 45: S5-12.
45. The National Diabetes Register (NDR). *Årsrapport 2007 [Annual Report 2007 (In Swedish)]*. Available from: <https://www.ndr.nu/> Accessed March 15 2007.
46. Berger, B., Stenstrom, G., Sundkvist, G. *Incidence, prevalence, and mortality of diabetes in a large population. A report from the Skaraborg Diabetes Registry*. Diabetes Care, 1999; 22: 773-8.
47. *1999 World Health Organization-International Society of Hypertension Guidelines for the Management of Hypertension. Guidelines Subcommittee*. J Hypertens, 1999; 17: 151-83.
48. Berg, C. M., Lissner, L., Aires, N., Lappas, G., Toren, K., Wilhelmsen, L., et al. *Trends in blood lipid levels, blood pressure, alcohol and smoking habits from 1985 to 2002: results from INTERGENE and GOT-MONICA*. Eur J Cardiovasc Prev Rehabil, 2005; 12: 115-25.
49. Wolf-Maier, K., Cooper, R. S., Banegas, J. R., Giampaoli, S., Hense, H. W., Joffres, M., et al. *Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States*. JAMA, 2003; 289: 2363-9.
50. Lundback, B. *Epidemiology of rhinitis and asthma*. Clin Exp Allergy, 1998; 28 Suppl 2: 3-10.

51. Backlund, A. B., Perzanowski, M. S., Platts-Mills, T., Sandstrom, T., Lundback, B., Ronmark, E. *Asthma during the primary school ages--prevalence, remission and the impact of allergic sensitization*. *Allergy*, 2006; 61: 549-55.
52. *From the Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA) 2006*. Available from: <http://WWW.ginasthma.org>. Accessed March 15 2007.
53. Dewar, M., Curry, R. W., Jr. *Chronic obstructive pulmonary disease: diagnostic considerations*. *Am Fam Physician*, 2006; 73: 669-76.
54. Pauwels, R. A., Buist, A. S., Ma, P., Jenkins, C. R., Hurd, S. S. *Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: National Heart, Lung, and Blood Institute and World Health Organization Global Initiative for Chronic Obstructive Lung Disease (GOLD): executive summary*. *Respir Care*, 2001; 46: 798-825.
55. Lundback, B., Lindberg, A., Lindstrom, M., Ronmark, E., Jonsson, A. C., Jonsson, E., et al. *Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies*. *Respir Med*, 2003; 97: 115-22.
56. *Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction*. *Eur Heart J*, 2000; 21: 1502-13.
57. Tunstall-Pedoe, H., Kuulasmaa, K., Mahonen, M., Tolonen, H., Ruokokoski, E., Amouyel, P. *Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease*. *Lancet*, 1999; 353: 1547-57.
58. Sowers, J. R. *Diabetes mellitus and cardiovascular disease in women*. *Arch Intern Med*, 1998; 158: 617-21.
59. Mercurio, G., Zoncu, S., Dragoni, F. *Gender differences in cardiovascular risk factors*. *Ital Heart J*, 2003; 4: 363-6.
60. Mak, K. H., Haffner, S. M. *Diabetes abolishes the gender gap in coronary heart disease*. *Eur Heart J*, 2003; 24: 1385-6.

61. Barrett-Connor, E. L., Cohn, B. A., Wingard, D. L., Edelstein, S. L. *Why is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? The Rancho Bernardo Study.* JAMA, 1991; 265: 627-31.
62. Huxley, R., Barzi, F., Woodward, M. *Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies.* BMJ, 2006; 332: 73-8.
63. Booth, G. L., Kapral, M. K., Fung, K., Tu, J. V. *Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study.* Lancet, 2006; 368: 29-36.
64. Centre of Epidemiology, National Board of Health and Welfare. *Hjärtinfarkter [Myocardial Infarctions].* Available from: http://www.socialstyrelsen.se/Statistik/statistik_amne/hjartsjukdomar/index.htm Accessed March 15 2007.
65. *The Register of Information and Knowledge about Swedish Heart Intensive care Admissions (RIKS-HIA)* Available from: <http://www.ucr.uu.se/rikshia/>. Accessed March 15 2007
66. The Swedish Medical Products Agency. *Prevention av aterosklerotisk hjärtkärlsjukdom med lipidreglerande läkemedel - Behandlingsrekommendation. [Guidelines for prevention with lipid lowering drugs on atherosclerotic heart disease (In Swedish)].* Stockholm, Sweden: The Swedish Medical Products Agency, 2005. Available from: http://www.lakemedelsverket.se/upload/Hälsa%20och%20sjukvård/behandlingsrek/ws_hyperlipidemi.pdf. Accessed March 15 2007
67. Kirkwood, B. R., Sterne, J. A. C., *Essential Medical Statistics.* Second ed. 2003, Oxford: Blackwell Science Ltd.
68. Grimes, D. A., Schulz, K. F. *An overview of clinical research: the lay of the land.* Lancet, 2002; 359: 57-61.
69. Cox, D. R., Snell, E. J., *Analysis of Binary Data.* Second ed. 1989, Boca Raton: Chapman & Hall.
70. Armitage, P., Berry, G., Matthews, J. N. S., *Statistical Methods in Medical Research.* Fourth ed. 2002, Oxford: Blackwell Science Ltd.

71. Rothman, K. J., Greenland, S., *Modern Epidemiology*. Second ed. 1998, Philadelphia: Lippincott-Raven Publ.
72. Collett, D., *Modelling Survival Data in Medical Research*. First ed. 1994, London: Chapman & Hall.
73. Baltussen, R., Niessen, L. *Priority setting of health interventions: the need for multi-criteria decision analysis*. *Cost Eff Resour Alloc*, 2006; 4: 14.
74. Drummond, M. F., O'Brian, B., Stoddart, G. L., Torrance, G. W., *Methods for the Economic Evaluation of Health Care Programmes*. Second ed. 1997, New York: Oxford University Press Inc.
75. Rice, D. P. *Cost of illness studies: what is good about them?* *Inj Prev*, 2000; 6: 177-9.
76. Luce, B. R., Elixhauser, A., *Standards for Socioeconomic Evaluation of Health Care Products and Services*. 1990, Berlin Heidelberg: Springer-Verlag.
77. Ettaro, L., Songer, T. J., Zhang, P., Engelgau, M. M. *Cost-of-illness studies in diabetes mellitus*. *Pharmacoeconomics*, 2004; 22: 149-64.
78. Drummond, M. *Cost-of-illness studies: a major headache?* *Pharmacoeconomics*, 1992; 2: 1-4.
79. Rice, D. P. *Cost-of-illness studies: fact or fiction?* *Lancet*, 1994; 344: 1519-20.
80. Evers, S. M., Struijs, J. N., Ament, A. J., van Genugten, M. L., Jager, J. H., van den Bos, G. A. *International comparison of stroke cost studies*. *Stroke*, 2004; 35: 1209-15.
81. Hodgson, T. A., Meiners, M. R. *Cost of illness methodology: a guide to current practices and procedures*. *Milbank Mem Fund Q Health Soc*, 1982; 60: 429-62.
82. Henriksson, F., Agardh, C. D., Berne, C., Bolinder, J., Lonnqvist, F., Stenstrom, P., et al. *Direct medical costs for patients with type 2 diabetes in Sweden*. *J Intern Med*, 2000; 248: 387-96.
83. Devore, J., Peck, R., *Statistics. The Exploration and Analysis of Data*. 1993, Belmont: Wadsworth Publishing Company.
84. Sorensen, H. T., Sabroe, S., Olsen, J. *A framework for evaluation of secondary data sources for epidemiological research*. *Int J Epidemiol*, 1996; 25: 435-42.

85. Jansson, J. H., Boman, K., Messner, T. *Trends in blood pressure, lipids, lipoproteins and glucose metabolism in the Northern Sweden MONICA project 1986-99*. Scand J Public Health Suppl, 2003; 61: 43-50.
86. Godfrey, K. M., Barker, D. J. *Fetal nutrition and adult disease*. Am J Clin Nutr, 2000; 71: 1344S-52S.
87. Statistics Sweden. *Flyttningar efter region, ålder och kön. År 1997-2006*. [Migration tables]. Available from: <http://www.ssd.scb.se/databaser/makro/SubTable.asp?yp=dddabbf&xu=96487001&omradekod=BE&huvudtabell=Flyttningar97&omradetext=Befolkning&tabelltext=Flyttningar+efter+region%2C+%E5lder+och+k%F6n%2E+%C5r+1997%2D&preskat=O&prodid=BE0101&starttid=1997&stopptid=2006&Fromwhere=M&lang=1&langdb=1> Accessed 15 March 2007.
88. National Board of Health and Welfare. *Vem får göra vad i hälso- och sjukvården och tandvården. Handböcker för handläggning*. [Allowed activities for different professionals in health- and dental care (In Swedish)]. Stockholm, Sweden: National Board of Health and Welfare, 2004. Available from: <http://www.ambulansforum.se/bb/files/20041015.pdf> Accessed 15 March 2007.
89. Szeto, H. C., Coleman, R. K., Gholami, P., Hoffman, B. B., Goldstein, M. K. *Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic*. Am J Manag Care, 2002; 8: 37-43.
90. Humphries, K. H., Rankin, J. M., Carere, R. G., Buller, C. E., Kiely, F. M., Spinelli, J. J. *Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review?* J Clin Epidemiol, 2000; 53: 343-9.
91. Wilchesky, M., Tamblyn, R. M., Huang, A. *Validation of diagnostic codes within medical services claims*. J Clin Epidemiol, 2004; 57: 131-41.
92. Lee, W. L., Cheung, A. M., Cape, D., Zinman, B. *Impact of diabetes on coronary artery disease in women and men: a meta-analysis of prospective studies*. Diabetes Care, 2000; 23: 962-8.
93. Berkson, J. *Limitations of the application of fourfold table analysis to hospital data*. Biometric Bulletin, 1946; 2: 47-53.
94. Eliasson, M., Lindahl, B., Lundberg, V., Stegmayr, B. *No increase in the prevalence of known diabetes between 1986 and 1999 in subjects 25-64 years of age in northern Sweden*. Diabet Med, 2002; 19: 874-80.

95. Lenzen, M., Ryden, L., Ohrvik, J., Bartnik, M., Malmberg, K., Scholte Op Reimer, W., et al. *Diabetes known or newly detected, but not impaired glucose regulation, has a negative influence on 1-year outcome in patients with coronary artery disease: a report from the Euro Heart Survey on diabetes and the heart.* Eur Heart J, 2006; 27: 2969-74.
96. Drummond, M. F., Botten, G., Hakkinen, U., Pedersen, K. M. *Assessing the quality of Swedish health economics research.* Scand J Public Health, 2006; 34: 566-7.
97. Svensson, L. (The CPP-group, Östergötland County council). *Personal communication.* A.B. Wiréhn. Linköping, Sweden; August 2007.
98. von Ferber, L., Koster, I., Hauner, H. *Medical costs of diabetic complications total costs and excess costs by age and type of treatment results of the German CoDiM Study.* Exp Clin Endocrinol Diabetes, 2007; 115: 97-104.