Risk factors for stroke - with special reference to diet, *Chlamydia pneumoniae* infection and use of non-steroidal anti-inflammatory drugs

Ph D thesis

Søren Paaske Johnsen

Department of Clinical Epidemiology, Aarhus University Hospital and Aalborg Hospital

Department of Epidemiology and Social Medicine, University of Aarhus

Department of Cardiology and Internal Medicine, Aarhus County Hospital

Faculty of Health Sciences, University of Aarhus, 2002

LIST OF ABBREVIATIONS

ATC	Anatomical therapeutical chemical classification system
CT	Computed tomography
FFQ	Semiquantitative food-frequency questionnaire
ICD	International Classification of Diseases
ICH	Intracerebral haemorrhage
IS	Ischaemic stroke
NSAID	Non-steroidal anti-inflammatory drug
OTC	Over-the-counter
SAH	Subarachnoid haemorrhage
TIA	Transient ischaemic attack
TOAST	Trial of Org 10172 in Acute Stroke Treatment

CONTENTS

This PhD thesis is based on the following papers:

I Johnsen SP, Overvad K, Sørensen HT, Tjønneland A, Husted SE. Predictive value of stroke and transient ischemic attack discharge diagnoses in The Danish National Registry of Patients. J Clin Epidemiol 2002; 55: 602-7.

II Johnsen SP, Overvad K, Stripp C, Tjønneland A, Husted SE, Sørensen HT. Intake of fruit and vegetables is associated with reduced risk of ischemic stroke in a cohort of Danish men and women. Submitted.

III Johnsen SP, Overvad K, Østergaard L, Tjønneland A, Husted SE, Sørensen HT. *Chlamydia pneumoniae* seropositivity and risk of ischemic stroke: a nested case-control study. Submitted.

IV Johnsen SP, Pedersen L, Friis S, Blot WJ, McLaughlin JK, Olsen JH, Sørensen HT. Non-aspirin non-steroidal anti-inflammatory drugs and risk of hospitalization for intracerebral hemorrhage: a population-based case-control study. Stroke 2003 (in press).

PREFACE

This PhD thesis is based on studies carried out during my employment at the Department of Clinical Epidemiology, Aarhus University Hospital and Aalborg Hospital, and the Department of Epidemiology and Social Medicine, University of Aarhus, during the period 1998-2002.

I am deeply indebted to a number of persons who have made this work possible. First of all to my supervisors: Kim Overvad for his never-failing engagement in my work and for always providing thoughtful and constructive feed-back, Henrik Toft Sørensen for inimitable inspiration and guidance into the exciting world of epidemiology, and Steen Elkjær Husted for his continuous support and for keeping me focussed on clinically relevant issues.

I am very grateful to Lars Pedersen for always providing skilful assistance and for fruitful discussions about the many challenges within biostatistics.

I am also most grateful to the entire staff running the "Diet, Cancer and Health Study" at The Danish Cancer Society, and in particular to Anne Tjønneland, Connie Stripp, Jytte Fogh Larsen, and Andrea Boll. The studies in this thesis would not have been possible without their invaluable work.

I want to express my sincere gratitude to my other co-authors for their important contributions to the studies in this thesis: Lars Østergaard, Søren Friis, William J. Blot, Joseph K. McLaughlin, and Jørgen H. Olsen.

Special thanks to Svend Juul for always being available with kind advice and useful solutions for problems with data management, to Antonio Sylvester Vethanayagam for help with computer problems, and to John Anderson and Janet Mikkelsen for excellent and careful attempts at making the drafts readable. I thank Jonna Guldberg for providing experienced help with the laboratory analyses of *Chlamydia pneumoniae* antibodies.

I thank the entire staffs at the Department of Clinical Epidemiology, Department of Epidemiology and Social Medicine, and the Danish Epidemiology Science Centre for having provided a unique and stimulating working environment throughout this PhD study, where help and good discussions were never in shortage.

Finally, my most sincere thanks to Cana Jørgensen for being who she is, and to my family for unconditional support and patience at all times.

Århus, July 2002

Sé, nu stiger Solen af Havets Skjød
Luft og Bølge blusser i Brand, I Glød;
hvilken salig Jubel, skjønt Alt er tyst,
medens Lyset lander paa Verdens Kyst

Sé, da stiger Solen af Hav paany,
alle Dødens Skygger for evig fly,
o, for Sejers-Jubel, for salig Lyst:
Lyset stander stille paa Livets Kyst!

Jakob Knudsen (1858-1917)

1. INTRODUCTION

"Our knowledge of disorders of the cerebral circulation and their manifestations is deficient in all aspects"

H. Oppenheim. Opening sentence from German textbook of neurology in 1913 (1)

1.1. What is a stroke ?

A stroke is a clinically defined syndrome "characterised by rapidly developing clinical symptoms and/or signs of focal, and at times global loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin" (2). Stroke is a heterogeneous disease entity that includes several pathologically different conditions (3). Stroke due to cerebral infarction, intracerebral haemorrhage (ICH), intraventricular haemorrhage, and most cases of subarachnoid haemorrhage (SAH) are thus included in the definition. SAH patients with headache but without abnormal neurological signs are according to this definition not considered as stroke patients. By convention, subdural haemorrhage, epidural haemorrhage, retinal infarction, and infarction due to trauma, infection, or tumour masses are not classified as stroke.

Based on the medical history, the clinical examination, and diagnostic tests, including brain imaging, a subclassification of a stroke is possible, e.g. cerebral infarction/ischaemic stroke (IS) may be subclassified as IS due to atherothromboembolism in large or medium-sized arteries, intracranial small vessel disease, cardioembolism, or other causes including rarities such as

arterial dissection, vasculopathies, metabolic disorders, etc.

1.2. The burden of stroke

-*Frequency*

Stroke has a huge impact on public health worldwide. There are about 12,000 new cases of stroke in Denmark each year, one million within the European Union, and 700,000 in the USA (4). About 25% of men and 20% of women can expect to suffer a stroke if they live to be 85 years (5). Stroke is the third commonest cause of death in Europe and the USA, trailing only coronary heart disease and cancer, and is assumed to be related to one in ten of all deaths (6). The exact figures are not available for the developing countries, but the worldwide estimate of five million stroke-related deaths in 1990 has been predicted to double by 2020, with most of the increase occurring in developing countries because of the expected major demographic and lifestyle changes, including increased prevalence of smoking, obesity and diabetes mellitus, in these countries (5).

-*Consequences*

Stroke is also a major cause of chronic disability. After one year, about a third of all stroke survivors are functionally dependent in daily activities due to sequelae, including hemiparesis, aphasia, etc., making stroke the most common cause of severe disability in the developed countries (7). Moreover, stroke is also associated with dementia, depression, epilepsy, and secondary medical problems including falls and fractures (6)

The economic consequences of stroke are considerable; about 4% of all expenses within Danish hospitals are used for treatment and care of patients with stroke (8). However, the total costs, i.e. direct and indirect costs associated with treatment and value of lost productivity, are much larger and have been estimated to exceed $40 billion each year in the USA alone (9).

-Prevention

The importance of prevention is of particular importance in relation to stroke since there is currently no widely applicable treatment for these patients apart from supportive care and treatment of acute complications in specialised hospital units (10). Thrombolysis is a potential curative treatment for patients with IS (11), however, although promising, widespread use of this treatment has proved to be difficult for a number of reasons including the short treatment time window, i.e. within 3 hours from onset of symptoms, and the need for early expert assessment (11,12).

1.3. Difficulties in studying risk factors for stroke

Stroke is a multifactorial disease entity when looking at the aetiology. Furthermore, the available knowledge about the pathophysiologic mechanisms is limited. This makes identification of factors causally related to the disease a challenge. Studies on stroke have furthermore been hampered by a number of problems:

1.3.1. Crude classification of stroke events

The traditional approach to observational epidemiologic studies and randomised trials has been to lump all types of stroke together. This approach is unfortunate because it is likely to weaken or even obscure relationships between a potential risk factor and a particular type of stroke. A positive association between a potential risk factor and a specific type of stroke may not be identified if no association or even a negative association is seen for other types of stroke. A dilution bias is consequently introduced into the study of in this situation, i.e. a bias that dilutes the "real" association between the putative risk factor and the disease. The apparent opposing associations between serum cholesterol and IS and ICH, respectively, provide a classic

example of this problem. Consequently, epidemiological studies carried out before brain imaging and valid classification of patients were available should be interpreted with caution since the crude classification of the stroke event probably introduced bias in several studies. There has been increasing awareness of this issue within stroke epidemiology in recent years, and a recognition of the need to use a more detailed classification of stroke. However, application of this new approach is not without problems, since the different classification systems rely on the availability of diagnostic tools that are sometimes missing (3). Furthermore, the validity of the stroke classification may in some situations be uncertain, eg due to co-existing causes of stroke, and splitting the cases in several subgroups may limit the sample size to such an extent that even large patient materials become inconclusive (3).

1.3.2. Risk of selection bias

Stroke patients who do not survive long, eg patients with SAH or ICH, may have a lower chance of being included in a study. This situation, which is seen particularly in cross-sectional and case-control studies, may cause problems in relation to studies on putative risk factors, if exposure to the factor in question is related to the prognosis after stroke. Thus, if exposure to a factor causes the stroke to be rapidly fatal, the prevalence of the exposure will be low among cases, and the exposure will appear protective, even though the factor has no effect on disease risk.

1.3.3. Risk of inverse causality

Stroke itself or the required treatment may in some situations change the risk factor, eg a temporary increase in both blood pressure and blood glucose are often seen following a stroke, whereas plasma cholesterol falls (13).

1.3.4. Problems with data collection

Using the case-control study design it may be difficult or impossible to obtain information on past activities and habits, e.g. diet, smoking, and physical activity, because of the patient's confusion or aphasia (13). This problem can be avoided in follow-up studies and case-control studies based on prospectively collected data.

1.3.5. Lack of blinding

Due to the frequent and often severe consequences of stroke, eg hemiparesis, aphasia and impairment of cognitive functions, it may be impossible to use blinding, typical of an interviewer, as a tool for reducing the risk of bias in studies on patients with stroke (13).

Some of these difficulties may be avoided by using prospectively collected data. However, availability of prospective data will not necessarily solve all problems since the issues of surveillance bias, exposure misclassification, and definition of the relevant exposure time window data will often be relevant.

1.4. Use of administrative registers in stroke research

The considerable burden of stroke on the patients, their families and the society, and the development of improved diagnostic tools within the last decades have prompted an epidemiological effort to describe the incidence of stroke, to examine risk factors, and to study the prognosis. A substantial part of this work has been done within large cohort studies in which at least some of the methodological problems mentioned above are more easily handled (14,15). In order to provide valid risk estimates for these cohort studies, it is essential to have complete follow-up or at least non-differential loss of follow up, and avoidance of surveillance

bias, ie systematic errors in methods of ascertainment, diagnosis, or verification of outcome events. Use of hospital discharge and other disease registers for follow-up of study participants has proved a valuable, cost-efficient, and complete way of collecting data on stroke events (16-18). The registries have the advantage of readily available data, often completeness of registration of people in the target population, and reduced likelihood of some types of bias, e.g. recall, non-response, and effect on the diagnostic process caused by the research question, although surveillance bias may also occur in clinical practice and thus become affect routinely collected data (16,19). However, the fact that the collection and quality of the data are not under the control of the researcher may sometimes represent important disadvantages of registry data (3). These disadvantages, which vary considerably from disease to disease, may have serious implications for the use of registry data in epidemiological research. Insufficient quality and completeness of data may be particularly troublesome when estimating absolute risk measures, whereas relative risk measures are usually less affected by non-differential shortcomings in completeness and quality, although the misclassification will usually tend to bias the associations toward the null hypothesis. Calibration studies, i.e. studies that evaluate the validity of data by comparison with independent external criteria, are essential to avoid invalid conclusions when using secondary data collected as part of routine daily clinical work (20). Sensitivity and specificity, ie the probabilities of correctly identifying diseased and non-diseased persons, respectively, are important properties when classifying events or persons. However, it is rarely possible to establish these measures for secondary data sources since typically there are no available reference data to compare with. Alternatively, the validity of a data source, ie the extent to which the data measures what they are intended to measure, may be assessed through evaluation of the predictive value of a positive registration.

A number of studies have examined the data quality of stroke or transient ischaemic attack

(TIA) diagnoses in hospital discharge registers and official mortality statistics (21-29) (**Table 1.1.**). The studies have primarily provided information on the positive predictive value of a stroke diagnosis, and seldom on the completeness of the data source. Although most studies have reported the positive predictive value of a stroke diagnosis as moderate to high, few studies have been able to provide information on the predictive value of diagnoses of specific types of stroke, including SAH, ICH, IS, and TIA (22,23,26,28,29). Furthermore, only few studies have provided information on the diagnostic tools used to ascertain the diagnoses (22). The diagnostic strategy and management of stroke patients have undergone important changes during the last decades, eg increased use of imaging procedures and formation of specialised stroke units. These changes presumably have implications for the diagnostic work up and validity of cerebrovascular diagnoses, including stroke and TIA. Thus, there is a need for up-to-date studies on the validity of cerebrovascular discharge diagnoses, in particular from the Nordic countries, which have some of the most valuable population-based registries for epidemiological research, due to the extensive possibilities of record linkage (30,31).

Table 1.1. Recent studies on the predictive value of stroke diagnoses within administrative registers.

Authors	Country	Study period	Data source	Method used	Number of cases	Predictive value
Iso et al (21)	USA	1970-80	Hospital discharge record/ death certificates	Record review	408 fatal events	stroke: 97.7%
Leppälä et al (22)	Finland	1985-92	Hospital discharge register	Record review	375 non-fatal events	SAH: 78.6% ICH: 82.6% IS: 90.1%
			Register of causes of death	Record review	218 fatal events	SAH: 95.2% ICH: 91.1% IS: 92.4%
Liu et al (23)	Canada	1990-91	Hospital discharge records	Record review	1,494 events	SAH: 92.9% ICH: 92.1% IS: 86.6%
Leibson et al (24)	USA	1970-89	Hospital discharge records	Linkage with stroke register/record review	377 events	stroke: 79.3%
Stegmayr, Asplund (25)	Sweden	1985-89	Hospital discharge register	Linkage with stroke register	5,101 events	stroke: 68.5%
Lindblad et al (26)	Sweden	1977-87	Hospital discharge register	Record review	251 non-fatal events	SAH: 78.3% ICH: 55.0% IS: 75.9%

			National mortality register	Record review	53 fatal events	stroke: 69.2%
Mähönen et al (27)	Finland	1983-89	Hospital discharge register	Linkage with stroke register	-	stroke: 87-93%
Authors	Country	Study period	Data source	Method used	Number of cases	Predictive value
Ellekjær et al (28)	Norway	1994-96	Hospital discharge records	Linkage with stroke register	508 events	SAH: 69.2% ICH: 71.4% IS: 65.8%
Gaist et al (29)	Denmark	1977-95	Hospital discharge register	Record review	210 events	SAH: 93.0%

1.5. Risk factors for stroke

A number of factors have consistently been shown to be associated with an increased risk of stroke. The association with stroke for these factors is supported by a large body of experimental and epidemiological scientific work. Although the indications for associations are therefore strong for these factors, the question of whether the associations are causal remains open for some of the factors. These well-documented risk factors for stroke are usually divided into modifiable and non-modifiable factors, ie factors that are amenable for intervention and factors that are not (**Table 1.2.**) (13,15).

Table 1.2. Established risk factors for stroke.

	Non-modifiable risk factors:	
	Age	
	Male gender	
	Ethnicity	
	Modifiable risk factors:	
	Hypertension	
	Smoking	
	Alcohol intake	
	Diabetes mellitus*	
	Atrial fibrillation*	
	Other cardiac diseases*	
	Carotid artery stenosis*	
	Oral anticoagulants#	
	Thrombolytic treatment#	

*Associated with IS

#Associated with ICH and SAH

Hypertension and smoking seem to be the most important of the modifiable risk factors, due to the relatively strong associations observed for these factors (13,15), even when controlling for other risk factors, and the high prevalence within most populations.

Despite the importance of the well-documented risk factors, other factors are also likely to be of importance. Thus, the well-established risk factors are not necessary components in the causation of all stroke events, as illustrated by the occurrence of stroke also in non-exposed subjects in the risk factor studies (13,15). Furthermore, it is common clinical experience that a substantial proportion of patients admitted with stroke are apparently not exposed to any of the established risk factors. Other factors and yet unknown or at least insufficiently studied mechanisms are thus likely to be of importance in relation to stroke. It is therefore not surprising that a large number of factors have been examined and in several studies reported to be associated with the risk of stroke (3,13). It is likely, that a high proportion of these putative causes of stroke are not on the causal pathway, ie the direct series of events or mechanisms linking cause and effect, or that they have been refuted later in larger and methodologically stronger studies, thereby contributing to what critics of modern epidemiology have termed "risk-factorology" or "black-box epidemiology" (32), i.e. the propensity of risk factor epidemiology indiscriminately to identify particular aspects of daily life as dangerous to health (33). However, several potential risk factors are under continual examination based on supportive findings from both experimental and observational studies and biologically plausible hypotheses regarding disease mechanisms. These potential risk factors include a range of dietary factors, e.g. intake of fruit and vegetables (34) and fish (35), other lifestyle factors, eg alcohol intake, physical activity and obesity (13,15,36), and various other factors, eg use of

different drugs (non-steroidal anti-inflammatory drugs (NSAIDs) (37), phenylpropanolamine (often found in appetite suppressants or cold remedies) (38), and hormone replacement therapy for post menopausal women (13,15)), infections (39), homocysteine levels (40), antiphospholipid antibodies (13,15), and migraine (13,15). Although modest relative risk estimates have been reported for most of these potential risk factors, in particular for lifestyle factors such as diet, physical activity, and obesity, the high prevalence of these factors within most populations indicates that the attributable risk of stroke is probably substantial for these potentially modifiable factors.

A short overview of the literature on the three potential risk factors examined in this thesis, ie intake of fruit and vegetables, infection with *Chlamydia pneumoniae*, and use of non-aspirin NSAIDs, is provided in the following section.

1.5.1. Intake of fruit and vegetables

The association between diet and the risk of stroke has been examined since migrant studies and studies of secular trends in rates of stroke provided evidence that lifestyle factors, possibly including diet, play an important role in the etiology of stroke (41-43).

Most analytical epidemiological studies to date have focused on specific dietary components rather than on foods (34,35,44-51). Among the dietary components most consistently reported to be associated with a reduced risk of stroke are antioxidants, ie vitamin C, β-carotene, and flavonoids, potassium, and fibre (44-48). When comparing people with a high dietary intake vs. people with a lower dietary intake of the food component in question, typical relative risk estimates in the observational studies have been in the range 0.3–1.0. These findings have led to increased interest in foods that are rich in these substances, such as fruit, vegetables, and

other plant foods (34,51-53). This interest in foods is further stimulated by previous experiences, eg the negative results from several large randomised trials on antioxidant vitamins, which have demonstrated that, when trying to unravel the complexity in the relation of foods to health, focusing on single nutrients might be too simple an approach (54,55).

The association between intake of fruits and vegetables and the risk of stroke has only been examined to date in a limited number of individual-level studies (**Table 1.3.**) (34,46,51,56,57). All of these studies have been based on the follow-up study design. The most detailed data have been reported from the well-known cohort studies: the Framingham Study and the Nurses' Health Study/ the Health Professionals' Follow-up Study (34,51). Gillman et al. analysed the association between intake of fruit and vegetables as assessed by a single 24-hour recall interview at base-line and the risk of stroke in 832 men during a 20-year follow-up period (34). Each increment of three servings per day of fruit and vegetables was associated with a relative risk of stroke of 0.75 (95% CI: 0.57-1.00) (34). Joshipura et al examined the association in 75,596 women and 38,683 men using dietary information from semiquantitative food frequency questionnaires. Each increment of one serving per day of fruit and vegetables was associated with a relative risk of IS of 0.94 (95% CI:0.90-0.99) (51). However, examining the findings from all the analytical epidemiological studies, both inverse (34,51,57) and no associations (46,56) have, however, been reported. Thus, there is still uncertainty as to the existence of any preventive effect of a dietary intake of fruit and vegetables. Furthermore, most of the studies have several limitations since they were based on crude assessments of type and amount of diet, and provided only limited adjustment for confounding factors. Only one study reported on the association between intake of specific fruits and vegetables and risk of stroke (51). The reduced risk of stroke in that study was most

Table 1.3. Follow-up studies on intake of fruits and vegetables and risk of stroke.

Authors	Country	Year	Dietary intake	Size	Results
Hirayama (56)	Japan	1986	Green and yellow vegetables daily vs. less than daily	Participants: 265,118 Stroke events: -	No association
Gillman et al (34)	USA	1995	Increment of three servings per day of fruit and vegetables	Participants: 832 Stroke events: 73	adj.RR: 0.75 (95% CI:0.57-1.00)
Key et al (57)	United Kingdom	1996	Fresh fruit daily vs. less than daily	Participants: 10,771 Stroke events: 147	adj.RR: 0.68 (95% CI:0.47-0.98)
Keli et al (46)	The Netherlands	1996	Top vs. bottom quartile of mean daily intake of fruit and vegetables	Participants: 552 Stroke events: 42	adj.RR (fruit): 0.52 (95% CI:0.21-1.31) adj.RR (veg.): 0.82 (95% CI:0.35-1.94)
Joshipura et al (51)	USA	1999	Top vs. bottom quintile of mean daily intake of fruit and vegetables	Participants: 114,279 Stroke events: 570	adj.RR: 0.69 (95% CI:0.52-0.92)

evident for intake of cruciferous vegetables, green leafy vegetables, citrus fruit including juice, and citrus fruit juice alone. Some studies reported only on fatal cases of stroke, and few studies classified the strokes as IS, ICH, or SAH. No study has provided information on the risk of different subtypes of stroke, eg IS due to atherothromboembolism in large or medium-sized arteries, intracranial small vessel disease, cardioembolism, etc. Finally, several studies included only few outcomes, resulting in low statistical precision of the risk estimates and further hampering the possibilities of sufficient confounder adjustment.

The issue of dietary intake of fruit and vegetables in relation to cardiovascular disease, including stroke and hypertension, has also been examined in a few clinical trials on patients with preexisting cardiovascular disease. The trials have provided evidence for a beneficial effect on cardiovascular risk of a diet with a high content of plant foods, but only very few of the patients developed a stroke in these trials resulting in imprecise risk estimates as indicated by the wide confidence intervals (58-60).

In summary, findings from different studies, including several observational and interventional studies, indicate that a high dietary intake of fruit and vegetables may be associated with a reduced risk of stroke. However, there is need for further relatively large studies including carefully designed and conducted observational epidemiological studies with detailed assessment of dietary exposures, potential confounders, and outcomes.

1.5.2. *Infection with Chlamydia pneumoniae*

There is increasing evidence that infectious agents, most consistently *Chlamydia pneumoniae*, but also *Helicobacter pylori*, herpes simplex virus, and cytomegalovirus, may be risk factors for atherosclerotic disease (61,62). These organisms may be involved in the atherogenesis and

act as triggers of acute clinical events, e.g. by destabilising pre-existing atherosclerotic plaques. Different pathophysiological mechanisms have been proposed to mediate the association between the various infections agents and the risk of atherosclerotic disease, including induction of a systemic or local inflammatory response, autoimmune reactions, and changes in conventional cardiovascular risk factors, eg lower levels of HDL cholesterol and higher levels of fibrinogen and triglycerides (62).

Most human studies have examined the role of *Chlamydia pneumoniae* in relation to coronary heart disease (62,63), whereas only few and less detailed data are available for the risk of cerebrovascular disease. Six seroepidemiological studies, primarily case-control studies, have reported on the association between seropositivity to *Chlamydia pneumoniae* and the risk of stroke (**Table 1.4.**) (39,64-68). Most of these studies (39,64-66), but not all (67,68), found that increased *Chlamydia pneumoniae* titers were positively associated with the risk of stroke. However, there is large variation in the reported relative risk estimates, ranging from 0.4 to 8.6. Likewise, some (69-72), but not all (73,74), studies have reported an association between seropositivity to *Chlamydia pneumoniae* or presence of *Chlamydia pneumoniae* and carotid atherosclerosis.

Table 1.4. Seroepidemiological studies on seropositivity to Chlamydia pneumoniae and risk of stroke

Authors	Year	Design	Cut-off value	Results
Wimmer et al. (39)	1996	Case-control (58 cases with IS/TIA, 52 controls)	IgA >= 1:16 IgG >= 1:32 IgG in IC >=1:8	adj. OR = 1.71 (95% CI:1.08-2.70) adj. OR = 0.81 (95% CI:0.49-1.33) adj. OR = 2.00 (95% CI:1.07-3.76)
Cook et al (64)	1998	Case-control (176 cases with stroke/TIA, 1518 controls)	IgG>=1:512 and/or IgM>=1:8 1:64=<IgG<1:256 or IgA>=1:8	adj. OR = 4.2 (95% CI:2.5-7.1) adj. OR = 4.4 (95% CI:3.0-6.5)
Fagerberg et al (65)	1999	Follow-up study (15 stroke events)	IgG>=1:512 and/or IgA>=1:64	adj. RR = 8.58 (95% CI:1.07-68.82)
Elkind et al (66)	2000	Case-control study (89 cases with IS, 89 controls)	IgA>=1:16 IgG>=1:16 IgG>=1:32	adj. OR = 4.51 (95% CI:1.44-14.06) adj. OR = 2.59 (95% CI:0.87-7.75) adj. OR = 2.50 (95% CI: 0.87-7.21)
Glader et al (67)	1999	Nested case-control (101 cases with IS, 201 controls)	IgA>=1:16 IgG>=1:32	adj. OR = 0.9 (95% CI: 0.5-1.6) adj. OR = 0.4 (95% CI: 0.2-0.9)
Heuschmann et al (68)	2001	Case-control (145 cases with IS, 260 controls)	IgG>=1:64	adj. OR = 0.86 (95% CI: 0.44-1.67)

This inconsistency in the reported associations between *Chlamydia pneumoniae* infection and cerebrovascular disease, in particular stroke, may be caused by several factors. Use of different study designs, eg follow-up, case-control and nested case-control design, is likely to have influenced the risk of bias in the individual studies. Different laboratory methods, uncertainties about the selection and interpretation of cut-off values for antibody titers, and differences in the available data for adjustment of confounding are also likely to have contributed to the inconsistency. Furthermore, changes in exposure status during the study period and use of antibiotics may have influenced the results. Most studies included only few outcomes, and the statistical precision of the findings was consequently moderate to low, particularly when attempts were made to distinguish between subtypes of stroke (64,68).

Furthermore, differences in the prevalence of the causal complement, ie factors that act together with *Chlamydia pneumoniae* infection in the putative causation of stroke, in the different study populations may also have given inconsistent risk estimates.

The potential risk of stroke, particularly IS, associated with *Chlamydia pneumoniae* infection therefore remains uncertain. Since the potential risk of stroke associated with *Chlamydia pneumoniae* infection may possibly be reduced or even removed by relevant antibiotic treatment, further data on the possible association between *Chlamydia pneumoniae* infection and the risk of stroke are important. Thus, there is need for further well-designed studies examining the associations, taking into consideration some of the weaknesses of the earlier studies, including the issues of subtypes of stroke, insufficient confounder control, and cut-off values for antibody titers.

1.5.3. *Non-steroidal anti-inflammatory drugs*

NSAIDs are among the most frequently used medications worldwide. NSAIDs are primarily

used for analgetic, antipyretic, and antithrombotic purposes; however, the use may be extended even further in the future since there is an increasing amount of scientific data suggesting that NSAIDs may have a role in the prevention of cancer (75).

The use of NSAIDs has been associated consistently with an increased risk of bleeding, particularly from the gastrointestinal tract (76-78). The increased risk is probably due to decreased synthesis of prostaglandins, which, among other consequences, leads to impairment of gastrointestinal mucosal integrity and platelet aggregation (76,79). Although most pronounced for aspirin, all NSAIDs affect platelet aggregation and bleeding time (80).

Several randomised trials and observational studies have investigated whether the use of NSAIDs increase the risk of ICH (37,81-84). Most of these studies have focused on the use of aspirin (81-84). An increased risk of ICH was seen in a recent meta-analysis based of 16 trials on aspirin treatment for primary or secondary prevention of cardiovascular disease, including stroke (81). In this analysis of 55,462 persons and 108 cases of ICH, use of aspirin was associated with a relative risk of ICH of 1.87 (95% CI: 1.24-2.74) and an absolute increase in ICH of 17 events per 10,000 patients per year (95% CI: 7-28) (81). It is unclear whether the risk was equally distributed over time, or was highest immediately after the start of treatment. However, the increased risk of ICH was outweighed by an even greater reduction in the risk of acute myocardial infarction and IS (fatal and non-fatal events) (81).

There are important differences between aspirin and the non-aspirin NSAIDs with respect to the pharmacological effects of the drugs and the indications for which they are used. Thus, extrapolating the findings on aspirin to other NSAIDs may not be possible. Even fewer data are available on the use of non-aspirin NSAIDs and risk of ICH. Thrift et al recently examined the association between intake of aspirin and non-aspirin NSAIDs and the risk of ICH in a case-

control study based on 331 consecutive patients hospitalised with primary ICH and 331 age and sex matched community based neighbours (37). Information on drug use within the 2 weeks preceding the ICH was obtained by a questionnaire administered to all subjects either directly or by proxy with the next of kin. The investigators found no evidence of a strong association between use of either aspirin (OR=1.00 ;95% CI:0.60-1.66) or non-aspirin NSAIDs (OR=0.85 ; 95% CI:0.45-1.61) and the risk of ICH (37). However, although the study was carefully planned and carried out, the number of cases was not sufficient to exclude the presence of a moderately increased risk.

Since non-aspirin NSAIDs are so widely used, even small risks of side effects, especially for serious conditions such as ICH, may have considerable clinical and public health implications. Furthermore, even if an overall increased risk of ICH can be excluded, substantial risks may still be present among subgroups of persons. Such a situation may obviously have important clinical implications for these persons and should therefore be examined further. Thus, further information is needed, including data from subgroups of persons with an increased base-line risk of ICH, eg elderly and hypertensive patients, before an increased risk of ICH associated with use of non-aspirin NSAIDs can be excluded.

1.5.4. Summary on putative risk factors

In conclusion, other factors apart from the already well-established risk factors are likely to be causally associated with the risk of stroke. These putative factors include dietary intake of fruit and vegetables, infection with *Chlamydia pneumoniae,* and use of non-asprin NSAIDs. However, further data are needed before the role of these factors can be determined and the potential of intervention against these factors can be further explored. Important weaknesses are present in several of the existing studies where these factors have been examined. These

weaknesses include a risk of selection and information bias, in particular in the studies on *Chlamydia pneumoniae* and NSAIDs, the likely presence of unadjusted confounding, most prominent in the studies on fruit and vegetables and *Chlamydia pneumoniae*, and a considerable statistical imprecision of the obtained risk estimates. Further efforts should be taken to minimise the risk of these weaknesses in future studies.

1.6. AIMS OF THESIS

The aims of this thesis were:

1. To examine whether routinely collected data from the health care system in Denmark could be used as a tool for stroke epidemiology, with special reference to the data quality of hospital discharge diagnoses of stroke and TIA.
2. To examine the association between intake of fruit and vegetables and the risk of IS.
3. To examine the association between previous/chronic infection with *Chlamydia pneumoniae* and the risk of IS.
4. To examine the association between use of non-aspirin NSAIDs and the risk of ICH.

2. DATA SOURCES AND DEFINITION OF OUTCOMES

2.1 Data sources

The studies included in this thesis were based on data from a number of different sources including the Danish follow-up study "Diet, Cancer and Health", the Danish National Registry of Patients, the Pharmaco-Epidemiological Prescription Database of North Jutland County, and the Civil Registration System. Linkage between these data sources is possible using the civil registry number, a unique, personal 10-digit identification number given to all Danish citizens at birth or on immigration since 1968.

"Diet, Cancer and Health"

The study "Diet, Cancer and Health" is a prospective cohort study with the primary aim of studying the etiological role of diet in cancer risk. The study design is described in detail elsewhere (85). From December 1993 to May 1997, 80,996 men and 79,729 women aged 50 to 64 years were invited to participate in the study; 27,177 men and 29,876 women accepted the invitation. Eligible cohort members were living in the Copenhagen and Aarhus areas, born in Denmark, and had no previous cancer diagnosis in the Danish Cancer Registry.

All cohort members completed a detailed 192-item semiquantitative food-frequency questionnaire (FFQ). Descriptions of the development and validation of the questionnaire have been published previously (86,87). All the participants also completed a questionnaire about other lifestyle factors such as smoking habits, alcohol intake, physical activity, medical history, education, and, for women, use of hormone replacement therapy. Body proportions, blood pressure, and total serum cholesterol were measured at baseline, and samples of biological material, including blood samples, were obtained and stored in a biological bank.

The Danish National Registry of Patients

The Danish National Registry of Patients, established in 1977, records 99.4% of all discharges from non-psychiatric hospitals in Denmark (88). Data on out-patients have been recorded since 1994. The data include the civil registry number, dates of admission and discharge, surgical procedures performed, and up to 20 discharge diagnoses, classified until 1993 according to the Danish version of the International Classification of Diseases, 8th Revision (ICD-8), and subsequently according to the corresponding national version of ICD-10 (88). All discharge diagnoses are assigned exclusively by the physician who discharges the patient.

The County Hospital Patient Register (HPR) of North Jutland County, which is used in study IV, is a patient administrative data system which collects data for administrative purposes at county level including data for the National Registry of Patients. This register is updated on a daily basis.

The Pharmaco-Epidemiological Prescription Database of North Jutland County

The population-based Pharmaco-Epidemiological Prescription Database of North Jutland County (89) covers the entire population of North Jutland County, which comprises about 500,000 inhabitants, approximately 9% of the total Danish population. All pharmacies in the county are equipped with a computerised accounting system by which data are sent to the Danish National Health Service as part of a national tax-supported health care programme. This programme refunds 50-75% of the costs associated with the purchase of most drugs prescribed by doctors. In North Jutland, this accounting system also provides key data on prescriptions for refundable drugs for the Pharmaco-Epidemiological Prescription Database.

This database, which was initiated on January 1, 1989 and has complete coverage from January 1, 1991, includes the civil registry number of the patient, type of drug prescribed according to

the anatomical therapeutical chemical (ATC) classification system (90), and date on which the drug was dispensed. The database does not contain information on in-hospital drug use, non-refundable drugs, or over-the-counter (OTC) drugs. However, certain non-reimbursable drugs can be reimbursed for pensioners and regular users of these drugs, e.g. patients with chronic diseases, and are thus partly registered in the database.

The Civil Registration System

The Civil Registration System has kept electronic records of all changes in vital status for the entire Danish population since 1968, including change in address, date of emigration, and date of death.

2.2. Definition of outcomes

Stroke (or subtypes of stroke) or TIA were the outcomes in all the studies included in this thesis:

The World Health Organization's definition of stroke was used when verifying discharge diagnoses by medical record review, i.e. an acute disturbance of focal or global cerebral function with symptoms lasting more than 24 hours or leading to death of presumed vascular origin (2). To distinguish between SAH, ICH, and IS in this thesis, a computed tomography (CT) or magnetic resonance (MR) scan, a spinal fluid examination, or an autopsy or operation description was necessary. A diagnosis of SAH was made when typical clinical symptoms and at least one of the following findings were described: CT findings compatible with blood in the subarachnoid space, xanthochromic cerebrospinal fluid, or SAH or aneurysm established by angiography or autopsy. ICH was defined as clinical symptoms of stroke combined with the presence of a CT/MR-verified ICH or ICH established by autopsy and not preceded by a

previous IS in the same area. A diagnosis of IS was made when an infarction or haemorrhage could not be visualised by CT/MR scanning but the symptoms were consistent with stroke.

We subclassified all cases of IS on the basis of the presumed etiology according to the Trial of Org 10172 (a low-molecular-weight heparinoid) in Acute Stroke Treatment (TOAST) classification: large-artery atherosclerosis, cardioembolism, small-vessel occlusion, stroke of other determined etiology, and stroke of undetermined etiology (91). This classification is based on clinical features, i.e. cortical or cerebellar dysfunction, lacunar syndrome, and on data collected by tests such as brain imaging, i.e. location and size of infarct, cardiac imaging, duplex imaging of extracranial arteries, arteriography, and laboratory assessments for a prothrombotic state.

A modified version of Kraaijeveld et al's criteria, which are based on the time course and the symptoms of the event, including symptoms explicitly not acceptable as TIA, was used for the diagnosis of TIA (92,93). Thus, the attack had to include an element of transient focal disturbance of the blood supply to the brain or eye. Transient global amnesia and migraine attacks were not considered as TIAs.

Only hospitalised events of stroke and TIA were considered as outcomes in the studies in this thesis. We made no distinction between fatal and non-fatal cases.

3. STUDY I: PREDICTIVE VALUE OF STROKE AND TIA DIAGNOSES

3.1. Subjects and methods

The study population in this study, where routinely coded hospital discharge diagnoses were compared with diagnoses based on information obtained from medical records, comprised all participants within the "Diet, Cancer and Health" study registered with a first-time discharge diagnosis of stroke/TIA (ICD-10 codes: I60-69.8 and G45) in the Danish National Registry of Patients during follow-up, i.e. until 31 December 1998 for participants living in the Copenhagen area and until 31 December 1999 for participants living in the Aarhus area. We excluded participants who had been hospitalised before enrolment with cardiovascular diseases, i.e. stroke, TIA, ischaemic heart disease, or peripheral arteriosclerosis (ICD-8 codes:410-414, 430-438, 440, ICD-10 codes: G45, I20-25, I60-70).

Patients not admitted to hospital, i.e. patients with a clinically "silent" stroke or patients dying of stroke before admission, were not included in the study. Furthermore, patients hospitalised with stroke were not included if the discharge diagnosis was not coded as a cerebrovascular condition.

Medical records and hospital discharge letters were retrieved and reviewed for all patients, using the previously defined criteria for stroke and TIA. The record reviews were done using a detailed standardised form without knowledge of the data collected at baseline in "Diet, Cancer and Health" (except for age and sex). The form was developed and tested in close collaboration with a consultant with extensive experience within the field of stroke medicine. All cases with an uncertain diagnosis based on the available information were discussed with the consultant. The review was based on all the available information in the medical records, including the written radiology reports, results from laboratory tests, etc. The actual brain imaging films were not re-interpreted. The available information was identical in most cases with the information

available to the physician at the time of discharge.

3.2. Data analysis

The positive predictive value of the ICD codes recorded in The National Registry of Patients was expressed as proportions, i.e. the numerator containing the number of participants with confirmed diagnoses after review of medical records using the above-mentioned criteria for a diagnosis of stroke as the reference standard, and the denominator containing the total number of participants registered in The National Registry of Patients with the specific diagnosis (16). Furthermore, the data were stratified by type of discharging department.

Patients whose records could not be found were excluded from the analyses. If a participant was registered with both a stroke and a TIA diagnosis in The National Registry of Patients, the stroke diagnosis had preference.

To estimate confidence intervals and compare proportions, we relied on the normal approximation of the binomial distribution.

3.3. Results

"Diet, Cancer and Health" included 57,053 persons. After exclusion of 2,500 participants (4.4%) with previously diagnosed cardiovascular disease, the study population included 25,494 men and 29,059 women. Median length of follow-up was 3.06 years (range 0.02-5.10 years). Medical records could be retrieved and validated for 377 of 389 (96.9%) stroke cases, 134 of 137 (97.8%) TIA cases, and 54 of 55 (98.2%) cases with other cerebrovascular diseases. We found no systematic characteristics for the patients for whom medical records could not be retrieved, i.e. these patients did not differ from the remaining patients with respect to age, sex, type of discharge diagnosis, and type of discharging department.

Among the 377 cases of stroke (SAH, ICH, IS, or unspecified stroke) coded in The National Registry of Patients, 79.3% (95% CI: 74.9-83.3%) of the diagnoses were confirmed after review. Among the 134 participants registered with TIA, 60.4% (95% CI: 51.6-68.8%) of the diagnoses were confirmed.

The predictive value differed between the stroke subgroups, i.e. discharge diagnoses of SAH and ICH were confirmed in 48.3 % (95% CI: 29.4-67.5%) and 65.7% (95% CI: 47.8-80.9%) of the cases, respectively, whereas IS and unspecified stroke were confirmed in 87.7% (95% CI: 80.1-93.1%) and 76.0% (95% CI: 69.5-81.7%) of the cases, respectively.

A wide range of medical conditions were identified among the cases for which a discharge diagnosis of stroke could not be verified, including cancer with cerebral metastases (n=11), migraine (n=3), epilepsy (n=4), subdural haematoma (n=3), meningitis (n=2), and dementia (n=2). Most cases with unverified TIA discharge diagnoses suffered from vertigo or other non-specific symptoms (n=35).

Overall, imaging procedures (CT or MR) had been used in 421 of 445 cases (94.6%). By contrast, autopsy was only conducted in 5 out of 21 fatal cases (23.8%).

The predictive value of the discharge diagnoses differed according to type of discharging department. Diagnoses from emergency rooms had a lower overall predictive value (48.8%, 95% CI: 39.9-57.8) than diagnoses from non-speciality departments, e.g. departments of internal medicine (and in some cases stroke units) (68.8%, 95% CI: 61.3-75.5), and speciality departments, i.e. departments of neurology or neurosurgery (77.9%, 95% CI: 72.3-82.7). This

trend was consistent within the different diagnostic subgroups.

The predictive value did not differ according to age or sex.

3.4. Strengths and weaknesses of the study

Interpretation of the findings from an analytical epidemiological study depends on a critical evaluation of potential alternatives to a causal association, i.e. selection problems, information problems, confounding, and statistical imprecision.

Only when these alternative explanations can be excluded, or at least thought to be of minor importance, are we able to consider causal inference.

The issues of selection and information problems, confounding and statistical precision will be discussed for each of the studies I-IV in this thesis. Please note that study I is not a conventional analytical epidemiological study, and that a thorough discussion of all of these issues is not applicable for this study.

Selection problems

Selection problems include selection mechanisms that may occur during sampling of the study subjects and selection biases. Selection problems may thus influence both the external and the internal validity of a study. The external validity is usually defined as the degree to which a study is able to produce unbiased inferences beyond the subjects in the study (95). Evaluation of the external validity of a study usually involves considerable subject-matter judgment based on an understanding of which conditions are relevant and which are irrelevant to the generalisation (95,96). The external validity is therefore not merely a technical question of the study population being representative of the target population, but also requires that other issues are considered, eg the patophysiological mechanisms underlying the disease in question and the presence or absence of relevant biological differences between the study population and the target population (96).

Selection problems might have been introduced at different levels of this study, ie when sampling the study population or during identification of the cases.

-Sampling

The study population was a cohort of middle-aged men and women living in urban areas and participating in a follow-up study on lifestyle factors. The observed number of cases of stroke in

the study population was approximately 35% lower than expected when comparing with age-specific incidence rates of stroke in the general population (97). This indicates that the participants were probably not a random sample of the general population. Thus, the study population may not be suited for estimating absolute measures of disease frequency, whereas relative risk measures are often less affected by this selection process. Furthermore, the validity of a discharge diagnosis of stroke or TIA could be different for patients not participating in studies, elderly patients, or patients from rural areas. A lower validity for these groups may be caused both by differences in coding practice, e.g. in rural areas, and by the amount of diagnostic work-up, e.g. among elderly patients. However, the coding of discharge diagnoses was probably independent of participation in "Diet, Cancer and Health". Moreover, specialised stroke units and diagnostic facilities, e.g. CT and MR scanners, have both become increasingly available in all parts of Denmark during the 1990s, making bias due to geography less likely. Finally, as mentioned above, age had no effect on the validity of the diagnosis of stroke in this study, although the relatively narrow age range of our cases should be kept in mind.

-Identification of cases

The degree of completeness of follow-up is unfortunately seldom characterised in studies that compare different data sources, including the present study, since we did not have access to an independent reference data source on stroke/TIA occurrence in the study population, e.g. a clinical database. Use of the civil registry number allowed safe linkage with the population-based National Registry of Patients, by which we were able to identify all participants of the "Diet, Cancer and Health" who were hospitalised with stroke or TIA in Denmark during follow-up. Given the age profile of our study base, it is likely that very few patients with clinical symptoms of acute stroke were not referred to hospital for further evaluation. Focusing our study on hospitalised cases, we may have missed very mild cases or patients who died before

they reached hospital. We had very limited possibilities for identifying these patients, particularly because the National Registry of Causes of Death is not yet updated to cover the entire follow-up period. However, the retrieval of medical records for probable hospitalised cases of stroke or TIA was almost complete, i.e. only 2.7% of the records could not be found, and, more important, we found no indication that the predictive value was different among the cases where medical records could not be retrieved. Thus, loss to follow-up was likely to be non-differential.

In conclusion, selection problems due either to selection into the study population or to loss to follow-up were likely to be of minor importance in this study.

Information problems

Information problems may occur in a closed cohort study such as "Diet, Cancer and Health" due to shortcomings in the completeness or quality of the exposure, the outcome, or the confounder data. To cause bias, these shortcomings should be distributed unevenly within the study population, i.e. sensitivity and specificity of the outcome classification should depend on the exposure classification.

The risk of information problems was reduced in this study by the use of a detailed standardised form for reviewing all available medical records and hospital discharge letters in probable cases of stroke or TIA. The reviews were conducted without knowledge of the data collected in the study at baseline (except age and sex). Detailed clinical information was available from the medical records, e.g. brain imaging, cardiac imaging, duplex imaging of extracranial arteries, arteriography, and laboratory tests were frequently used among these patients. Thus, the diagnostic basis on which assessment of cases was made was extensive, making it possible to diagnose and subclassify cases according to internationally accepted criteria (2,91-93).

Confounding and effect modification

Confounding, i.e. mixing of effects, was not an issue in this study due to the somewhat atypical design. However, we were able to examine whether age and sex acted as effect modifiers in relation to the predictive value of a discharge diagnosis of stroke or TIA, i.e. if the predictive value varied across age and sex strata. None of these factors was associated with the predictive value of the discharge diagnoses. By contrast, the type of discharging department, i.e. emergency room, non-speciality departments, or speciality departments, were clearly associated with the predictive value.

Chance- statistical precision

Random error or chance is inherent in all observations. The role of chance may be assessed through hypothesis testing using p-values or through estimation. Estimation is the process of using calculated sample values to determine the probable value of a population parameter using either point estimates, confidence intervals or confidence interval functions, i.e. functions of the p-value versus the possible range of relative risk estimates. Whereas the p-value is primarily used to determine the simple question of presence or absence of statistical significance, more detailed information may be obtained when using confidence intervals or confidence interval functions. Confidence intervals and confidence interval functions are particularly useful when examining associations between exposures and outcomes, since these measures separate two important aspects of the data, i.e. the strength of the association between the exposure and the outcome, and the precision with which that relation was measured (96). Throughout this thesis we have assessed the statistical precision of all measures using confidence intervals.

When studying the confidence intervals for the predictive values for some of the diagnoses in study I, in particular SAH and ICH, it is clear that the statistical precision of these estimates is moderate. Due to a much larger number of cases with stroke of any kind, IS or unspecified stroke, the statistical precision of the estimated predictive value for these diagnoses was much better.

Although the number of cases was also low in some subgroups when examining the predictive values according to type of discharging department, our finding of a higher predictive value of diagnoses from the more specialised departments is supported by the consistent trend within all the different diagnostic subgroups.

Conclusion

The most important threat to the accuracy of this study, ie the validity and precision of the findings, was probably the low statistical precision of some of the estimated predictive values. By contrast the risk of major sources of selection or information problems was low.

3.5. Discussion

We evaluated the predictive value of the stroke and TIA diagnoses in the The Danish National Registry of Patients. Overall, a discharge diagnosis of stroke or TIA had a low to moderate predictive value. The predictive value varied considerably for different diagnostic subgroups and according to the type of discharging department. For some diagnoses the estimates of the predictive value were based on few cases, making them rather imprecise.

The high predictive value of the IS diagnosis in this study is in agreement with other recent studies on the validity of stroke diagnoses (21-29). By contrast, we could not confirm the high validity of SAH and ICH diagnoses found in those studies, but, since the absolute number of

cases is small in all the studies, it is difficult to draw conclusions from the apparent differences. Detailed clinical information was available from the medical records in the present study, which is positive from a clinical point of view, since a valid diagnosis is essential when treating and caring for these patients. However, the frequent use of a wide range of diagnostic tools did not result in a high predictive value for several of the discharge diagnoses in our study. There may be several explanations for this finding, including variations in coding practice, errors in coding, limitations in the specificity of the available codes and errors, and variation in clinical diagnoses (16). In agreement with our study, a higher validity of the stroke discharge diagnoses has also previously been reported from specialised hospitals and departments of neurology or neurosurgery (23,29). This may be explained by the more extensive experience and interest in diagnosing, treating, and discharging stroke patients at the specialised departments and hospitals. Handling stroke patients in the emergency room may, however, represent a separate logistic problem in relation to making the correct discharge diagnosis, since time is often short and the patient is often transferred to another department before information from brain imaging etc. is available to the physician treating the patient.

Studies on the predictive value of other common discharge diagnoses have reported substantial variation, e.g. moderate to high predictive values (>80%) have been found for diagnoses of coronary heart disease (98), diabetes mellitus (99), venous thromboembolism (100), chronic obstructive pulmonary disease (101), rheumatoid arthritis (102), and upper gastrointestinal bleeding (103), whereas low predictive values (<60%) have been reported for diagnoses of hypertension (104), rupture of the uterus (105), and septicaemia (106). Thus, the moderate predictive value reported for several discharge diagnoses of stroke reported in this study is not an isolated problem for these diagnoses. Given that the interest in and the possibilities for using

large routine health care databases for surveillance and research is rapidly increasing, it becomes more important to consider and document the quality of the available data, taking into account the specific circumstances under which they are collected, and the purposes for which they should be used. Further improvement in the completeness and quality of data from routine registries and databases may depend on more systematic use of data for scientific and administrative purposes, and more direct feed back to the staff and departments collecting the data.

3.6. Conclusion

The predictive value of stroke and TIA diagnoses in The Danish National Registry of Patients was low to moderate for several diagnostic subgroups, and it differed among different types of department. This misclassification of the outcome, being differential or non-differential, could seriously weaken the possibility of making causal inferences in epidemiological studies if due caution is not taken. Thus, use of registry-based stroke and TIA diagnoses for epidemiological research should be done with caution, and for several subdiagnoses it should probably be preceded by assessment of the registered cases.

4. STUDY II: INTAKE OF FRUIT AND VEGETABLES AND RISK OF ISCHAEMIC STROKE

4.1. Subjects and methods

The study population in this follow-up study included participants in "Diet, Cancer and Health" not registered before enrolment in the Danish National Registry of Patients with a diagnosis of cardiovascular disease, including stroke, TIA, ischaemic heart disease or peripheral arterial disease, and not leaving 10 or more items blank in the FFQ, nor having seven or more items with implausible values. The outcome was IS as verified in Study I.

We estimated the intake of fruit and vegetables from the FFQ, in which the participants reported their average intake of different food items during the previous year within one of 12 possible categories, ranging from never to 8 or more times per day.

The daily intake of specific foods and nutrients, including 42 different types of fruit and vegetables, was calculated from the FFQ for each participant, using the software program Food Calc (107). Standard recipes and portion sizes were applied using data from different sources, i.e. the 1995 Danish National Dietary Survey (108), 24-hour diet recall interviews from 3,818 of the participants in "Diet, Cancer and Health" (109), and various cookery books.

Fruit and vegetables were grouped into 10 categories: leafy vegetables (e.g. lettuce and spinach), fruiting vegetables (e.g. cucumber, peas, and tomatoes), root vegetables other than potatoes (e.g. carrots), cruciferous vegetables (e.g. broccoli and cauliflower), mushrooms, onion and garlic, stalk vegetables (e.g. leeks and asparagus), citrus fruit (e.g. oranges and grapefruit), other fruits (e.g. apples and pears), fruit and vegetable juices. Potatoes were not included in the analyses.

4.2. Data analysis

We categorised the intake of fruit and vegetables into quintiles, and the risks of IS were compared using the lowest quintile as reference level.

We used Cox proportional hazards regression to obtain risk ratios (RRs) adjusted for potential confounders, i.e. factors that are more or less established as independent risk factors for IS and may also be associated with intake of fruit and vegetables. These factors include sex, total energy intake, smoking, blood pressure, serum cholesterol, diabetes mellitus, body mass index (BMI), alcohol intake, intake of red meat and omega-3 polyunsaturated fatty acids, physical activity, and education.

We used two approaches for fitting the multivariable models. First, going through the existing scientific literature, we tried to identify all potential confounders in the relationship between intake of fruit and vegetables and the risk of IS, which were available in our dataset. We then used the change-in-estimate method to identify the most important confounders and hence estimate risk ratios adjusted for these factors. In the change-in-estimate method, selection of variables is based on changes in the estimated exposure effect, ie according to the degree to which inclusion or exclusion of the variable affects the association between intake of fruit and vegetables and the risk of IS (110). This approach has the advantages of maximising the statistical precision of the risk estimates, while identifying the most influential confounding factors. Apart from the diet, the variables in these models included sex, total energy intake, and smoking. Second, we extended this adjusted model by adding the additional potential confounders identified by the literature search (Full model), i.e. systolic and diastolic blood pressure, serum cholesterol at baseline, diabetes mellitus, BMI, alcohol intake, intake of red meat and omega-3 polyunsaturated fatty acids, physical activity, and education. We used this two-step strategy in order to obtain more detailed information on the identity and strength of

the confounders in the association between intake of fruit and vegetables and the risk of IS.

Age was used as the time axis, and we corrected for delayed entry at the time of enrolment, since the persons were first considered at risk from their age at inclusion. The assumption of proportional hazards in the Cox models was evaluated using graphical assessment, and it was found appropriate in all the models.

Energy-adjusted food intakes were derived by adding the median nutrient intake to the residuals from regression analyses of intake of nutrients on energy intake (111). The total energy intake was also included in the models.

We also carried out stratified analyses with separate Cox models according to sex, smoking status, and length of follow-up to assess possible variation in the relative risk estimates across the strata. Any potential effect modification introduced by these factors might be related to differences in sex hormones, oxidant load, and presence of undiagnosed cardiovascular disease.

Finally, the analyses were repeated according to different subtypes of IS, ie large-artery atherosclerosis, cardioembolism, small-vessel occlusion, and stroke of undetermined etiology.

4.3. Results

The study population included 54,506 partcipants in “Diet, Cancer and Health”. Data on all variables used in this study were available for 53,035 participants (97.3%).

Intake of fruit and vegetables was positively associated with the proportion of females and the proportions with higher education, diabetes mellitus, and total energy intake. By contrast, the total intake of fruit and vegetables was negatively associated with the proportion of current smokers, and the proportions with hypertension, high cholesterol, and high alcohol intake.

266 study participants were hospitalised with IS during follow-up.

Table 4.1 shows risk ratios of IS according to intake of fruit and vegetables. The total intake of fruit and vegetables, as well as fruit and vegetables separately, was associated with a reduced risk of IS in the analyses based on crude data; the risk ratios were between 0.40 and 0.63 when comparing the top with the bottom quintile. The association appeared linear for total intake of fruit and vegetables, whereas a less systematic pattern across the quintiles was seen for fruit and vegetables separately. Adjustment for sex, total energy intake, and smoking weakened the association (risk ratios between 0.52 and 0.85); however, the directions of the associations remained unchanged. The risk ratios did not change appreciably after additional adjustment for systolic and diastolic blood pressure, serum cholesterol at baseline, diabetes mellitus, BMI, alcohol intake, intake of red meat and omega-3 polyunsaturated fatty acids, physical activity, and education. The lowest RR was seen in the top quintile of fruit intake (RR= 0.60, 95% CI: 0.39-0.93).

Regarding specific types of fruit and vegetables, there was a reduced risk of IS for all items except mushrooms, stalk vegetables, and onion and garlic, with crude RRs between 0.45 and 0.69. Adjustment for the most influential confounding factors weakened these associations, and RRs were now between 0.60 and 0.81 (except for mushrooms, stalk vegetables, and onion and garlic). Additional adjustment for potential confounding factors had a minor impact on the risk estimates. The strongest inverse associations were seen in the top quintiles of intake of citrus fruit (RR= 0.63, 95% CI: 0.41-0.96) and other fruit (RR= 0.67, 95% CI: 0.43-1.04). A decreased risk was seen in most quintiles, but wide confidence intervals for most estimates made it difficult to assess the precise form of the association across quintiles of intake.

When we stratified by smoking status (never/former smokers vs. current smokers), we found an indication of variation in the relative risk estimates across the strata when comparing top and bottom quintiles, e.g. of total intake of fruit and vegetables (non-smokers: RR= 0.93, 95% CI:

0.47-1.82 vs. smokers: RR= 0.65, 95% CI 0.35-1.20). No clear differences were seen after stratifying by sex (men: RR= 0.88, 95% CI 0.49-1.60 vs.women: RR= 0.68, 95% CI 0.34-1.40).

The associations between intake of fruit and vegetables and IS were also evaluated for different subtypes of IS separately. No differences were evident; however, some of the risk estimates were imprecise due to the relatively low number of outcomes. The lowest RRs were found for strokes due to small-vessel occlusion and strokes of undetermined etiology in the top quintile of fruit intake; RRs of 0.67 (95% CI: 0.34-1.31) and 0.50 (95% CI: 0.23-1.08), respectively. However, inverse associations were found for all types of IS.

In searching for any preclinical disease at baseline that might affect the diet or other lifestyle factors, exclusion of cases from the first year of follow-up did not change the risk estimates.

TABLE 4.1. Risk ratios (RR) of IS according to dietary intake of fruit and vegetables (Study II).

Dietary item (quintiles)	Median intake (g/day)	Unadjusted (95% CI)	Adjusted model* (95% CI)	Full model† (95% CI)
All fruit and vegetables				
1	146.9	1.00	1.00	1.00
2	252.8	0.82 (0.58-1.14)	0.80 (0.57-1.13)	0.85 (0.60-1.21)
3	345.9	0.68 (0.48-0.97)	0.82 (0.57-1.17)	0.88 (0.61-1.27)
4	459.5	0.58 (0.40-0.85)	0.66 (0.44-0.99)	0.73 (0.49-1.11)
5	673.2	0.44 (0.29-0.66)	0.63 (0.41-0.96)	0.72 (0.47-1.12)
All fruit				
1	41.1	1.00	1.00	1.00
2	107.0	0.62 (0.44-0.88)	0.84 (0.60-1.18)	0.74 (0.52-1.07)
3	167.4	0.71 (0.51-0.99)	0.92 (0.65-1.30)	0.97 (0.68-1.38)
4	249.7	0.47 (0.32-0.69)	0.66 (0.44-0.99)	0.66 (0.44-1.00)
5	423.4	0.40 (0.27-0.60)	0.52 (0.34-0.81)	0.60 (0.39-0.93)
All vegetables				
1	65.7	1.00	1.00	1.00
2	117.0	0.77 (0.54-1.09)	0.91 (0.64-1.29)	0.99 (0.69-1.42)

3	162.1	0.71 (0.49-1.02)	0.93 (0.65-1.34)	1.04 (0.71-1.52)
4	214.8	0.71 (0.49-1.03)	0.97 (0.66-1.40)	1.08 (0.73-1.60)
5	312.3	0.63 (0.43-0.92)	0.85 (0.57-1.26)	0.97 (0.64-1.48)

*Adjusted for sex, total energy intake, and smoking.

†Adjusted for sex, total energy intake, smoking, systolic blood pressure, diastolic blood pressure, total serum cholesterol, history of diabetes, body mass index, alcohol intake, intake of red meat, intake of omega-3 polyunsaturated fatty acids, physical activity, and education.

4.4. Strengths and weaknesses of the study

Selection problems

-Sampling

Different selection mechanisms might have hampered our findings.

Selection into the study at baseline and exclusion of participants with previously diagnosed cardiovascular disease may have resulted in a study population with a more healthy lifestyle and consequently a lower risk of stroke, including IS. Indeed, the incidence rate of stroke was lower than expected in our study population. This might have limited the range of variation in the dietary exposure variables. However, looking at the substantial variation in median intake of fruit and vegetables between the top and bottom quintiles, this appears not to have been a major problem.

-Selection bias

Selection into the study at baseline may also under certain circumstances have introduced selection bias. This bias may occur if the likelihood of participation in the study varied not only with respect to exposure or risk of disease, but instead varied according to specific combinations of exposure and risk of disease (112). This situation is most likely to occur in situations when potential study participants are aware of the details of the hypotheses that are going to be examined in the study, which was not the case in our study.

Selection bias might also have occurred after inclusion into the study if loss to follow-up was related to both the intake of fruit and vegetables and the risk of stroke. The complete follow-up of the study population through record linkage with population-based nationwide registries made this type of bias less likely. However, we cannot completely exclude that the probability of being admitted to hospital differ according to intake of fruit and vegetables among persons experiencing symptoms of stroke.

Information problems

Information problems may have been present in all types of data used for this study, including exposure data, outcome data, and confounder data. Misclassification of the intake of fruit and vegetables, may have occurred for several reasons. In general, assessment of diet of free-living individuals, and in particular long-term diet rather than intake on any specific day or limited number of days, constitutes a myriad of methodological challenges to which the available solutions may often be imperfect and yet still useful. Assessment of diet including intake of fruit and vegetables was done in "Diet, Cancer and Health" using a FFQ, which is usually considered the most appropriate method for dietary assessment in large epidemiologic studies (113). The validity of the FFQ was compared with information from diet recall interviews of the participants, which can be considered as a better but still imperfect standard, and was found to be acceptable (86,87,109). Although detailed information about type and frequency of fruit and vegetable servings was collected, a certain degree of inaccuracy of the FFQ is inevitable due to substantial variation in day-to-day and seasonal intake, crude frequency categories, and standardised portion sizes, etc. The inaccuracy of the dietary data provided at baseline compared with the current intake will furthermore tend to increase according to length of follow-up due to changes in dietary practices among the participants during follow-up. These sources of non-differential misclassification of the exposure variables would lessen our ability to detect any association between dietary intake of fruit and vegetables and the risk of IS. Correction for at least some of this misclassification is theoretically possible since the FFQ has been calibrated against a 24-hour recall interview. However, correction was not used in our study.

Misclassification of potential confounders, e.g. smoking, blood pressure, and diabetes, might also have influenced our findings. Depending on the prevalence of the confounder and the

direction and strength of the association with the exposure and the outcome, non-differential misclassification of the potential confounder variables may lead to residual confounding and could result in both attenuation and inflation of the relative risk estimates. However, looking at the relatively minor effect on the risk ratios of adjusting for potential confounders, any major role of residual confounding due to misclassification of confounders appeared less likely. Furthermore, misclassification of the IS may have occurred. However, as described in study I, this was probably of minor importance in our study due to the standardised assessment of all outcome events.

By contrast with the likely non-differential misclassification of several variables, the risk of information bias was probably minimal since the assessment of end points, i.e. stroke, was standardised and done without knowledge about the exposure or confounder status.

Confounding

We adjusted for a number of potential confounders in the regression analyses, which weakened the association between intake of fruit and vegetables and risk of IS. However, our risk estimates may still be affected by potential confounders not included in the analyses. Thus, a high intake of fruit and vegetables, and in particular fruit, might be an indicator of an overall more healthy lifestyle (114) that, besides measurable behavioural elements such as being a non-smoker, having a moderate alcohol intake, etc., also includes a wide range of more subtle behavioural and psychosocial elements. Furthermore, knowledge about diet and attitudes towards health and food are likely to be associated with intake of fruit and vegetables (115). Selection into "Diet, Cancer and Health" may have reduced this confounder, since the study population, at least on average, was likely to have a healthier lifestyle than the general population.

Chance- statistical precision

The follow-up period and thus number of outcomes was modest in this study, in particular when examining subtypes of IS or specific types or quantities of fruit and vegetables, and several of the associations examined were therefore not statistically significant. However, several of the point estimates in the study were remarkably low, and the direction of the findings in almost all the analyses was towards a risk reduction with increasing intake of fruit and vegetables.

The association between intake of fruit and vegetables and the risk of IS did not apparently follow a simple log-linear dose-response curve in the multiplicative models. However, the lack of statistical precision made it difficult to make further detailed analyses on the association using the dietary exposure variables as continuous variables, although this could theoretically be done with spline-regression or curve-smoothing techniques, i.e. regression techniques that are able to fit non-linear associations.

Conclusion

The risk of uncontrolled confounding and the low to moderate statistical precision of several of the risk estimates were important weaknesses in this study. This makes it difficult to make conclusions on the precise size of the potential risk reduction associated with an increasing intake of fruit and vegetables within the study population.

4.5. Discussion

In this cohort study, including persons aged 50-64 years at baseline, the intake of fruit and vegetables was associated with a reduced risk of IS. This protective effect was present for most categories of fruit and vegetables, in particular citrus and other fruit. The reduced risks were

most evident in relation to stroke due to small-vessel occlusion and stroke of undetermined etiology.

Our findings are consistent with the few previous prospective studies providing data on intake of fruit and vegetables and risk of cerebrovascular morbidity and mortality (34,46,51,56,57). Based on the Nurses' Health Study and Health Professionals' Follow-up Study, Joshipura et al. (51) reported a RR of 0.69 (95% CI: 0.52-0.91) when they compared the top quintile of fruit intake (median servings per day: men= 4.54, women= 4.33) with the bottom (median servings per day: men=0.86, women=0.72), which is similar to the RR of 0.60 found in our study if servings are assumed to be about 100 g each. They also found a weaker inverse association between intake of vegetables alone and risk of IS (RR=0.90) when they compared top quintiles of intake (median servings per day: men=6.21, women=5.37) with bottom quintiles (median servings per day: men=1.60, women=1.36). By contrast, Gillman et al. (34) did not find a stronger protective effect of fruit compared with vegetables in a study based on the Framingham cohort, i.e. each increment of three daily servings of fruit or vegetables was associated with a RR of 0.81 (95% CI: 0.56-1.19) and 0.74 (95% CI: 0.54-1.02), respectively. However, the study was based on only 97 cases, including cases of IS, ICH and TIA.

Several biological mechanisms may be involved in the apparent protective effect of fruit and vegetables on the risk of IS. A number of fruit and vegetable constituents, including micronutrients, antioxidants, phytochemicals, and fiber, have been related to a decreased risk of stroke and other cardiovascular diseases in animal models and observational studies (44-48). However, the evidence about the role of the individual constituents has so far been inconclusive, and the results from testing several of these potentially bioactive compounds

alone or in combination in randomised trials on cardiovascular disease and cancer have been disappointing, in particular in persons eating a Western diet (53,116-120). There may be various explanations for these findings, e.g. inexpedient dosages, short study duration, etc. Furthermore, it may well be that the role of single nutrients cannot easily be isolated from the complex biochemical content of plant foods. Attempts at making causal inferences in nutritional epidemiology between specific constituents of fruit and vegetables and later risk of disease is, furthermore, hampered by the varying content of biologically active constituents in fruit and vegetables, e.g. vitamin C, β-carotene, and flavonoids, depending on the geographical origin, season of the year, methods of storing and cooking, etc. These issues make it difficult to study the specific nutrients, and they are also troublesome in attempts to classify fruits and vegetables by levels of specific constituents and micronutrients. For these reasons, focusing future research efforts on intake of specific foods or even on food patterns rather than on single food components or nutrients may prove to be a more useful approach when examining the association between diet and health.

4.6. Conclusion

Previous studies have suggested a protective effect of a high intake of fruit and vegetables on the risk of IS. Our results from this large Danish cohort study support the hypothesis that a high dietary intake of fruit and vegetables is independently associated with a reduced risk of IS.

5. STUDY III: *CHLAMYDIA PNEUMONIAE* SEROPOSITIVITY AND RISK OF

ISCHAEMIC STROKE

5.1. Subjects and methods

The study population in this nested case-control study comprised participants in "Diet, Cancer and Health" not registered in the Danish National Registry of Patients with a diagnosis of cardiovascular disease, including stroke, TIA, ischaemic heart disease and peripheral arterial disease, before enrolment and not leaving 10 or more items blank in the FFQ, nor having seven or more items with implausible values.

Cases were defined as participants with verified first time IS. We selected one control for each case matched by age (within 5 years) and sex using the incidence-density technique (121). The control thus had to be alive and at risk of first stroke, i.e., no previous stroke discharge diagnosis registered in The Danish National Registry of Patients at the time the case was diagnosed.

We used the LOY-EIA (Labsystems, Helsinki, Finland) for measuring *Chlamydia pneumoniae* IgG and IgA plasma antibodies from the blood samples collected and stored at baseline (122). The test is an indirect solid-phase enzyme immunoassay with horseradish peroxidase as a marker enzyme (122). The results are expressed as enzyme immunounits, which are equal to antibody titers. The antibody titers were classified into a series of IgA (range, >1:8 to >1:64) and IgG titers (range, >1:16 to >1:256). A priori, IgA titers >1:16 and IgG titers >1:64 were considered appropriate cut-off values, since these values have been used traditionally in seroepidemiological studies, although the rationale for this approach is not well-established (68,123). All analyses were done without knowledge of the case-control status of the samples.

5.2. Data analysis

The association between the presence of IgA and IgG antibodies at baseline and subsequent risk of any IS, and different subtypes of IS, was evaluated across a series of antibody titers.

We used conditional logistic regression to obtain odds ratios (ORs) adjusted for available potential confounders that have been linked with the risk of IS and may possibly also be associated with the risk of being infected with *Chlamydia pneumoniae*, i.e., smoking, history of hypertension, history of diabetes mellitus, hypercholesterolaemia, BMI, alcohol intake, and education. The obtained ORs in these analyses were direct estimates of incidence rate ratios, which can otherwise only be estimated in follow-up studies.

We also made stratified analyses according to age and sex and calculated ratios of the ORs to assess possible effect modification, i.e. variation in the strength of the association between exposure and outcome according to a third factor, as indicated by departure from the multiplicative model.

5.3. Results

Among the 266 study participants in "Diet, Cancer and Health" who were hospitalised with IS during follow-up, 254 (95.5%) had provided plasma samples at baseline. This study thus included 254 cases and 254 controls.

Information on all variables was available for 498 persons (98.0%). Cases were more likely than controls to be current smokers, have a history of hypertension or diabetes, or hypercholesterolaemia at baseline.

Table 5.1. shows ORs for all ISs, and for different subtypes of IS according to a positive IgA (>= 1:16) or IgG (>=1:64) *Chlamydia pneumoniae* titer. Positive IgA or IgG titers were

associated with increased risks of IS of 1.45 (95% CI; 0.98-2.15) for IgA and 1.28 (95% CI: 0.89-1.82) for IgG. Adjustment for smoking, history of hypertension, hypercholesterolaemia, diabetes mellitus, BMI, alcohol intake, and education did not change the risk estimates appreciably. When both IgA and IgG titers were increased, an even further increase in the risk of IS was seen (adjusted OR = 1.77, 95% CI: 1.04-3.00). We found the strongest association for stroke due to large-artery atherosclerosis, i.e. ORs were 3.87 (95% CI: 0.43-35.10) and 6.32 (95% CI: 0.76-52.61), whereas more modest associations were found for stroke due to cardioembolism, small-artery occlusion, and undefined cause, i.e. ORs were between 0.53-1.59.

The associations between seropositivity against *Chlamydia pneumoniae* and the risk of IS differed according to the cut-off value used for the antibody titers, in particular for IgG. Thus, *Chlamydia pneumoniae* seropositivity was associated with both an increased risk of IS (IgG >= 1:32: adjusted OR=2.04, 95% CI 1.23-3.37) and a reduced risk (IgG >= 1:128: adjusted OR=0.56, 95% CI :0.31-1.00) depending on the chosen cut-off value. Variation in the association between *Chlamydia pneumoniae* seropositivity and risk of IS was also seen for the subtypes of IS; however, some of the risk estimates were imprecise, particularly for stroke due to large-artery atherosclerosis and cardioembolism.

When we stratified by sex, we found varying risk estimates across the strata. The variation was most evident for IgA >=1:16 (men: adjusted OR=2.23, 95% CI: 1.19-4.17 vs. women: adjusted OR=0.58, 95% CI 0.21-1.60). The ratio OR men/OR women was 3.84 (95% CI: 0.85-17.43). No clear differences were seen after stratifying by age, for IgA >=1:16 for example (age <60 years: adjusted OR=1.44, 95% CI: 0.68-3.08 vs. age >=60 years: adjusted OR=1.96, 95% CI 0.95-4.04).

TABLE 5.1. *Chlamydia pneumoniae* titer and risk of IS (Study III).

	Cases/controls	IgA titer ≥ 1:16		IgG titer ≥ 1:64	
		Crude OR* (95% CI)	Adjusted OR† (95% CI)	Crude OR* (95% CI)	Adjusted OR† (95% CI)
Total	254/254	1.45 (0.98-2.15)	1.54 (0.96-2.47)	1.28 (0.89-1.82)	1.28 (0.83-1.95)
Stroke subtype:					
Large-artery atherosclerosis	26/26	3.00 (0.97-9.30)	3.87 (0.43-35.10)	6.50 (1.47-28.80)	6.32 (0.76-52.61)
Cardioembolism	22/22	1.00 (0.32-3.10)	0.53 (0.07-4.08)	0.86 (0.29-2.55)	0.91 (0.15-5.45)
Small-artery occlusion	112/112	1.40 (0.72-2.72)	1.59 (0.71-3.52)	1.12 (0.65-1.92)	1.11 (0.59-2.08)
Other determined etiology	2/2	-	-	-	-
Undetermined etiology	92/92	1.29 (0.69-2.44)	1.54 (0.59-4.01)	1.05 (0.57-1.94)	1.36 (0.55-3.36)

*Matched for age and sex.

†Matched for age and sex; adjusted for smoking, self-reported history of hypertension, self-reported history of diabetes, serum cholesterol, body mass index, alcohol intake, and length of education.

5.4. Strengths and weaknesses of the study

Selection problems

-Sampling

Selection problems may have occurred at several levels of this study. As in study II, selection into the study may have limited the variation of the exposure variables, in this case IgA and IgG antibody titers against *Chlamydia pneumoniae*. It may be difficult to assess whether this was the case because the prevalence of antibody titers is known to vary over time in the population according to season of the year, epidemics, etc. However, recruitment of participants into "Diet, Cancer and Health" was carried out during a five-year period, which reduced the risk of any seasonal or calendar time effect. Furthermore, with respect to the actual variation of the antibody titers, it seems less likely that any major restriction on the variation of the exposure variable has occurred.

-Selection bias

The risk of selection bias was probably of minor importance in this study due to the nested case-control design with prospectively collected data on the exposure, potential confounders, and outcome. Consequently, the exposure, ie IgA and IgG antibody titers to *Chlamydia pneumoniae*, did not influence the chance of being included in this case-control study as a case or a control. Furthermore, the complete follow-up of the study population minimised loss to follow-up, and thus the risk of not identifying all potential cases and controls. Plasma samples were not available for 12 of 266 cases (4.5%), but there was no indication that these cases differed from the rest, i.e. age, sex and prevalence of smoking, hypertension, diabetes, BMI, and length of education were similar in the two groups.

Information problems

As discussed in studies I and II, information problems may affect exposure, confounder and

outcome data. Assessment of exposure status in this study was made by measurement of *Chlamydia pneumoniae* antibody titers using a serological test in an enzyme-linked immunoassay format that has previously been shown to perform well by comparison with the more widely used complement fixation and microimmunofluorescence tests (122). Thus, a sensitivity and specificity of 96% and 99%, respectively, of the LOY-EIA test, found in a recent study, were competitive with the performance of the complement fixation test (sensitivity: 69%, specificity: 99%) and the microimmunofluorescence test (sensitivity: 88%, specificity: 99%) (119). The latter tests have been questioned due to low sensitivity and specificity, and substantial inter-observer and inter-laboratory variation (122,124). Any non-differential misclassification of the antibody titers would probably result in conservative risk estimates and could thus not explain the direction of the associations found in this study.

We defined a priori IgA titers >1:16 and IgG titers >1:64 as cut-off values for seropositivity against *Chlamydia pneumoniae*. However, it is currently uncertain to which extent these cut-off values (or any other values) actually indicate prior/chronic infection with *Chlamydia pneumoniae* and the study subjects may therefore not have been classified into to the biological most meaningful categories. A range of cut-off values was used in order to further examine this issue, however, we did not have a reference standard of *Chlamydia pneumoniae* infection status to compare with.

Potential misclassification of confounder data has been discussed in study II. The risk of information bias, ie a situation where the exposure data are affected by the outcome data, was minimal in this study because of the collection of information on exposures and potential confounders before the time of the stroke events, and because of the blinding of the laboratory staff, who conducted the measurement of antibody titers.

Confounding

Although we adjusted for a number of available confounders, the risk estimates may still be affected by potential confounders not included in the analyses, eg diet, or by residual confounding due to misclassification or use of crude categories for several of the included confounders. The latter seems less likely, because adjusting for potential confounders using crude categories had only a minor and non-uniform influence on the risk estimates in our analyses.

Unfortunately, further adjustment for potential confounding is still somewhat hampered by the lack of detailed knowledge on the etiology of subtypes of ischemic stroke.

Chance-statistical precision

Although to our knowledge this study is the largest of its kind, the statistical precision of several of the risk estimates was modest. The limited size of some of the subgroups furthermore made it difficult to ensure efficient control for confounding at all times.

Conclusion

Use of the nested case-control study design was an efficient approach in this study where additional information was needed, i.e. *Chlamydia pneumoniae* antibody titers that were not obtained for the whole "Diet, Cancer and Health" cohort. The nested case-control design thus provided us with some of the advantages of the follow-up study, i.e. it took place within a well-defined cohort and was based on prospectively collected data, and the case-control study, i.e. we only had to obtain and analyse data on cases and relevant controls instead of on the entire cohort.

The potentially uncontrolled confounding was an important weakness in this study.

Furthermore, the limited statistical precision of the risk estimates should be taken into consideration. It should be emphasised that we examined the association between antibody titers to *Chlamydia pneumoniae* and the risk of IS. Raised antibody titers were thus used as a proxy for infection with *Chlamydia pneumoniae*, since validated microbiological methods for detecting *Chlamydia pneumoniae* in vivo are not yet available for large scale epidemiological studies (121).

5.5. Discussion

In this nested case-control study, including persons aged 50-64 years at baseline, a combination of increased IgA and IgG antibody titers against *Chlamydia pneumoniae* was associated with an overall increased risk of IS. However, the association between *Chlamydia pneumoniae* antibody titers and risk of IS differed according to gender, subtype of IS, and cut-off value of antibody titers.

Our findings agree in part with a number of previous sero-epidemiological studies, which have reported positive associations between *Chlamydia pneumoniae* infection and the risk of stroke or TIA (39,64-66). However, several previous studies have been hampered by methodological weaknesses, including crude assessment of outcomes and exposures, limited possibilities for confounder adjustment, and use of small sample size. Studying the findings from our study, an association between infection with *Chlamydia pneumoniae* and the risk of IS seems probable, whether causal or not. The association may however appear more complex than previously thought. Comparing our findings with previous studies on this issue, it becomes evident that interpretation and generalisation of the results from sero-epidemiological studies on *Chlamydia pneumoniae* infection and risk of cerebro- and cardiovascular disease is difficult for a number of

reasons (125,126). First, there is no standardisation of the available *Chlamydia pneumoniae* diagnostic tests, and the extent to which the titers of *Chlamydia pneumoniae* IgA or IgG antibodies reflect previous infection, chronic active infection, or reinfection is unclear (126). Studies on patients with pneumonia have suggested that high IgG titers (1:512 or 1:1024) reflect a recent infection (126), but the extent to which high IgA or IgG titers are also measures of persistent, chronic active or reinfection is still a matter of debate (126). Still, despite this uncertainty about the interpretation of raised antibody titers, large randomised trials have been initiated using specific cut-off values for the anti-body titers as inclusion criteria (127). The findings from our study underline these problems, i.e. both an increased and a reduced risk could be found depending on the chosen cut-off value for the *Chlamydia pneumoniae* antibody titer. Only one previous study has examined the association between *Chlamydia pneumoniae* seropositivity and risk of stroke over a broad range of titers, but the results were inconclusive because of a small number of cases (68). Use of repeated measurements of antibody titers of the same individuals over a prolonged follow-up period, or perhaps measurement of *Chlamydia pneumoniae* antigens in circulating monocytes, might prove useful for a more correct classification of the infectious status of individuals (126).

Second, the inconsistency among the reported studies may be related to the fact that the outcomes studied within the different studies may not be comparable, particularly with respect to studies on stroke. Thus, if the risk associated with *Chlamydia pneumoniae* varies for different subtypes of IS, as indicated by our study, this may be an important contributor to the inconsistency because differences in the distribution of stroke subtypes are likely to be present between the studies reported so far, taking into account the differences in ethnicity, age, sex prevalence of hypertension, etc. between the study populations (39,64-68).

Third, the lack of consistency might also be related to other issues. For instance, it is still

unclear whether *Chlamydia pneumoniae* infection is related primarily to the initiation and early progression of atherosclerosis, or to the much later occurring thrombotic complications such as stroke or myocardial infarction (125,126). Thus, it may well be that studies using *Chlamydia pneumoniae* infection status 10 or 20 years before the time of stroke as an exposure are not comparable with studies that use exposure information reflecting a more recent infection status. Fourth, based on the findings in the present and other studies (125), one may speculate that the host response to infection, e.g. type of immune response (humoral or cellular driven), is a decisive factor when evaluating the role of *Chlamydia pneumoniae* in relation to stroke and cardiovascular disease. Although further studies with more detailed data on these parameters are needed to confirm this, not taking this potentially important effect modifier into account may also have led to inconsistent findings.

Before causal inference can be made from sero-epidemiological studies, we have to improve the design of the studies and the microbiological methods used for detecting *Chlamydia pneumoniae.* Furthermore, other explanations for an association should be considered and further studied, including the possibility that *Chlamydia pneumoniae* infection might just be a risk marker rather than an actual cause of disease. Thus, an impressive accumulation of data within recent years has documented the role of inflammatory processes in the development of atherosclerosis, including the involvement of macrophages and T lymphocytes (61). Furthermore, there is evidence that pulmonary macrophages infected with *Chlamydia pneumoniae* can disseminate the organism throughout the body, including to atheromatous tissue (128). Consequently, it has been suggested that the presence of *Chlamydia pneumoniae* in atheromatous tissue is merely a marker of the ongoing inflammatory process, and not part of a causal mechanism. Likewise, raised total immunoglobulin levels, including antibodies against

Chlamydia pneumoniae, may be markers of ongoing inflammatory processes (129).

Although several large randomised clinical trials that test the effect of antibiotic treatment on the risk of new cardiovascular events in patients with preexisting cardiovascular disease are currently ongoing and are expected to be completed within the coming years, it is not likely that the question of a potential causal link between infection with *Chlamydia pneumoniae* and development of cardiovascular disease including stroke will be resolved by these studies. Further epidemiological and experimental studies will be needed to clarify the clinical impact of this putative risk factor and the basic mechanisms underlying the potential association. Ideally, epidemiological studies should be performed, of sufficient sample size and based on long-term follow-up of young individuals with repetitive assessments of *Chlamydia pneumoniae* infectious status using a standardised and accurate test, and with detailed data on potential confounders, e.g. lifestyle factors, and effect modifiers, e.g. immune response. Development of an effective vaccine against *Chlamydia pneumoniae* would be another alternative way of obtaining strong arguments for or against a causal role of the organism in the development of cardiovascular disease including stroke.

5.6. Conclusion

Raised IgA and IgG antibody titers to *Chlamydia pneumoniae* may be associated with the risk of IS. However, the risk may differ according to subtype of IS, cut-off value of antibody titers, and gender. These issues should be considered when planning and interpreting studies on *Chlamydia pneumoniae* and the risk of stroke.

6. STUDY IV: NSAID DRUGS AND RISK OF INTRACEREBRAL HAEMORRHAGE

6.1. Subjects and methods

The study population in this nested case-control study comprised all persons who were living in North Jutland County and not registered in the The County Hospital Patient Register (HPR) of North Jutland County with a discharge diagnosis of stroke including ICH during the period 1980-1989. The HPR collects data for the National Registry of Patients at county level, and the type of data included is thus identical in these two registries.

Cases

We defined cases as persons with a first diagnosis of ICH (ICD-8 codes: 431.00, 431.08-431.90, 431.98, 431.99; ICD-10 codes: I61.0-9) in the period 1991-1999.

We evaluated the positive predictive value of a discharge diagnosis of ICH through medical record review in an approximately 10% random sample of the cases using the previously mentioned definition of ICH. The overall predictive value for ICH was 75.0% (95% CI: 65.4%-83.0%) among 92 cases. A clear difference was seen when the predictive value was stratified according to availability of CT/MR scan at the treating hospital, i.e. 89.7% (95% CI: 80.7-95.4%) (61 of 68 cases) when imaging was available at the hospital, compared with 33.3% (95% CI: 16.8-53.6%) (8 of 24 cases) when imaging was not available.

Controls

Using the Civil Registration System we planned to select 10 random controls for each case matched by age (same date of birth), sex, and place of residence (North Jutland County). Since identification of controls could be done at very low cost in this study design, we chose to include more controls for this study than were used in study III in order to achieve the best

possible statistical precision of our relative risk estimates. The controls were selected using the incidence density sampling technique (121).

Exposure and confounder data

Using the Pharmaco-Epidemiological Prescription Database, we identified all prescriptions for non-aspirin NSAIDs (ATC-code M01A) during the study period 1990-1999 among cases and controls before the date of hospital admission of cases. All non-aspirin NSAIDs, except low-dose ibuprofen (200 mg per tablet), are available only on prescription in Denmark. Although low-dose ibuprofen is available without prescription, pensioners and regular users of this drug, e.g. patients with chronic disease or pain requiring prolonged treatment, are presumably all registered in the Prescription Database, because they receive a 50% refund when the ibuprofen is prescribed by a physician.

Data on potential confounding factors identified from the litterature were collected from the HPR and the Pharmaco-Epidemiologic Prescription Database. These included discharge diagnoses of hypertension, chronic bronchitis and emphysema (a proxy measure of smoking), alcoholism, liver cirrhosis, diabetes mellitus, prescriptions for insulin or oral hypoglycaemic agents (a proxy measure of diabetes mellitus), antihypertensive agents, lipid-lowering agents, low-dose aspirin (75-150 mg), high-dose aspirin (100-500 mg), and oral anticoagulants (warfarin, phenprocoumon) before the date of ICH diagnosis. Hypertension was considered present when a person was registered with a discharge diagnosis of hypertension and/or had redeemed a prescription for an antihypertensive drug. Likewise, diabetes mellitus was considered present when a person was registered with a discharge diagnosis of diabetes mellitus and/or had redeemed a prescription for insulin or an oral hypoglycaemic agent.

6.2. Data analyses

We used conditional logistic regression to calculate ORs for ICH among users and non-users of non-aspirin NSAIDs since our dataset was matched. Initially, we conducted bivariate analyses on the association between the prescription of non-aspirin NSAIDs and the risk of ICH, and then on the association between the potential confounding variables, i.e. discharge diagnoses of hypertension, chronic bronchitis and emphysema, alcoholism, liver cirrhosis and diabetes mellitus, or prescriptions of insulin or oral hypoglycaemic agents, antihypertensive agents, lipid-lowering agents, low-dose aspirin, high-dose aspirin and oral anticoagulants, and ICH. We then calculated ORs adjusted for the potential confounders, which were included in the models as dichotomous variables. Separate analyses were conducted with different exposure time windows according to typical supplies prescribed, i.e. us of medication within 30, 60, or 90 days before the date of admission for ICH. Furthermore, stratified analyses were performed according to age, sex, discharge diagnosis of hypertension, and availability of imaging at the discharging hospital.

Finally, sensitivity analyses were conducted to assess the potential influence of misclassification of outcome status based on the finding of an overall predictive value for the ICH discharge diagnoses of 75%. Thus, we calculated ORs for the association between use of non-aspirin NSAIDs and ICH, assuming that the misclassified cases, i.e. 228 cases (25%), had either been exposed to non-aspirin NSAIDs to the same degree as the controls or had not been exposed at all.

6.3. Results

The study population included 912 cases of ICH identified from the HPR, and 9,059 population-based controls. A higher proportion of cases than controls had discharge diagnoses

of hypertension, chronic bronchitis/emphysema, alcoholism, and liver cirrhosis. Similarly, a higher proportion of cases than controls had a history of diabetes mellitus as defined by a discharge diagnosis and/or use of insulin or oral hypoglycaemic agents. Low-dose aspirin and oral anticoagulants were also more frequently used by cases than controls.

Table 6.1. gives crude and adjusted ORs for ICH according to prescription of non-aspirin NSAIDs within 30, 60, or 90 days before the date of hospitalisation for ICH. The proportion of cases and controls with a prescription for non-aspirin NSAIDs differed only moderately regardless of which time window was used. For instance, 5.3% of cases and 4.2% of controls had received a prescription within 30 days before hospitalisation, yielding a crude OR for ICH in this time window of 1.25 (95% CI: 0.91-1.70). After adjustment for potential confounding factors the OR remained virtually unchanged, 1.12 (95% CI: 0.81-1.55). When restricting the analyses to the cases with ICH verified through record review (n=59) and their controls (n=589), we found a crude OR of 0.91 (95% CI: 0.21-3.96) for prescription of non-aspirin NSAIDs within 30 days before the date of ICH. Adjustment for potential confounders had only minor influence, i.e. the OR was 0.81 (95% CI: 0.81 (95% CI: 0.17-3.84).

Changing the time window to 60 or 90 days before the date of hospitalisation for ICH did not show any association between prescription of non-aspirin NSAID and risk of ICH.

To evaluate the possibility that prescription of non-aspirin NSAIDs might be associated with an increased risk of ICH only in certain subgroups of the population, we stratified our analyses according to age, sex, and a previous discharge diagnosis of hypertension. These factors may all be related to the presence of structural changes in the cerebral blood vessels; including

lipohyalinosis, cerebral amyloid angiopathy, intracranial saccular aneurysms, and cerebral arteriovenous malformations, which again are associated with an increased risk of ICH (10). No increased risk of ICH was found in any of the subgroups for non-aspirin NSAID prescriptions within 30 days prior to hospitalisation.

Due to the apparent marked difference in the predictive value of the ICH diagnosis according to availability of brain imaging facilities at the discharging hospital, we also conducted analyses restricted to patients discharged from hospitals with easy access to CT/MR scanning. None of the relative risk estimates were materially altered.

Using sensitivity analyses, we estimated how sensitive our findings were to variations in the assumptions underlying the use of routinely collected registry-based data. In the primary analyses we assumed that the exposure did not differ between true cases, i.e. patients correctly coded with an ICH diagnosis, and misclassified cases, i.e. patients falsely coded with an ICH diagnosis. In the sensitivity analyses we tested the impact of this assumption on the study findings. If we assumed that the misclassified cases had been exposed like the controls, i.e. 4.2% within 30 days before the date of the ICH diagnosis, we found a crude OR of 1.37 (95% CI: 0.96-1.94). In the even more unlikely situation that all the misclassified cases had not received prescriptions for non-aspirin NSAIDs, we found a crude OR of 1.77 (95% CI: 1.29-2.44). Lower ORs were seen when using 60 or 90 days time windows.

TABLE 6.1. Crude and adjusted odds ratios for ICH according to prescription of non-aspirin NSAIDs within 30, 60 or 90, days before the date of hospitalisation (Study IV).

Prescription of drug	Cases (n=912)	Controls (n=9059)	Crude Odds Ratio (95% CI)	Adjusted Odds Ratio* (95%)
Non-aspirin NSAIDs in previous 30 days:				
No	864	8674	100	100
Yes	48	385	1.25 (0.91-1.70)	1.12 (0.81-1.55)
Non-aspirin NSAIDs in previous 60 days:				
No	840	8366	100	100
Yes	72	893	1.03 (0.80-1.33)	0.91 (0.69-1.18)
Non-aspirin NSAIDs in previous 90 days:				
No	814	8199	100	100
Yes	98	860	1.15 (0.92-1.43)	1.02 (0.81-1.28)

*Adjusted for discharge diagnoses of hypertension, chronic bronchitis and emphysema, alcoholism, liver cirrhosis and diabetes mellitus, and prescriptions for insulin or oral hypoglycaemic agents, antihypertensive agents, lipid-lowering agents, low-dose aspirin, high-dose aspirin, and oral anticoagulants before the date of admission for ICH.

6.4. Strengths and weaknesses of the study

Selection problems

Like study III, this study was a case-control study and the study subjects were therefore sampled on outcome status, i.e. ICH or not, instead of exposure status. The case-control design has some important strengths, including high statistical efficiency, but the findings may be influenced by the enhanced potential for introducing selection problems when sampling, and the risk of exposure and confounder ascertainment being more prone to error and bias compared with other study designs.

-Sampling

Due to the uniformly organised health care system in Denmark and the complete and detailed administrative registries, we were able to use a population-based design for the identification of both cases and controls. Thus, the study base was essentially unrestricted, which supports the validity of the findings.

Focussing our study on hospitalised cases, we may have missed the very mild cases or cases who died before they reached hospital. Likewise, old people living in nursing homes and with a short life expectancy may have been less likely to be admitted to hospital. Thus, our findings may not necessarily be extended to all cases of ICH. However, as indicated by the age distribution in our study (mean 68 years, range 19-98 years), even very old patients were hospitalised with ICH in the study area and hence included in the study. Furthermore, we found no indication of effect modification by age among the cases and controls included in our study. Focussing on hospitalised cases of ICH, i.e. patients with a medium grade of disease severity, would not in itself affect the validity of the findings with respect to patients with milder or more severe disease unless the exposure, ie having redeemed a prescription of non-asprin NSAIDs, was also a prognostic factor. We have no reason to believe that this was the case. Thus,

selection problems due to non-differential loss of potential cases and controls may have occurred in this study, but the extent of these selection problems was probably modest.

-Selection bias

Moreover, since non-asprin NSAIDs are not generally believed to be a major cause of ICH at the individual level, it is less likely that receipt of a prescription for non-NSAIDs should influence the chance of being diagnosed with ICH and consequently included in this study as a case. The risk of selection bias was therefore small in this study.

Information problems

Misclassification of exposure, outcome, and confounder data may also have influenced our findings.

Data on exposures, outcome, and potential confounders were obtained from population-based registries within the health care system. Data from the Pharmaco-Epidemiological Prescription Database are collected prospectively through routine procedures at all the pharmacies in the county. The completeness of the data is probably high due to the reimbursement system that encourages the pharmacists to register the data correctly. Routine validity examinations have shown errors in the registration of the prescriber in 4% of the prescriptions (personal communication: Kirsten Nielsen, Department of Health Insurance, North Jutland County), but information on inaccuracies for other types of data is not available. It is also important to notice that we had no information on compliance or duration of use, i.e. redeeming a prescription was used as a proxy for actual use of a drug in our study, which is obviously not always a valid assumption. Furthermore, we lacked data on OTC use of non-aspirin NSAIDs. These issues may both lead to misclassification of the exposure. OTC use of non-aspirin NSAIDs was, however, limited to low-dose ibuprofen (200 mg per tablet), for which the sales constituted only

approximately 13.5% of the total non-aspirin NSAID sales in Denmark during the study period (personal communication: Janne Kampmann, The Danish Medicines Agency). OTC use was however probably modest among persons being prescribed non-NSAIDs since these persons receive a 50% refund when the ibuprofen is prescribed by a physician.

Data from the HPR are also collected routinely and may thus vary in completeness and data quality as described in study I. The validity of the ICH hospital discharge diagnoses has previously been reported to be high (22,23,28), but there may be substantial variation in the predictive value between specialised and non-specialised departments (23,Study I). We validated the register diagnosis of ICH in a random sample of our cases and found a high predictive value among patients discharged from departments with easy access to imaging facilities, i.e. a CT- or MR scanner, but a low predictive value for patients discharged from departments without easy access to such facilities. However, limiting the analyses to verified cases only or patients discharged from hospitals with easy access to imaging facilities, did not change the results. Furthermore, sensitivity analyses showed only modest changes in the risk estimates.

Combining data from the Pharmaco-Epidemiological Prescription Database and the HPR is, however, likely to improve the completeness and predictive value of a number of conditions, eg diabetes mellitus and hypertension (99). In summary, non-differential misclassification of both exposure and, in particular, confounder data may have occurred. As previously mentioned, non-differential misclassification of the exposure status would probably draw the relative risk estimates toward the null hypothesis, whereas non-differential misclassification of the potential confounder variables may lead to residual confounding and could result in both attenuation and inflation of the relative risk estimates.

The risk of recall bias is usually an important issue when interpreting the results from case-control studies based on data from questionnaires or interviews. However, this was not an issue

in our study, which was based on registry data. Thus, since all data on exposures and confounders used in this study were collected prospectively and before the date of admission for ICH, the collection of exposure data was not influenced by the outcome status. The risk of information bias was consequently minimal, which illustrates an important advantage of using registry-based data in case-control studies.

Confounding

Using the population-based registries, we were able to adjust for a number of conditions and factors predisposing to ICH. It is somewhat uncertain to which degree all of these conditions and factors are acting as "true" confounders, ie independently associated with both non-aspirin NSAID use and risk of ICH, when looking at the data published so far. Adjustment for these potential confounders also had only a minor effect on the relative risk estimates. We used proxy measures to control for a number of the potential confounders, eg discharge diagnoses of hypertension and chronic bronchitis and emphysema were used as proxies for blood pressure and smoking, respectively. The possibility of residual confounding could therefore not be excluded since the sensitivity of some of these proxy measures may not be high.

Since by definition cases had been hospitalised at least once, and information on potential confounding factors registered during the admission for ICH were also included in the analyses, we cannot exclude the possibility that more complete and maybe also more detailed data on potential confounders were available for cases than for controls as. To assess the impact on this potential nonequivalence of available data between cases and controls, we re-analysed the data based only on information obtained at admissions before the ICH event. Our results were be almost identical to the original analyses, ie the adjusted OR for receiving a prescription of non-aspirin NSAIDs 30 days before the ICH event was 1.14 (95% CI: 0.83-1.57) in the new

analyses vs. 1.12 (95% CI: 0.81-1.55) in the original analyses. These findings indicate that nonequivalence of data available for cases and control may not be a major problem in our study. Furthermore, by contrast with the extensive data available on previous hospital discharges and prescriptions, we had no data on other potential confounders, e.g. the indication on which the non-aspirin NSAIDs were prescribed and lifestyle factors such as diet, physical activity, etc. The overall credibility of the methodology used in our study was also supported by the fact that the risk estimates obtained for known risk factors for ICH, eg a history of hypertension and treatment with oral anticoagulants, were in line with estimates from the existing literature (130,131).

Chance- statistical precision

This study is to our knowledge the largest of its kind to date, but statistical imprecision was still seen for several of the risk estimates. The overall finding of no risk of being admitted to hospital for ICH among patients prescribed non-aspirin NSAIDs was also supported by the consistent subgroup analyses.

Conclusion

The most important strength of this study was the nested case-control design based on population-based registries with prospective data collection. The nested case-control design provided the possibility of calculating ORs, which could be interpreted as estimates of the incidence rate ratio. This measure of association can otherwise only be estimated in follow-up studies. Furthermore, the risk of selection and information bias appeared low in this study. Important weaknesses of the study included the risk of non-differential misclassification of outcome and non-aspirin-NSAID status, the risk of non-equivalent confounder information, and

the statistical imprecision of some of the relative risk estimates.

6.5. Discussion

In this population-based case-control study we found no substantial increase in overall risk of being admitted to hospital for ICH among patients prescribed non-aspirin NSAIDs. The absence of an association was evident in all subgroups examined, including both men and women, elderly, and hypertensive patients.

The results of the study are consistent with those of a recent case-control study by Thrift et al., in which no overall association between use of non-aspirin NSAIDs in the preceding fortnight and risk of ICH was found among 331 cases of ICH and 331 controls (OR = 0.85; 95% CI: 0.45-1.61) (37). Likewise, no increased risk of ICH was seen when stratifying for age, sex, or hypertension in the study by Thrift et al. The two studies differed in several ways; our study was based on prospectively collected population-based data and estimated the risk of ICH using different exposure time windows, whereas Thrift et al obtained detailed data using medical records and a questionnaire survey after the ICH event and focussed on a predefined exposure time window. In this perspective, the almost identical risk estimates obtained in these two studies, indicating no substantially increased risk of ICH associated with use of non-aspirin NSAIDs, is striking and reassuring.

Both aspirin and non-aspirin NSAIDs exert their anti-platelet actions by inhibiting cyclooxygenase and thereby blocking the formation of platelet-activating thromboxane A_2. Unlike aspirin, which causes a permanent inhibitory effect on cyclooxygenase persisting for the lifespan of the aspirin-exposed platelets (approximately 7-10 days), non-aspirin NSAIDs are reversible inhibitors of cyclooxygenase (132). Function of the enzyme, and thus the platelets, is

therefore restored as the drugs are cleared from circulation. Normal platelet function is consequently more rapidly re-established after discontinuation of non-asprin NSAIDs with shorter half-lives (133), whereas platelet effects of long-acting non-aspirin NSAIDs such as piroxicam may persist for several days after discontinuation (134). Indeed, substantial variability in extent and duration of the platelet inhibitory effect have been found among non-aspirin NSAIDs as indicated by ex vivo platelet aggregation (135). These aspects made it difficult to assess the relevant exposure time window for non-aspirin NSAIDs in our study. Furthermore, although no overall risk of ICH has been found in users of non-aspirin NSAIDs in our study and the study by Thrift et al (37), increased risk of ICH may still exist for specific drugs.
Further well-designed and well-carried out studies of sufficient size are needed on specific types of non-aspirin NSAIDs before more final conclusions on the safety of non-aspirin NSAIDs can be made.

6.6. Conclusion

In conclusion, patients prescribed non-aspirin NSAIDs were not at increased risk of being hospitalised for ICH. This reassuring finding was seen in all examined subgroups, including those with a higher baseline risk for ICH, such as the elderly and patients with a previous discharge diagnosis of hypertension. Although an increased risk of ICH in users of non-aspirin NSAIDs cannot be ruled out, the published data to date, including the present study, indicate that non-aspirin NSAIDs are at least not major contributors to the risk of ICH.

7. PERSPECTIVES

The clinical burden and costs of stroke are already impressive and are estimated to increase even further over the next few decades. Unfortunately, the gaps in our knowledge about the etiology, effective treatment and rational preventive measures for this complex disease entity are still

substantial. Thus, a dedicated combined effort from different scientific disciplines, including epidemiology, clinical medicine, pharmacology, and molecular biology, will be urgently needed in the near future. It is hoped that important contributions to this work may be provided by epidemiologists around the world, including the Scandinavian countries.

Although the potential fallacies and weaknesses of the observational study designs used in the studies of this thesis should be kept in mind, the failure of the randomised trials, and indeed of other research disciplines, to provide data on the issues studied in this thesis illustrates the need for and potential of further epidemiological efforts within this field.

Looking specifically at the potential risk factors studied in this thesis, ie dietary intake of fruit and vegetables, infection with *Chlamydia pneumoniae*, and use of non-aspirin NSAIDs, a range of unsolved issues remains to be solved before the role of these factors is clear.

If reduced intake of fruit and vegetables is to become further established as a risk factor for stroke, there will probably be a need for a well-designed and carefully carried out randomised trial. Furthermore, an important issue for future studies on diet and stroke will be the role of diet for the prognosis of patients with preexisting cerebrovascular disease, and in particular whether dietary changes after a stroke, e.g. increase in intake of fruit and vegetables, are associated with an improved outcome.

With respect to the existing methodological problems in sero-epidemiological studies of *Chlamydia pneumoniae* infection, it appears quite clear that the role of this factor cannot be clarified based on epidemiology alone. Based on the experience from the existing epidemiological studies, much stronger methodological studies may now be designed and carried out. However, the ideal sero-epidemiological study is most likely not feasible, at least not at the moment. Instead, experimental studies on animal models and randomised trials on humans using

antibiotics, and perhaps in the future a vaccine, will probably be of great importance.
By contrast, further data on the role of non-aspirin NSAIDs as a potential risk factor for ICH will be obtained most likely through observational studies on large populations. Future studies should focus on obtaining more detailed and valid data on actual use, potential confounders, and outcome status, preferably within a population-based setting.

Thus, given that the necessary planning is made and precautions are taken, stroke is amenable to epidemiological research. Combining data from different sources, e.g. primary data from cohort, case-control, or survey studies and secondary data from hospital discharge registries, mortality files, social registries, prescription registries, or clinical databases, may well prove an efficient and useful approach, that should be further explored in the future. The Scandinavian countries, and in particular Denmark, will probably continue to have unique possibilities, and therefore also the obligation, to pursue this line of research in order to provide answers to some of the many long-standing questions about risk factors for stroke, e.g. the role of diet, physical activity, obesity, psychological stress, drugs, etc. By collaborating with geneticists, molecular biologists and biochemists and others, it is hoped that this work may also be extended to include studies on genetic and biochemical risk factors/markers.

Although the impact of several risk factors, eg an adverse effect of a drug or polymorphism of a gene, may seem modest from a population perspective, it may be substantial for the persons exposed, and may be even higher for specific subgroups of the exposed, eg persons with a high susceptibility or persons exposed to unusual levels or combinations of exposures. The aim of this continued effort to unravel the aetiology of stroke should then be to establish more detailed knowledge about the mechanisms that may lead to stroke, and to make this knowledge available

for clinicians and the general population in order to ensure that reasonable and more efficient preventive measures may be taken at both a population level and an individual level.

8. SUMMARY

The aims of this thesis were; 1) to examine whether routinely collected data from the health care system in Denmark could be used as a tool for stroke epidemiology, with special reference to the data quality of hospital discharge diagnoses of stroke and TIA (study I), and 2) to examine the role of three potential risk factors for stroke through studies on the association between intake of fruit and vegetables and the risk of IS (study II), on the association between previous/chronic infection with *Chlamydia pneumoniae* and the risk of IS (study II), and on the association between use of non-aspirin NSAIDs and the risk of ICH (study IV).

The studies included in this thesis were based on data from a number of different sources. Linkage between these data sources was done using the civil registry number, a unique, personal identification number given to all Danish citizens at birth or upon immigration since 1968. Studies I-III were based on data from the follow-up study "Diet, Cancer and Health". This cohort of 57,055 Danish men and women was established during 1993-1997 with the primary aim of studying the aetiological role of diet in cancer risk. Detailed data about diet and other lifestyle factors, including smoking habits, alcohol intake, physical activity, medical history, and education, were collected at baseline. Data on body proportions, blood pressure, and total serum cholesterol were also measured at baseline and samples of biological material, including blood samples, were obtained and stored in a biological bank.

We retrieved all probable incident cases of stroke and TIA (n=581) within the cohort that were registered in the Danish National Registry of Patients. Medical records and hospital discharge summaries were retrieved and reviewed using a standardised form. The overall predictive value of the discharge diagnoses of stroke were reasonable, though, we found substantial variation in the predictive value according to subtypes of stroke and types of hospital department. We conclude that stroke and TIA diagnoses in the Danish National Registry of Patients should be

used with caution in epidemiological research since the low predictive value for some diagnostic subgroups may lead to serious misclassification and biased results (study I).

The association between intake of fruit and vegetables and the risk of IS was examined in a follow-up study based on data from the 54,506 men and women included in "Diet, Cancer and Health", who had not previously been diagnosed with cardiovascular disease at baseline. We identified 266 hospitalised cases of IS during follow-up. After adjustment for potential confounders, persons in the top quintile of fruit and vegetable intake had a reduced risk of IS compared with the persons in the bottom quintile. The risk was primarily reduced among persons with a high intake of fruit (study II). An increased intake of fruit and vegetables, and in particular fruit, may reduce the risk of IS.

The association between previous/chronic infection with *Chlamydia pneumoniae* and the risk of IS was examined in a case-control study nested within the "Diet, Cancer and Health" cohort. The study included 254 cases with IS, i.e. from which blood samples were available, and 254 controls matched by age and sex. Titers of IgA and IgG antibodies specific for *Chlamydia pneumoniae* were measured by an indirect solid-phase enzyme immunoassay. The combination of a positive IgA titer (>= 1:16) and IgG titer (>=1:64) was associated with an increased risk of IS. The highest risk was found for IS due to large-artery atherosclerosis. The risk of IS varied according to sex and the chosen cut-off values for the antibody titers. These issues should be considered when planning and interpreting studies on *Chlamydia pneumoniae* and the risk of stroke (study III).

The association between the use of non-aspirin NSAIDs and the risk of ICH was examined in a

population-based nested case-control study in North Jutland County, Denmark. We identified 912 cases with a first diagnosis of ICH during the period 1991-1999 in the County Hospital Patient Register and 9059 population-based controls matched on age and sex. All redeemed prescriptions for non-asprin NSAIDs before the date of hospital admission for ICH were identified for both cases and controls using the Pharmaco-Epidemiological Prescription Database of North Jutland County. Data on potential confounders were also obtained from the County Hospital Patient Register and the Pharmaco-Epidemiological Prescription Database of North Jutland County. There was no increase in overall risk of being admitted to hospital for ICH among patients prescribed non-aspirin NSAIDs. The absence of an association was evident in all subgroups examined, including men and women, the elderly, and hypertensive patients. Although an increased risk of ICH among users of non-aspirin NSAIDs cannot be ruled out, the published data to date, including the present study, indicate that non-aspirin NSAIDs are at least not major contributors to the risk of ICH (study IV).

The data sources used in this thesis together constitute a potentially important research resource when examining risk factors for stroke. However, the results from this thesis also illustrate the complexity of stroke as a disease entity and the substantial methodological challenges that remain within this research area.

9. DANSK RESUMÉ

Ph.D. afhandlingen omhandler potentielle risikofaktorer for apopleksi og er baseret på fire originalarbejder. Formålet med afhandlingen var at undersøge 1) den prædiktive værdi af apopleksi og transitorisk cerebral iskæmi diagnoser i Landspatientregisteret (LPR) [studie I], 2) associationen mellem indtag af frugt og grønt og risikoen for iskæmisk apopleksi [studie II], 3) associationen mellem infektion med *Chlamydia pneumoniae* og risikoen for iskæmisk apopleksi [studie III], samt 4) associationen mellem brug af non-steroide antiinflammatoriske farmaka (NSAID) og risikoen for intracerebral hæmorrhagi [studie IV].

De tre første studier er baseret på data fra befolkningsundersøgelsen "Kost, kræft og helbred". Denne prospektive kohorte, som omfatter 57.055 mænd og kvinder, blev etableret mellem 1993 og 1997 i København og Århus med det primære formål at analysere kostens ætiologiske rolle ved en række kræftformer. For alle deltagere blev der indsamlet detaljerede oplysninger vedr. kost samt andre faktorer inkl. rygevaner, alkoholindtag, fysisk aktivitet, medicinsk anamnese samt uddannelse ved inkludering i undersøgelsen. Der blev endvidere registreret en række antropometriske mål, helkrops-impedans, blodtryk og serum-cholesterol og samt indsamlet biologisk materiale, inkl. blodprøver, som opbevares i en biologisk bank.

Via kobling til LPR blev alle kohortemedlemmer med en incident cerebrovaskulær udskrivelsesdiagnose identificeret (n=581). Diagnoserne blev herefter valideret v.h.a. en standardiseret gennemgang af journaler. Apopleksidiagnosen i LPR havde overordnet en acceptabel positiv prædiktiv værdi. Vi fandt dog betydelig variation i den prædiktive værdi mellem forskellige subdiagnoser og forskellige typer af sygehusafdelinger [studie I].

Associationen mellem indtag af frugt og grønt og risikoen for iskæmisk apopleksi blev undersøgt i et follow-up design blandt de 54.506 mænd og kvinder i "Kost, kræft og helbred", som ikke havde været hospitaliseret med apopleksi eller kardiovaskulær sygdom forud for inklusion i kohorten. Vi fandt 266 verificerede tilfælde af iskæmisk apopleksi. Kohorten blev opdelt i kvintiler efter indtag af frugt og grønt. Efter justering for mulige confoundere, havde personer i den øverste kvintil af frugt og grønt indtag en nedsat risiko for iskæmisk apopleksi sammenlignet personer i den nederste kvintil. Den nedsatte risiko for iskæmisk apopleksi var mest fremtrædende blandt personer med det højeste indtag af frugt [studie II].

Associationen mellem infektion med *Chlamydia pneumoniae* og risikoen blev undersøgt i et case-kontrol studie indlejret i et follow-up studie. I studiet indgik 254 cases med iskæmisk apopleksi, d.v.s. cases hvor der var biologisk materiale til rådighed, samt 254 køns- og aldersmatched kontrolpersoner fra "Kost, kræft og helbred". IgA og IgG antistoffer mod *Chlamydia pneumoniae* blev bestemt på plasmaprøver fra den biologiske bank v.h.a. ELISA teknik. Kombinationen af en positiv IgA (>= 1:16) og IgG (>=1:64) titer var associeret med en øget risiko for iskæmisk apopleksi. Risikoen syntes særligt øget for iskæmisk apopleksi p.g.a. stor-kars aterosklerose. Associationen mellem *Chlamydia pneumoniae* og risikoen for iskæmisk apopleksi var dog afhængig af hvilken cut-off værdi der anvendtes for antistoftiteren. Endvidere var køn tilsyneladende en effekt-modifikator [studie III].

Associationen mellem brugen af NSAID og risikoen for intracerebral hæmorrhagi blev undersøgt i et populationsbaseret case-kontrol studie indlejret en kohhorte bestående af hele befolkningen i Nordjyllands Amt. Studiet var baseret på data fra det Regionale Nordjydske Landspatientregister og Lægemiddeldatabasen i Nordjyllands Amt. Vi identificerede 912 tilfælde

af første gangs intracerebral hæmorrhagi og 9059 køns- og aldersmatchede populationsbaserede kontrolpersoner i perioden 1991-1999. På en ca. 10% stikprøve af cases blev diagnosen valideret ved journalgennemgang og fundet tilfredsstillende. Alle recepter på NSAID forud for indlæggelsesdatoen for intracerebral hæmorrhagi blev herefter identificeret. Information om potentielle confoundere blev indhentet fra det Regionale Nordjydske Landspatientregister og Lægemiddeldatabasen i Nordjyllands Amt. Vi fandt generelt ingen sammenhæng mellem brugen af NSAID præparater og risikoen for intracerebral hæmorrhagi uanset hvilket tidsvindue som blev anvendt. Ej heller fandtes nogen øget risiko efter stratificering for køn, alder og tidligere udskrivelsesdiagnoser for hypertension [studie 4].

Datakilderne anvendt i denne Ph.D. afhandling udgør tilsammen en potentiel vigtig forskningsressource med henblik på udforskning af risikofaktorer for apopleksi. Resultaterne af de ovenævnte studier illustrerer dog også, hvor kompleks en sygdomsenhed apopleksi er og de heraf følgende betydelige metodologiske udfordringer som resterer indenfor dette forskningsfelt.

REFERENCES

1. Oppenheim H. Lehrbuch der Nervenkrankheiten für Ärtzte unde Studierende. S. Karger, 1913, Berlin.

2. Hatano S. Experience from a multicentre stroke register: a preliminary report. Bulletin of the World Health Organization 1976; 54: 541-3.

3. Warlow CP. Epidemiology of stroke. Lancet 1998; 352 (suppl III): 1-4.

4. Sudlow CLM, Warlow CP, for the International Stroke Incidence Collaboration. Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. Stroke 1997; 28: 491-9.

5. Bonita R. Epidemiology of stroke. Lancet 1992; 339: 342-4.

6. Rothwell PM. The high cost of not funding stroke research: a comparison with heart disease and cancer. Lancet 2001; 357: 1612-6.

7. Murray CJL, Lopez AD. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Boston: Harvard University Press, 1996.

8. Sundhedssektoren i tal- 2000. Sundhedsstyrelsen, København 2001.

9.Taylor TN, Davis PH, Torner JC et al. Lifetime cost of stroke in the United States. Stroke 1996;27:1459-66.

10. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. The Cochrane Library, Issue 3, 2001. Oxford: Update Software.

11. Wardlaw JM, del Zoppo G, Yamaguchi T. Thrombolysis for acute ischaemic stroke. The Cochrane Library, Issue 4, 2001. Oxford: Update Software.

12. Jørgensen HS, Nakayama H, Kammersgaard LP, Raaschou HO, Olsen TS. Predicted impact of intravenous thrombolysis on prognosis of general population of stroke patients: simulation model. BMJ 1999; 319: 288-9.

13. Warlow CP, Dennis MS, van Gijn J et al. Stroke - a practical guide to management. Second Edition. Blackwell Science, Oxford 2001.

14. Willett WC, Colditz GA. Approaches for conducting large cohort studies. Epidemiol Rev 1998; 20: 91-9.

15. Elkind MS, Sacco RL. Stroke risk factors and stroke prevention. Semin Neurol 1998; 18: 429-40.

16. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol 1996; 25; 435-42.

17. Lauderdale DS, Furner SE, Miles TP, Goldberg J. Epidemiologic uses of Medicare data. Epidemiol Rev 1993; 15: 319-27.

18. Bright RA, Avorn J, Everitt DE. MEDICAID data as a resource for epidemiologic studies: strengths and limitations. J Clin Epidemiol 1989; 42: 937-45.

19. Roos LL, Roos NP, Fisher ES, Buholz TA. Strengths and weakness of health insurance data systems for assessing outcomes. In: Gelijns AC, Ed. Medical Innovation at the Crossroads. Volume I. Modern Methods of Clinical Investigations. Washington DC: National Academy Press, 1990: 47-67.

20. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev 1980; 2: 210-20.

21. Iso H, Jacobs DR Jr, Goldman L. Accuracy of death certificate diagnosis of intracranial hemorrhage and nonhemorrhagic stroke. Am J Epidemiol 1990; 132: 993-8.

22. Leppälä JM, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. Eur J Epidemiol 1999; 15: 155-60.

23. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. Cerebrovasc Dis 1999; 9: 224-30.

24. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994; 25: 2348-55.

25. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. Neuroepidemiology 1992; 11: 204-13.

26. Lindblad U, Råstam L, Ranstam J, Peterson M. Validity of register data on acute myocardial infarction and acute stroke: The Skaraborg Hypertension Project. Scand J Soc Med 1993; 21: 3-9.

27. Mähönen M, Salomaa V, Keskimäki I et al. The feasibility of combining data from routine Hospital Discharge and Causes-of-Death Registers for epidemiological studies on stroke. Eur J Epidemiol 2000; 16: 815-7.

28. Ellekjær H, Holmen J, Krüger Ø, Terent A. Identification of incident stroke in Norway. Stroke 1999; 30: 56-60.

29. Gaist D, Væth M, Tsiropoulos I et al. Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. BMJ 2000; 320: 141-5.

30. Sørensen HT. Regional administrative health registries as a resource in clinical epidemiology. Int J Risk Safe 1997; 10: 1-22.

31. Frank L. When the entire country is a cohort. Science 2000; 287: 2398-9.

32. Skrabanek P. Risk-factor epidemiology: science or non-science? In: Anderson D (editor). Health, lifestyle and environment. Social Affairs Unit, London, 1991.

33. Davey Smith G. Reflections on the limitations to epidemiology. J Clin Epidemiol 2001; 54: 325-31.

34. Gillman MW, Cupples LA, Gagnon D et al. Protective effect of fruits and vegetables on development of stroke in men. JAMA 1995; 273: 1113-7.

35. Iso H, Rexrode KM, Stampfer MJ et al. Intake of fish and omega-3 fatty acids and risk of stroke in women. JAMA 2001; 285: 304-12.

36. Truelsen T, Grobaek M, Schnohr P, Boysen G. Intake of beer, wine, and spirits and risk of stroke. Stroke 1998; 29: 2467-72.

37. Thrift AG, McNeil JJ, Forbes A, Donnan GA. Risk of primary intracerebral haemorrhage associated with aspirin and non-steroidal anti-inflammatory drugs: case-control study. BMJ 1999; 318: 759-64.

38. Kernan WN, Viscoli CM, Brass LM et al. Phenylpropanolamine and the risk of hemorrhagic stroke. N Engl J Med 2000; 343: 1826-32.

39. Wimmer MLJ, Sandmann-Strupp R, Saikku P, Haberl RL. Association of chlamydial infection with cerebrovascular disease. Stroke 1996; 27: 2207-10.

40. Perry IJ, Refsum H, Morris RW, Ebrahim SB, Ueland PM, Shaper AG. Prospective study of serum total homocysteine concentration and risk of stroke in middleaged British men. Lancet 1995; 346: 1395-8.

41. Syme SL, Marmot MG, Kagan A, Kato H, Rhoads G. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: Introduction. Am J Epidemiol 1975; 102: 477-80.

42. Kimura N. Changing patterns of coronary heart disease, stroke, and nutrient intake in Japan. Preventive Medicine 1983; 12: 222-7.

43. Omura T, Hisamatsu S, Takizawa, Minowa M, Yanagawa H, Shigematsu I. Geographical distribution of cerebrovascular disease mortality and food intakes in Japan. Soc Sci Med 1987; 24: 401-7.

44. Khaw KT, Barrett-Connor E. Dietary potassium and stroke-associated mortality: a 12-year prospective population study. N Engl J Med 1987; 316: 235-40.

45. Gale CR, Martyn CN, Winter PD, Cooper C. Vitamin C and risk of death from stroke and coronary heart disease in cohort of elderly people. BMJ 1995; 310: 1563-6.

46. Keli SO, Hertog MGL, Feskens EJM, Kromhout D. Dietary flavonoids, antioxidant vitamins, and incidence of stroke. Arch Intern Med 1996; 154: 637-42.

47. Hirvonen T, Virtamo J, Korhonen P, Albanes D, Pietinen P. Intake of flavonoids, carotenoids, vitamin C and E, and risk of stroke in male smokers. Stroke 2000; 31: 2301-6.

48. Ascherio A, Rimm EB, Hernán MA et al. Intake of potassium, magnesium, calcium, and fiber and risk of stroke among US men. Circulation 1998; 98: 1198-204.

49. Gillman MW, Cupples LA, Millen BE, Ellison RC, Wolf PA. Inverse association of dietary fat with development of ischemic stroke in men. JAMA 1997; 278: 2145-50.

50. Iso H, Stampfer MJ, Manson JE et al. Prospective study of fat and protein intake and risk of intraparenchymal hemorrhage in women. Circulation 2001; 103: 856-63.

51. Joshipura KJ, Ascherio A, Manson JE et al. Fruit and vegetable intake in relation to risk of ischemic stroke. JAMA 1999; 282: 1233-9.

52. Ness AR, Powles JW. Fruit and vegetables, and cardiovascular disease: a review. Int J Epidemiol 1997; 26: 1-13.

53. Ness AR, Powles JW. The role of diet, fruit and vegetables and antioxidants in the aetiology of stroke. J Cardiovasc Risk 1999; 6: 229-34.

54. Jacobs DR Jr, Murtaugh MA. It's more than an apple a day: an appropriately processed plant-centered dietary pattern may be good for your health. Am J Clin Nutr 2000; 72: 899-900.

55. Steinberg D, Witztum JL. Is the oxidative modification hypothesis relevant to human atherosclerosis ? Circulation 2002; 105: 2107-11.

56. Hirayama T. Nutrition and cancer-- a large cohort study. Prog Clin Biol Res 1986; 206: 299-311.

57. Key TJA, Thorogood M, Appleby PN, Burr ML. Dietary habits and mortality in 11000 vegetarians and health conscious people: results of a 17 year follow up. BMJ 1996; 313: 775-9.

58. Singh RB, Rastogi SS, Rakesh V et al. Randomised controlled trial of cardioprotective diet in patients with recent myocardial infarction: results of one year follow up. BMJ 1992; 304: 1015-9.

59. Appel LJ, Moore TJ, Obarzanek E et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N Engl J Med 1997; 336: 1117-24.

60. De Lorgeril M, Salen P, Martin JL, Monjaud I, Delaye J, Mamelle N. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction. Circulation 1999; 99: 779-85.

61. Ross R. Atherosclerosis: an inflammatory disease. N Engl J Med 1999; 340: 115-26.

62. Danesh J, Collins R, Pet R. Chronic infections and coronary heart disease: is there a link? Lancet 1997; 350: 430-6.

63. Danesh J, Whincup P, Walker M et al. Chlamydia pneumoniae IgG titres and coronary heart disease: prospective study and meta-analysis. BMJ 2000; 321: 208-13.

64. Cook PJ, Honeybourne D, Lip GYH, Beevers DG, Wise R, Davies P. *Chlamydia pneumoniae* antibody titers are significantly associated with acute stroke and transient cerebral ischemia. Stroke 1998; 29: 404-10.

65. Fagerberg B, Gnarpe J, Gnarpe H, Agewall S, Wikstrand J. Chlamydia pneumoniae but not cytomegalovirus antibodies are associated with future risk of stroke and cardiovascular disease. Stroke 1999; 30: 299-305.

66. Elkind MSV, Lin I-F, Grayston JT, Sacco RL. Chlamydia pneumoniae and the risk of first ischemic stroke. Stroke 2000; 31: 1521-5.

67. Glader CA, Stegmayr B, Boman J et al. *Chlamydia pneumoniae* antibodies and high lipoprotein(a) levels do not predict ischemic cerebral infarctions. Stroke 1999; 30: 2013-8.

68. Heuschmann PU, Neureiter D, Gesslein M et al. Association between infection with *Helicobacter pylori* and *Chlamydia pneumoniae* and risk of ischemic stroke subtypes. Stroke 2001; 32: 2253-8.

69. Melnick SL, Shahar E, Folsom AR et al. Past infection by *Chlamydia pneumoniae* strain TWAR and asymptomatic carotid atherosclerosis. Am J Med 1993; 95: 499-504.

70. Grayston JT, Kuo CC, Coulson AS et al. Chlamydiae pneumoniae (TWAR) in atherosclerosis of the carotid artery. Circulation 1995; 92: 3397-400.

71. Schmidt C, Hulthe J, Wikstrand J et al. Chlamydia pneumoniae seropositivity is associated with carotid artery intima-media thickness. Stroke 2000; 31: 1526-31.

72. Sander D, Winbeck K, Klingelhöfer J, Etgen T, Conrad B. Enhanced progression of early carotid atherosclerosis is related to *Chlamydia pneumoniae* (Taiwan Acute Respiratory) seropositivity. Circulation 2001; 103: 1390-5.

73. Markus HS, Sitzer M, Carrington D, Mendall MA, Steinmetz H. *Chlamydia pneumoniae* infection and early asymptomatic carotid atherosclerosis. Circulation 1999; 100: 832-7.

74. Coles KA, Plant AJ, Riley TV, Smith DW, McQuillan BM, Thompson PL. Lack of association between seropositivity to Chlamydia pneumoniae and carotid atherosclerosis. Am J Cardiol 1999; 84: 825-8.

75. Baron JA, Sandler RS. Nonsteroidal anti-inflammatory drugs and cancer prevention. Annu Rev Med 2000; 51: 511-23.

76. Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 1999; 340: 1888-99.

77. Raskin JB. Gastrointestinal effects of nonsteroidal anti-inflammatory therapy. Am J Med 1999; 106(5B): 3S-12S.

78. Mellemkjær L, Blot WJ, Sørensen HT et al. Upper gastrointestinal bleeding among users of NSAIDs: a population-based cohort study in Denmark. Br J Clin Pharmacol 2002; 53: 173-81.

79. Schoen RT, Vender RJ. Mechanisms of nonsteroidal anti-inflammatory drug induced gastric damage. Am J Med 1989; 86: 449-58.

80. Schafer AI. Effects of nonsteroidal anti-inflammatory therapy on platelets. Am J Med 1999; 106(5B): 25S-36S.

81. He J, Whelton PK, Vu B, Klag MJ. Aspirin and risk of hemorrhagic stroke: a meta-analysis of randomized controlled trials. JAMA 1998; 280: 1930-5.

82. The Steering Committee of the Physicians Health Study Research Group. Preliminary report: findings from the aspirin component of the ongoing Physicians Health Study. N Engl J Med 1988; 318: 262-4.

83. Peto R, Gray R, Collins R et al. Randomised trial of prophylactic daily aspirin in British male doctors. BMJ 1988; 296: 313-6.

84. Antiplatelet Trialists Collaboration. Secondary prevention of vascular disease by prolonged antiplatelet treatment. BMJ 1988; 296: 320-31.

85. Tjønneland A, Overvad K. Diet, cancer and health - a prospective cohort study and biological bank in Denmark. Ugeskr Laeger 2000; 162: 350-4.

86. Overvad K, Tjønneland A, Haraldsdóttir J, Bang S, Ewertz M, Jensen OM. Development of a semiquantitative food frequency questionnaire to assess food, energy and nutrient intake in Denmark. Int J Epidemiol 1991; 20: 900-5.

87 Tjønneland A, Overvad K, Haraldsdóttir J, Bang S, Ewertz M, Jensen OM. Validation of a semiquantitative food frequency questionnaire developed in Denmark. Int J Epidemiol 1991; 20: 906-12.

88. Andersen TF, Madsen M, Jørgensen J, Mellemkjær L, Olsen JH. The Danish National Hospital Register. Dan Med Bull 1999; 46: 263-8.

89. Nielsen GL, Sørensen HT, Zhou W, Steffensen FH, Olsen J. The pharmacoepidemiologic prescription database of North Jutland - a valid tool in pharmacoepidemiological research. Int J Risk Safety Med 1997; 10: 203-5.

90. Guidelines for the ATC and DDD. World Health Organization Collaborative Centre for Drugs Statistics Methodology. Oslo, 1996.

91. Adams HP Jr, Bendixen BH, Kappelle LJ et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. Stroke 1993; 24: 35-41.

92. Kraaijeveld CL, van Gijn J, Schouten HJA, Staal A. Interobserver agreement for transient ischemic attacks. Stroke 1984; 15: 723-5.

93. Sandercock PAG. Recent developments in the diagnosis and management of patients with transient ischemic attacks and minor ischemic strokes. Q J Med 1991; 78:101-12.

94. Fletcher RH, Fletcher SZ, Wagner EH. Clinical Epidemiology. The essentials. Third

edition.Williams & Wilkins. Baltimore, 1996.

95. Last JM (Ed.). A dictionary of epidemiology. Fourth edition. Oxford University Press. New York, 2001.

96. Rothman KJ, Greenland S. Modern Epidemiology. Second edition. Lippincott-Raven. Philadelphia, 1998.

97. Truelsen T, Prescott E, Grønbæk M, Schnohr P, Boysen G. Trends in stroke incidence. Stroke 1997; 28: 1903-7.

98. Mähönen M, Salomaa V, Brommels M et al. The validity of hospital discharge register data on coronary heart disease in Finland. Eur J Epidemiol 1997; 13: 403-15.

99. Nielsen GL, Sorensen HT, Pedersen AB, Sabroe S. Analyses of data quality in registries concerning diabetes mellitus - a comparison between a population based hospital discharge and an insulin prescription registry. J Med Syst 1996; 20: 1-10.

100. Lawrenson R, Todd JC, Leydon GM, Williams TJ, Farmer RD. Validation of the diagnosis of venous thromboembolism in general practice database studies. Br J Clin Pharmacol 2000; 49: 591-6.

101. Rawson NS, Malcolm E. Validity of the recording of ischaemic heart disease and chronic

obstructive pulmonary disease in the Saskatchewan health care datafiles. Stat Med 1995; 14: 2627-43.

102. Tennis P, Bombardier C, Malcolm E, Downey W. Validity of rheumatoid arthritis diagnoses listed in the Saskatchewan Hospital Separations Database. J Clin Epidemiol 1993; 46: 675-83.

103. Cattaruzzi C, Grazia Troncon M, Agostinis L, Garcia Rodriguez LA. Positive predictive value of ICD-9th codes for upper gastrointestinal bleeding and perforation in the Sistema Informativo Sanitario Regionale Database. J Clin Epidemiol 1999; 6: 499-502.

104. Nielsen HW, Tuchsen F, Jensen MV. Validity of the diagnosis essential hypertension in The National Patient Registry. Ugeskr Laeger 1996;158:163-7.

105. Devantier A, Kjer JJ. The National Patient Register- a research tool ? Ugeskr Laeger 1991;153:516-7.

106. Madsen KM, Schonheyder HC, Kristensen B, Nielsen GL, Sorensen HT. Can hospital discharge diagnosis be used for surveillance of bacteremia? A data quality study of a Danish hospital discharge registry. Infect Control Hosp Epidemiol 1998;19:175-80.

107. Lauritsen J. FoodCalc 1.3. 1998. World Wide Web: http://www.foodcalc.dk (accessed 2 January 2002).

108. Warming DL, Fagt S. Danskernes kostvaner, Teknisk rapport 2. Copenhagen. The

Danish Veterinary and Food Administration, 1997.

109. Slimani N, Deharveng G, Charrondiere RU et al. Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. European Prospective Investigation into Cancer and Nutrition. Comput Methods Programs Biomed 1999;58:251-66.

110. Greenland S. Modelling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-9.

111. Willett WC, Stampfer MJ. Total energy intake: implications for epidemiologic analysis. Am J Epidemiology 1986;124:17-27.

112. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. Second edition. Oxford University Press, 1996. Oxford.

113. Willett W. Nutritional Epidemiology. Second edition. Oxford University Press, 1998. Oxford.

114. Tjønneland A, Grønbæk M, Stripp C, Overvad K. Wine intake and diet in a random sample of 48763 Danish men and women. Am J Clin Nutr 1999; 69: 49-54.

115. Tompson RL, Margetts BM, Speller VM, McVey D. The Health Education Authority's health and lifestyle survey 1993: who are the low fruit and vegetable consumers ? J Epidemiol Community Health 1999; 53: 294-9.

116. Hennekens CH, Buring JE, Manson JE et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. N Engl J Med 1996; 334: 1145-9.

117. The Alpha-Tocopherol Beta Carotene cancer prevention study group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N Engl J Med 1994; 330: 1029-35.

118. Omenn GS, Goodman GE, Thornquist MD et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med 1996; 334: 1150-5.

119. Blot WJ, Li J, Taylor PR, Guo W et al. Nutrition intervention trials in Linxian, China: supplementation with specific vitamin/mineral combinations, cancer incidence and disease specific mortality in the general population. J Natl Cancer Inst 1993; 85: 1483-92.

120. Mark SD, Wang W, Fraumeni JF Jr et al. Lowered risks of hypertension and cerebrovascular disease after vitamin/mineral supplementation: the Linxian Nutrition Intervention Trial. Am J Epidemiol 1996; 143: 658-64.

121. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies, I: principles. Am J Epidemiol 1992; 135: 1019-28.

122. Persson K, Boman J. Comparison of five serologic tests for diagnosis of acute infections by *C. pneumoniae*. Clin Diagn Lab Immunol 2000; 7: 739-44.

123. Nieto FJ, Folsom AR, Sorlie PD, Grayston JT, Wang SP, Chambless LE. Chlamydia pneumoniae infection and incident coronary heart disease: the Atherosclerosis Risk in Communities Study. Am J Epidemiol 1999; 150: 149-56.

124. Boman J, Hammerschlag MR. Chlamydia pneumonia and atherosclerosis: critical assessment of diagnostic methods and relevance to treatment studies. Clin Microbiol Rev 2002; 15: 1-20.

125. Epstein SE, Zhu J. Lack of association of infectious agents with risk of future myocardial infarction and stroke. Definitive evidence disproving the infection/coronary artery disease hypothesis ? Circulation 1999; 100: 1366-8.

126. Siscovick DS, Schwartz SM, Caps M, Wang SP, Grayston JT. Chlamydia pneumoniae and atherosclerotic risk in populations: the role of seroepidemiology. J Infect Dis 2000; 181(Suppl 3): S417-20.

127. Dunne MW. Rationale and design of a secondary prevention trial of antibiotic use in patients after myocardial infarction: The WIZARD (Weekly Intervention with Zithromax [Azithromycin] for Atherosclerosis and Its Related Disorders) Trial. J Infect Dis 2000; 181(Suppl 3): S572-8.

128. Moazed TC, Kuo CC, Grayston JT, Campbell LA. Evidence of systemic dissemination of Chlamydia pneumoniae via macrophages in the mouse. J Infect Dis 1998; 117: 1322-5.

129. Kovanen PT, Mänttäri M, Palosuo T, Manninen V, Aho K. Prediction of myocardial infarction in dyslipidemic men by elevated levels of immunoglobulin classes A, E, and G, but not M. Arch Intern Med 1998; 158: 1434-9.

130. Brott T, Thalinger K, Hertzberg V. Hypertension as a risk factor for spontaneous intracerebral hemorrhage. Stroke 1986; 17: 1078-83.

131. Hart RG, Boop BS, Anderson DC. Oral anticoagulants and intracranial hemorrhage. Stroke 1995; 26: 1471-7.

132. Schafer AI. Effects of nonsteroidal antiinflammatory drugs on platelet function and systemic hemostasis. J Clin Pharmacol 1995; 35: 209-19.

133. Small RE, Johnson SM. Consideration of platelet effects in the selection of an anti-inflammatory agent. Clin Pharm 1987; 6: 756-7.

134. Weintraub M, Case KR, Kroening B. Effects of piroxicam on platelet aggregation. Clin

Pharmacol Ther 1978; 23: 134-5.

135. Cronberg S, Wallmark E, Soderberg I. Effect on platelet aggregation of oral administration of 10 non-steroidal analgesics to humans. Scand J Haematol 1984; 33: 155-9.