From DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIOSTATISTICS
Karolinska Institutet, Stockholm, Sweden

CAN PARKINSON'S DISEASE BE PREVENTED?

- EPIDEMIOLOGICAL EVIDENCE ON LIFESTYLE FACTORS

Fei Yang

杨 飞

Stockholm 2017

All previously published papers were reproduced with permission from the publisher.
Cover photo by Barbro Sandin.
Published by Karolinska Institutet.
Printed by Eprint AB 2017

ISBN 978-91-7676-639-2

Can Parkinson's disease be prevented?

- Epidemiological evidence on lifestyle factors

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Fei Yang

Principal Supervisor:

Associate Professor Karin Wirdefeldt
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):

Professor Nancy Pedersen
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Associate Professor Fang Fang
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:

Professor Beate Ritz
University of California, Los Angeles
School of Public Health
Department of Epidemiology

Examination Board:

Associate Professor Shahram Bahmanyar
Karolinska Institutet
Department of Medicine, Solna
Centre for Pharmacoepidemiology

Associate Professor Lars Rylander
Lund University
Department of Laboratory Medicine, Lund
Division of Occupational and Environmental Medicine

Professor Johan Lökk
Karolinska Institutet
Department of Neurobiology, Care Sciences and Society
Division of Clinical Geriatrics

To my beloved family

献给我最爱的家人

(By Shuyang Yao)

特别是陪伴着我的傻宝、傻蛋和蛋卷 <3

ABSTRACT

The etiology of Parkinson's disease (PD) is largely unknown, but epidemiological studies have provided robust evidence that environmental factors have a central role in PD pathogenesis and progression. The studies included in this thesis examined associations between several lifestyle factors and subsequent risk of PD.

In **study I**, we prospectively investigated the association between physical activity and risk of PD in a Swedish cohort of 43,368 men and women. We found that higher levels of physical activity were associated with lower risk of PD. In addition, when pooling our results and all previously prospective cohort studies on this topic using meta-analysis, there was a 34% lower risk of PD for the highest level of physical activity compared with the lowest level.

In the second study (**study II**), we examined the effect of Swedish moist smokeless tobacco (i.e. snus in Swedish) on future risk of PD using individual participant data from seven Swedish cohorts, consisting of 348,601 men. Our pooled results confirmed that non-smoking men who used snus had a substantially lower risk of PD compared with never snus users. In addition, a dose-response relationship between snus use and risk of PD was indicated.

A third series of analyses evaluated the effect of dietary antioxidants on PD risk (**study III**). Using comprehensive dietary information on 45,837 men and 38,937 women from two population-based cohorts, we concluded that intake of dietary vitamin E and β-carotene was associated with a lower risk of PD.

In the last study (**study IV**), we analyzed Swedish census data with more than 4.6 million individuals regarding the potential effect of socioeconomic status on PD risk, as well as all-cause mortality before and after PD diagnosis. PD risk was higher in individuals with higher socioeconomic status, and all-cause mortality was higher in individuals with lower socioeconomic status regardless of PD status. However, the impact of socioeconomic status on all-cause mortality was weaker in PD patients than in PD-free individuals.

In conclusion, our findings contribute to the existing epidemiological evidence on the associations between lifestyle factors and risk of PD. Our results from prospective cohort studies suggest that lifestyle factors can influence future risk of PD. Interventions on these lifestyle factors may have the possibility to prevent the disease eventually.

LIST OF SCIENTIFIC PAPERS

I. **Yang F**, Trolle Lagerros Y, Bellocco R, Adami HO, Fang F, Pedersen NL, Wirdefeldt K. Physical activity and risk of Parkinson's disease in the Swedish National March Cohort. *Brain*. Feb 2015;138(Pt 2):269-275. doi: 10.1093/brain/awu323

II. **Yang F**, Pedersen NL, Ye W, Liu Z, Norberg M, Forsgren L, Trolle Lagerros Y, Bellocco R, Alfredsson L, Knutsson A, Jansson JH, Wennberg P, Galanti MR, Lager A, Araghi M, Lundberg M, Magnusson C, Wirdefeldt K. Moist smokeless tobacco (Snus) use and risk of Parkinson's disease. *International Journal of Epidemiology*. Dec 2016. doi: 10.1093/ije/dyw294

III. **Yang F**, Wolk A, Håkansson N, Pedersen NL, Wirdefeldt K. Dietary antioxidants and risk of Parkinson's disease in two population-based cohorts. (Submitted)

IV. **Yang F**, Johansson AL, Pedersen NL, Fang F, Gatz M, Wirdefeldt K. Socioeconomic status in relation to Parkinson's disease risk and mortality: A population-based prospective study. *Medicine*. Jul 2016;95(30):e4337. doi: 10.1097/MD.0000000000004337.

CONTENTS

LIST OF ABBREVIATIONS

BMR	Basal metabolic rate
CDR	Cause of death register
CI	Confidence interval
CNS	Central nervous system
DAG	Directed acyclic graph
FFQ	Food-frequency questionnaire
HR	Hazard ratio
ICD	International classification of diseases
MET	Metabolic equivalent
MET-hours/day	MET hours per day
NPR	National patient register
ORAC	Oxygen radical absorbance capacity
PD	Parkinson's disease
PIN	Personal identification number
PYR	Person-years
RCT	Randomized controlled trial
RR	Relative risk
SES	Socioeconomic status
SEI	Swedish socioeconomic classification index
SNpc	Substantia nigra parts compacta
STR	Swedish Twin Registry
TAC	Total antioxidant capacity

1 INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder in the world, topped by Alzheimer's disease. PD is a complex movement disorder that results in primarily motor dysfunction, but non-motor symptoms are also common. The disease was first clinically described in detail by the London doctor, James Parkinson, 200 years ago in his landmark "*An Essay on the Shaking Palsy"*.[1] In this essay, Parkinson defined the *shaking palsy* as a combination of rest tremor, lessened muscular power, abnormal truncal posture, and festinant and propulsive gait.[2] These features are still used today to establish a diagnosis of PD clinically. In particular, he documented six illustrative cases of PD, though these observed cases were actually a mixture of different disorders according to current knowledge of neurology.

Undoubtedly, Parkinson's monograph encourages other doctors to advance the knowledge of the disease, yet both the essay and Parkinson himself received little immediate attention in British medical circles at the time.[3] About 60 years later, a French neurologist Jean Martin Charcot distinguished slowness of movement from weakness and added other clinical manifestations to the description of the disease, including non-tremulous form and facial motor involvement. In fact, Charcot was the first who acknowledged the significance of Parkinson's essay and named the disease after him.[4]

At the time of Parkinson's essay, there was no effective treatment for the disease. There is still no cure today. It is also unclear when people start developing PD since mechanisms of the disease are unknown. However, the interest of the scientific community in PD is constantly growing, triggered by recent discoveries from genetic studies.[5] Indeed, genetic findings may pinpoint potential pathogenesis and etiology of the disease,[6] but they only explain a small fraction of the disease since over 90% of PD cases are sporadic with no identifiable genetic cause.[7]

Observational findings have provided compelling evidence on behavioral and environmental factors in relation to PD,[8] with the most robust finding being an inverse association between smoking and risk of PD. As a result, other lifestyle aspects such as physical activity and diet have been suggested to play an important role to further the understanding of the pathogenesis and progression of PD. Large, well-designed and population-based

epidemiological prospective cohort studies are probably the best ways to investigate potential risk factors for PD, as well as the outcomes that develop during a long period of the disease.[9]

In this thesis, I provide an update on the epidemiological evidence of PD along with four epidemiological studies that I conducted during the period of my PhD studies, focusing on four specific lifestyle factors. PD refers to idiopathic PD unless specified otherwise. Epidemiological studies regarding potential genetic causes of PD or parkinsonism/Parkinsonian disorders other than PD are briefly mentioned but fall beyond the scope of this thesis.

2 BACKGROUND

PD involves both the dopamine system and Lewy pathology. In the central nervous system (CNS), dopamine functions as a neurotransmitter sending signals from one neuron to another. In the brain, the dopamine system includes several pathways that play important roles in neuronal activities, for example the nigrostriatal pathway that is responsible for motor control. More specifically, patients with PD have substantial loss of dopaminergic neurons (i.e. dopamine-secreting neurons) in the substantia nigra parts compacta (SNpc) located in the mesencephalon (i.e. midbrain of the brainstem). These dopaminergic neurons project to the striatum of the basal ganglia, which interconnect with other brain areas including the motor cortex (frontal lobe). The cell death of dopaminergic neurons leads to depletion of dopamine in the striatum, resulting in motor dysfunction.

In addition to dysfunction of the dopaminergic system, PD is associated with presence of ubiquinated misfolded proteins (i.e. α-synuclein) in the cytoplasm of neurons.[10] In the misfolded situation, the soluble protein α-synuclein aggregates abnormally and becomes insoluble to form intracellularly spherical inclusions, called Lewy bodies and Lewy neurites.[11] Lewy bodies are found extensively in dopaminergic neuros of SNpc in patients with PD, but also in other parts of the CNS. Based upon this observation of Lewy bodies in PD patients, Braak and colleagues proposed a hypothesis that the Lewy pathology starts from peripheral nervous system and progresses to the CNS in a caudal-to-rostral direction.[12] The Lewy pathology in PD includes six stages (Table 1.1). Although it is still controversial, the Braak staging seems to explain largely the clinical course of PD.

Table 1.1 Braak staging of Lewy pathology in PD (adapted from Braak et al.[12])

Stage 1 • Lesions found in the peripheral nervous system (autonomic neurons), olfactory system, and medulla (dorsal motor nuclei of vagal and glossopharyngeal nerves)
Stage 2 • Caudal raphe nuclei, gigantocellular reticular nucleus, and coeruleus–subcoeruleus complex
Stage 3 • Midbrain (substantia nigra pars compacta)
Stage 4 • Anteromedial temporal mesocortex, thalamus, and limbic system (accessory cortical and central nuclei of amygdala, and ventral claustrum)
Stage 5 • High order sensory association areas and prefrontal areas (anterior cortex)
Stage 6 • First order sensory association and premotor areas, as well as primary sensory and motor areas.

2.1 Symptoms and Diagnosis of Parkinson's Disease

2.1.1 Symptoms of Parkinson's disease

PD is the most common (about 80%) neurodegenerative cause of parkinsonism, which is a clinical syndrome and includes other parkinsonian disorders such as cerebrovascular parkinsonism and multiple system atrophy, etc. The symptoms of PD are characterized by three cardinal motor dysfunctions:

- *Tremor (shaking)*
  An involuntary shaking movement in a part of the body. In comparison to an "*essential tremor/action tremor*" that happens during the action of movement, patients with PD are more likely to have a "*resting tremor*", which happens when the body is relaxed, such as sitting down with hands resting on the lap.

- *Bradykinesia (Slowness of movement)*
  Slowness of movement during, for example, walking or using hands and arms. Patients with PD may find it takes longer to do regular things, such as turning around, getting out of chairs, or writing.

- *Rigidity (stiffness or inflexible muscles)*
  Stiffness of muscles and joints of the body.

These motor symptoms start slowly and develop gradually with no particular order. They usually begin in one hand and spread to limbs on the same side, then to the other side of the body. Notably, resting tremor can be the first noticeable sign of PD. However, tremor may be caused by other conditions, and not all PD patients develop the resting tremor.

In addition, modern symptomatology of PD considers PD heterogeneous meaning that the disease pathology involves multiple regions of the nervous system, several neurotransmitters, and different protein aggregates besides the Lewy bodies. Thus, non-motor symptoms are also of clinical significance to patients with PD. Moreover, non-motor features have been suggested as early clinical manifestations of PD for up to 20 years prior to the occurrence of cardinal motor symptoms (Figure 1.1). Such non-motor features include olfactory dysfunction, cognitive impairment, psychiatric symptoms, sleep disorders, fatigue, and others.

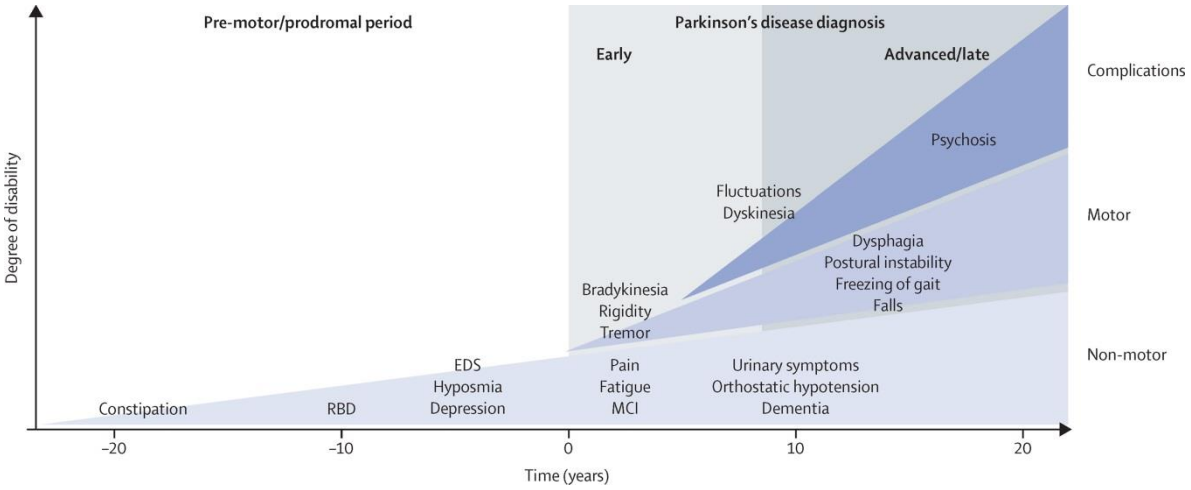

Figure 1.1 Clinical symptoms and time course of Parkinson's disease progression (Reprinted from Kalia LV & Lang AE,[6] with permission from Elsevier)

A premotor or prodromal phase of PD (characterized by specific non-motor symptoms) can be 20 years or more prior to PD diagnosis made at the onset of motor symptoms (time 0 years). EDS = excessive daytime sleepiness; MCI = mild cognitive impairment; RBD = REM sleep behavior disorder.

2.1.2 Diagnosis of Parkinson's disease

There is no biological marker for the diagnosis of PD. At present, a clinical diagnosis of PD relies on motor symptoms and definite PD depends on post-mortem neuropathological examination. Several clinical diagnostic criteria for PD have been proposed.[13-17] The most commonly used criteria by physicians are the *UK Parkinson's disease Society Brain Bank Clinical Diagnostic Criteria* (UK-Criteria, Table 1.2).

The UK-Criteria were suggested more than 20 years ago. As the understanding of PD advances, the UK-Criteria are thought to be insufficient for the clinical research requirements nowadays.[18] Therefore, the Movement Disorder Society refined the UK-Criteria and presented the *Movement Disorder Society Clinical Diagnostic Criteria for Parkinson's Disease* (MDS-PD Criteria) in 2015 to replace the UK-Criteria for use in clinical research and potentially in the clinical setting.[19] The MDS-PD Criteria reinforce the importance of early non-motor manifestations compared to the UK-Criteria,[20] yet they have not been validated for accuracy against the definite diagnosis of PD. Nevertheless, MDS-PD criteria reported higher sensitivity than the UK-Criteria.[21]

Table 1.2 UK Parkinson's disease Society Brain Bank Clinical Diagnostic Criteria (adapted from Hughes et al.[17])

Step 1: Inclusion criteria for possible PD

- Bradykinesia (Slowness of initiation of voluntary movement with progressive reduction in speed and amplitude of repetivtive actions)
- And at least one of the following:
  Muscular rigidity
  4-6 Hz rest tremor
  Postural instability not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction

Step 2: Exclusion criteria for PD

- History of repeated strokes with stepwise progression of parkinsonian features
- History of repeated head injury
- History of definite encephalitis
- Oculogyric crises
- Neuroleptic treatment at onset of symptoms
- More than one affected relative
- Sustained remission
- Strictly unilateral features after 3 years
- Supranuclear gaze palsy
- Cerebellar signs
- Early severe autonomic involvement
- Early severe dementia with disturbances of memory, language, and praxis
- Babinski sign
- Presence of cerebral tumour or communicating hydrocephalus on CT scan
- Negative response to large doses of levodopa (if malabsorption excluded)
- MPTP exposure

Step 3: Supportive prospective positive criteria for diagnosis of definite PD

(Three or more required**)**

- Unilateral onset
- Rest tremor present
- Progressive disorder
- Persistent asymmetry affecting side of onset most
- Excellent response (70–100%) to levodopa
- Severe levodopa-induced chorea
- Levodopa response for 5 years or more
- Clinical course of 10 years or more

PD = Parkinson's disease; CT = computed tomography; MPTP = 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.

2.2 Prevalence, Incidence and Mortality of Parkinson's Disease

The overall prevalence of PD was estimated at 0.3% for all ages and at 1% for people over 60 years of age in high-income countries.[9] A slightly increased prevalence of PD was found in men than in women.[22] Importantly, age is the greatest contributor to the development of PD, since the disease occurs rarely before age 50 and the prevalence can reach up to 1,903 per 100,000 in individuals over the age of 80.[22] Compared to developed countries, prevalence of PD seems to be lower in developing countries, where prevalence for all ages was estimated at 10-43 per 100,000, 15-119 per 100,000, and 27-43 per 100,000 in Asian, African, and Arabic countries, respectively.[6]

Prevalence of a disease is prone to bias such as differences in survival among various age groups, whereas incidence rate is theoretically not affected by survival and therefore a better measure of the risk of disease.[23] Overall, incidence rates of PD estimated for all ages range from 1.5 to 22 per 100,000 person-years (PYR).[24] Similar to the prevalence, the incidence of PD increases rapidly with age, especially after 60 years of age and it peaks at around 80 years of age.[25] Age-standardized incidence of PD was estimated at between 16 and 19 per 100,000 PYR.[26] In high-income countries such as the United States, median age-standardized incidence of PD was estimated at 14/100,000 PYR in the total population and up to 160/100,100 PYR for individuals over the age of 65.[27] Men seemed to have a higher risk of PD than women, with the male-to-female incidence ratio ranging from around 1.3 to 2.0.[28,29] Similarly, cumulative incidence of PD adjusted for competing risk of death (i.e. lifetime risk of PD), was higher in men (2%) than in women (1.3%).[30]

Patients with PD have a shorter life expectancy than healthy individuals. Most epidemiological studies suggested mortality hazard ratios (HR) at about 2 (ranging between 1.3 and 5.7) for patients with PD compared to the general population.[24]

2.3 Risk Factors for Parkinson's Disease

The exact cause of PD is largely unknown, but risk factors proposed for the development of PD include both genetic and non-genetic/environmental factors.

2.3.1 Genetic factors

A genetic component in PD was suggested by discoveries of monogenic forms of PD (Table 1.3).[31-34] In addition, a heritability estimate of 34% was suggested in a previous twin study[35] but others reported low concordance rates in monozygotic and dizygotic twins,[36-40] suggesting the importance of non-genetic or environmental factors in PD etiology. A recent meta-analysis reported a higher risk of PD [relative risk (RR): 2.9, 95% confidence interval (CI): 2.2-3.8] for having a first-degree relative with PD.[41]

Although highly penetrant monogenic mutations have been identified, these findings were primarily based on family data, whereas most PD patients (90%) are sporadic cases.[7] Thus, environmental factors may play an important role in the etiology of sporadic PD.

Table 1.3 Monogenic forms of Parkinson's disease, by gene (Adapted from Kalia LV & Lang AE,[6] with permission from Elsevier)

	Protein	**Pathogenic mutation(s)**
Autosomal dominant		
SNCA	α-synuclein	Missense mutations (Ala18Thr, Ala29Ser, Ala30Pro, Glu46Lys, His50Gln, Gly51Asp, Ala53Glu, Ala53Thr); multiplications (duplications, triplications)
LRRK2	Leucine-rich repeat kinase 2	Missense mutations (Ile1371Val, Asn1437His, Arg1441Cys, Arg1441Gly, Arg1441His, Tyr1699Cys, Gly2019Ser [most common], Ile2020Thr)
VPS35	Vacuolar protein sorting 35	Missense mutation (Asp620Asn)
EIF4G1	Eukaryotic translation initiation factor 4-γ 1	Missense mutations (Arg1205His, Ala502Val)
DNAJC13	Receptor-mediated endocytosis 8 (REM-8)	Missense mutation (Asn855Ser)
CHCHD2	Coiled-coil-helix-coiled-coil-helix domain containing 2	Missense mutations (Thr61Ile, Arg145Gln); splice-site alteration

(Table 3 Continued)

Autosomal recessive		
Parkin	Parkin	Exon rearrangements, including exon deletions or multiplications (most common); missense mutations, nonsense mutations, small deletions or insertions; splice-site alterations
PINK1	PTEN-induced putative kinase 1	Missense or nonsense mutations (most common); exon rearrangements, including exon deletions or duplications
DJ-1	DJ-1	Missense mutations or exon rearrangements (most common); splice-site alterations

2.3.2 Non-genetic/Environmental factors

Compared with genetic studies, associations between non-genetic/environmental factors and risk of PD were consistently observed in epidemiological research.[42] Four specific lifestyle factors primarily investigated in this thesis are introduced below.

2.3.2.1 Physical activity

Physical activity is defined as "*any bodily movement produced by skeletal muscles resulting in energy expenditure*".[43] It is a multidimensional behavior that includes different forms of activities of the body as long as it results in energy expenditure. Physical exercise is a type of physical activity that is planned, structured and repetitive for the purpose of improvement or maintenance of the body.[43]

Total energy expenditure of the human body can be divided into three parts: basal metabolic rate (BMR), diet induced thermogenesis, and physical activity. BMR, the rate at which the body metabolizes energy in order to stay alive, is a function of both body size and composition, and is affected mainly by age, gender, weight and height.[44] BMR accounts for around 60-75% of total energy expenditure. Diet induced thermogenesis is the energy required to digest food and absorb nutrients, accounting for about 10% of total energy expenditure. These two parts are rather stable compared to physical activity. Physical activity is the most modifiable part of the total energy expenditure.

Measurement of physical activity

Energy expenditure from physical activity varies by frequency, duration and intensity of the activity. Therefore, assessment of physical activity can be diverse and difficult. Metabolic equivalent (MET) value measures energy expenditure and represents the intensity of a specific physical activity.[45,46] More precisely, one MET (e.g. sitting quietly or lying down) represents the ratio of metabolic rate as a standard BMR of a healthy adult with a normal weight, which equals to an oxygen consumption of 3.5 ml O_2 * kg^{-1} body weight * min^{-1} or an energy expenditure of 1 kcal * kg^{-1} body weight * $hour^{-1}$. A higher MET-value means that the more intense the activity is, the more energy is expended during performance of the activity. For example, the energy expenditure of performing a physical activity with an intensity of 7 MET (e.g. running) equals to 7 * 1 kcal * kg^{-1} body weight * $hour^{-1}$. Thus, total energy expenditure for a specific physical activity can be calculated given an individual's body weight and duration of the performed activity.[47]

Associations between physical activity and PD

An inverse association between physical activity and risk of PD was first reported by Sasco and colleagues in 1992.[48] They reported a lower risk of PD for men who played sports in college and adulthood. However, no association between regular exercise or competitive sports and risk of PD was found in two later case-control studies.[49,50] A recent population-based case-control study, on the contrary, showed that individuals engaged in overall physical activity and leisure time competitive sports at high levels had lower risk of PD.[51]

The few prospective cohort studies investigating the effect of physical activity on risk of PD have yielded inconsistent findings.[52-56] Recreational physical activity (e.g. climbing stairs, walking or jogging) was not associated with PD in the Harvard Alumni Health Study.[52] However, moderate to vigorous level of recreational physical activity was associated with lower risk of PD in the Cancer Prevention Study II Nutrition cohort[53] and NIH-AARP Diet and Health Study cohort.[55] Similarly, higher levels of physical exercise was associated with lower risk of PD in the Health Professionals' Follow-up Study and Nurses' Health Study cohorts, although the association was only significant for men but not for women.[54] Lastly, a recent study showed that heavier leisure-time physical activity was related to a stronger reduction of PD risk.[56]

Gender differences regarding the effect of physical activity on PD are plausible, but previous studies mainly assessed physical exercise in PD and did not consider possible differential effects between men and women. In addition, it is particularly important to study the effect of total physical activity (i.e. all activities together) instead of physical exercise on PD risk, as the total physical activity may be more realistically achievable for many.

2.3.2.2 Tobacco

Use of tobacco products, or more specifically, smoking of cigarettes, is one of the most studied lifestyle factors related to PD. An inverse association between cigarette smoking and PD was first observed more than 50 years ago, with a standardized morality ratio of 0.36 in regular cigarette smokers compared to non-smokers.[57] Subsequently, more than 40 retrospective case-control studies have been conducted to investigate the association between cigarette smoking and risk of PD. Their results were summarized in a meta-analysis, showing a pooled relative risk of PD of 0.59 (95% CI 0.55-0.65) for ever smokers, 0.88 (95% CI 0.75-1.02) for past smokers, and 0.41 (95% CI 0.32-0.51) for current smokers, compared to never smokers.[58]

However, the inverse association between smoking and PD has been criticized regarding potential biases related to the study design nature of case-control studies.[59,60] These biases include recall bias, selection of inappropriate controls or selective mortality of smokers among PD-free individuals (selection bias), inaccurate recording of PD diagnoses in smokers (information bias), and that preclinical PD patients may be less prone to smoke or more likely to quit (reverse causation).[61] Notably, results from prospective studies including nested case-control[62-64] and large cohort studies[61,65-69] confirmed the strong inverse association between cigarette smoking and risk of PD, in which potential biases such as selective mortality have been addressed.[70] More importantly, a dose-dependent risk reduction of PD was suggested, with more intense and longer duration of smoking related to reduced PD risk.[61,66,69,71]

Among numerous chemicals responsible for the potential protective effect, nicotine is considered the most plausible candidate, primarily because it stimulates dopaminergic systems and provides a neuroprotective effect in experimental models of PD.[72,73] Accordingly, research on different tobacco products, in particular nicotine concentrated products without combustion of tobacco such as chewing tobacco or snuff, may better explore what constituent of tobacco smoke is responsible for the reduced risk of PD. Yet only

two studies with small sample sizes are available, both of which have uncovered a strong inverse relationship between smokeless tobacco use and PD.[74,75]

In Sweden, there is a long history of producing and using a wet snuff (i.e. an oral moist, finely ground and smokeless tobacco product), "snus" in Swedish (Figure 1.2).[76] Production of snus once reached to 7000 tons (1.2 kilograms/capita) in 1919, but was almost diminished during the 1950s and 60s because of high prevalence of cigarette smoking among Swedish men.[77] Nevertheless, snus use has increased since the mid-1970s as the prevalence of smokers has decreased, and in 2015 about 19% of men and 4% of women were daily snus users in Sweden.[78,79]

Figure 1.2 Illustrations of snus sold in the market in Sweden (private picture from the author)

The health consequences of using snus have not been studied in detail. Besides a few studies that have reported harmful effects of snus use on cardiovascular diseases and cancers,[80,81] most epidemiological evidence have shown that snus use does not lead to increased risk of cancers, cardiovascular diseases, or mortality.[82-85] Many have considered snus an example of a "Swedish experience" that is perceived as a less harmful tobacco product compared to cigarettes,[86-89] although the less harmful nature of snus use has been debated among both researchers and the general public.[90-94] No one has investigated relationship between snus use and PD.

2.3.2.3 Antioxidants

Oxidative stress has long been hypothesized to be involved in the pathogenesis of PD.[95,96] Experimental and clinical evidence showed that oxidative stress can induce mitochondria damage and modulate intracellular signaling,[97] leading to production of cytotoxic free radicals and dopaminergic neuron death by necrosis and apoptosis.[98] Therefore, antioxidants have been suggested as neuroprotective agents in PD based on their property of reducing oxidative damage, presumably by suppressing radical generation.[99-101]

Conversely, evidence from experiments and observational studies on the relationship between antioxidant vitamins and PD is inconclusive.[102] Among previous epidemiological studies conducted on dietary antioxidant vitamins and risk of PD, most have been case-control studies,[103-112] while only a few studies were prospective.[63,67,113,114] A lower risk of PD was found for higher total intake of vitamin E in the Singapore Chinese Health Study, but not for total intake of vitamin C or A.[67] In the Health Professionals' Follow-up Study and Nurses' Health Study cohorts, no association was found between risk of PD and total intake of antioxidant vitamins, including vitamin C and E, and carotenoids.[114] Interestingly, an inverse association between vitamin E intake and risk of PD was found when considering intake of vitamin E from foods only.[114] Vitamin B6, B16, and folate were not related to PD in two cohort studies,[115] but vitamin B6 was associated with a decreased risk of PD in the Rotterdam cohort study.[116] Moreover, a recent large prospective study revealed an inverse association between vitamin C and PD risk.[117] In summary, previous evidence suggests that vitamin E and possibly vitamin C and beta-carotene may lower the risk of PD.[118,119]

Given the plausible neuroprotective effect of antioxidants on PD and that the major source of antioxidants is various food items, it is important to investigate the overall effect of dietary antioxidants in relation to PD risk. Total antioxidant capacity (TAC), a single measure of the free-radical-reducing capacity from all antioxidants present in foods, is probably the best measure in this context.

TAC of diet can be estimated using an oxygen radical absorbance capacity (ORAC) assay [µmol Trolox equivalents per 100 g] that takes into account the synergistic effects of thousands of compounds in foods.[120-122] The use of TAC has been assessed in relation to risk of myocardial infarction,[123] cataract,[124] and other medical conditions,[125-127] but no one has investigated the relationship between TAC and risk of PD.

2.3.2.4 Socioeconomic status

Socioeconomic status (SES) has been studied in relation to many health consequences, including both morbidity and mortality. Defining SES, however, is challenging. The main issue remains in the measurement of SES, which can be broadly covered by both social class and social indicators.[128] These two aspects somehow define the SES as 'socioeconomic position', a combined measure of social status and economic situation.

According to well-known sociologists, Karl Marx and Max Webber, the society can be stratified into different classes, where an occupational class or a group of people can share 'life-changes' and own the means of production.[128] Therefore, social class represents 'social position (status)' that indicates a particular structure location within society.[129] Social indicators, such as educational level and income, are strongly related to social class (measured by occupation).[130] Since occupation, education and income are closely correlated with each other, they are commonly used in epidemiological research and often interchangeable. However, these socioeconomic indicators possess unique qualities and each of them can influence people's health and mortality accordingly.[131]

SES is often considered as a measure of family income or highest education level in studies from North America,[132,133] whereas SES is often characterized as individual occupational class in European studies.[134,135] In Sweden, sociologist Robert Erikson and colleagues constructed the Swedish Socioeconomic Classification Index (SEI) in the 1970s, by classifying occupations into different social class categories.[136,137] The SEI distinguishes between employers and employees (see appendix 1). According to the SEI, employees are divided into manual and non-manual workers, and classified further into high- and low-level sub-categories based on educational level required for the occupation.[137]

Notably, the SEI classification is coded differently in the Swedish censuses from 1960 to 1990. For example, SEI in the 1980 census has 17 different social and occupational classes, whereas in the 1990 census 12 classifications are available together with 3 additional categories that cannot be assigned to a social class (see appendix 2).[138] The latter 3 categories are referred to as "unclassifiable employed", "missing SEI information", and "not in the labor force" such as students and pensioners.

Social epidemiology is the "*study of relations between social factors and disease in populations*".[139] A few studies have examined the association between SES and PD but most of them used a single social indicator such as education or occupation.[132,140] Limited numbers of studies on a higher hierarchical measure of SES and PD are, in part, due to the difficulties of measuring SES and explaining the plausible mechanisms behind the observed association. However, it has been shown that SEI is a good indicator for the combined measure of SES and individuals above age of 30 years have a stable occupation during 20th century in Sweden.[141-143] Moreover, many Swedish studies have investigated associations between SES (classified according to SEI) and medical conditions, including neurodegenerative diseases.[144]

Therefore, SES classified according to the SEI, can be of interest in epidemiological research both for descriptive purposes and for formulating new hypotheses about the role of environmental and social factors in PD etiology. Such research may also help develop equitable healthcare service available to PD patients.

3 AIMS

The overall aim of this thesis is to understand better the relationship between several lifestyle factors and subsequent risk of PD.

In particular, in different studies, we aim:

Study I: to study whether physical activity is associated with risk of PD;

Study II: to examine the association between smokeless tobacco (Snus) use and risk of PD;

Study III: to investigate the effect of dietary antioxidants on PD risk; and

Study IV: to explore the association between SES and PD risk, as well as the effect of SES on all-cause mortality in PD patients and PD-free individuals.

4 MATERIAL AND METHODS

4.1 Swedish Population Registers

Sweden has more than 300 years of history on population statistics that originated from parish registrars. Later in the 19th century, Statistics Sweden (Swedish: Statistiska Centralbyrån) and National Board of Health and Welfare (Swedish: Socialstyrelsen) were established as governmental agencies for the purposes of registering population statistics and health records.[145,146]

In 1947, personal identification number (PIN) was introduced to the Swedish society and has become a crucial part of Swedish healthcare administration since then. The PIN is assigned to every person who is born in Sweden, as well as to immigrants who intend to stay in Sweden for more than one year. This number is unique to everyone and enables record linkage with all administrative and health registers in Sweden.[147]

4.1.1 National Patient Register

The national patient register (NPR) started in 1964 and became nationwide in 1987. The NPR covers all hospital discharge diagnoses in Sweden.[148] Since 2001, the NPR also includes outpatient specialist visits at clinics. Diagnoses given in primary healthcare settings are not covered by the NPR. Diagnoses in the NPR are coded by the Swedish revisions of the *International Classification of Diseases* (ICD). In addition, dates of diagnosis, admission to, and discharge from hospital are also recorded in the NPR. The NPR is managed by the National Board of Health and Welfare.

4.1.2 Cause of Death Register

In Sweden, death records dates back to the 18th century. This system gradually formed the cause of death register (CDR) and reached nationwide coverage in 1961. The CDR was managed by Statistics Sweden in the early 20th century but migrated to the National Board of Health and Welfare in 1994.[149] The CDR contains data on date of death as well as main and secondary diagnoses of death for all individuals registered in Sweden, regardless of whether

the death occurs inside or outside the country. Causes of death in the CDR are also coded according to the Swedish revisions of the ICD.

4.1.3 Swedish Twin Registry

The Swedish Twin Registry (STR) contains more than 170,000 twins born in Sweden from 1886 and onwards.[150,151] In addition to the PIN that everyone has in the Swedish administrative registers, twins in the STR have a unique twin identification number enabling record linkage with all sub-cohorts of the register. Zygosity information on same-sexed twins are assessed by questions in the questionnaires and easily identified through twin identification number. The questions used for determining zygosity have been validated with an accuracy of 99% compared with genotyping.[152]

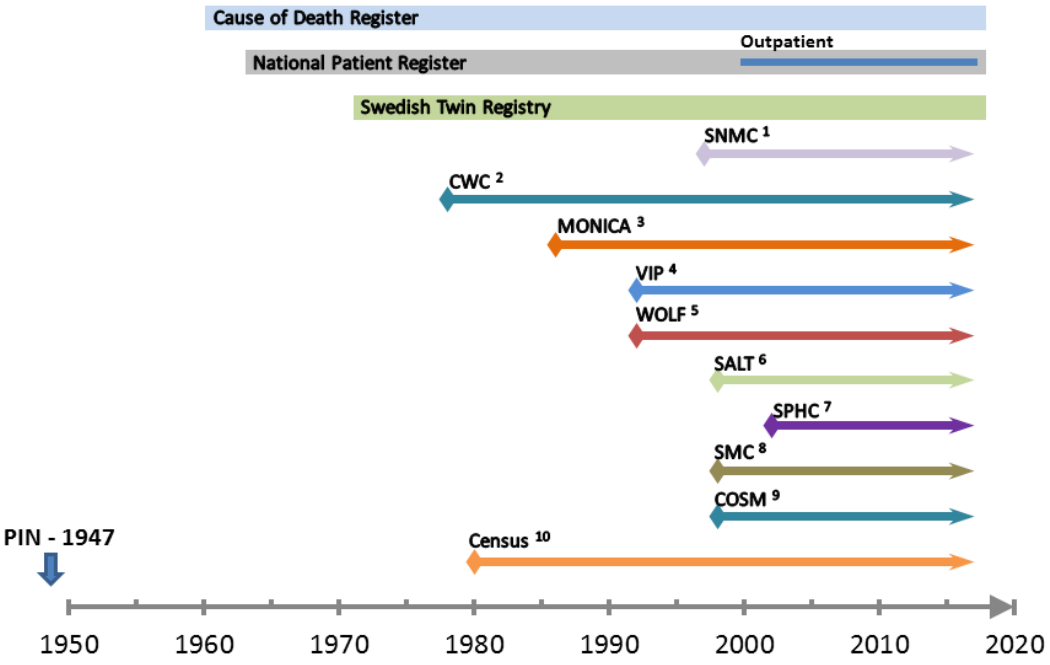

Figure 4.1 Swedish register data and cohort data used in this thesis

1. SNMC = Swedish National March Cohort; 2. CWC = Construction Workers Cohort; 3. MONICA = Multinational Monitoring of Trends and Determinants in Cardiovascular Disease; 4. VIP = Västerbotten Intervention Program; 5. WOLF = Work, Lipids and Fibrinogen Study; 6. SALT, Screening Across the Lifespan Twin Study; 7. SPHC = Stockholm Public Health Cohort; 8. SMC = Swedish Mammography Cohort; 9. COSM = Cohort of Swedish Men; 10. Census data; PIN = Personal Identification Number

4.2 Types of Study Design

4.2.1 Swedish National March Cohort (Study I)

4.2.1.1 Study participants

The Swedish National March Cohort (SNMC) was assembled during a 4-day promotional and fund-raising event for the Swedish Cancer Society in almost 3600 cities and villages in Sweden in September 1997. During the event, all participants were invited to fill out a 36-page questionnaire with background data on lifestyle and medical history.

The SNMC is a general population cohort, in which participants provided their PIN, enabling identification of health status and possible follow-up monitoring through record linkage to all nationwide administrative registers in Sweden.[153]

Due to the specific characteristic of this event, the total number of participants who received a questionnaire was unknown. In total 43,880 participants completed the questionnaire in the SNMC. After exclusion of incorrect PIN, 43,852 participants were included in the cohort. We further excluded participants who had a record of PD diagnosis (n=10), death (n=8), or emigrated out of Sweden (n=466) before start of follow-up. Therefore, the final sample used in **study I** included 43,368 participants, consisting of 15,505 (35.7%) men and 27,863 (64.3%) women. Study participants were followed from baseline on October 1, 1997 until date of PD, death, emigration, or end of follow-up on December 31, 2010, whichever came first (Figure 4.2).

4.2.1.2 Assessment of physical activity

To investigate the association between physical activity and risk of PD, we analyzed most types of physical activity assessed in the questionnaire of SNMC in **study I**. These activities include household and commuting activity, leisure time exercise, occupational activity, and total daily physical activity.

Regarding household and commuting physical activity, participants reported how many hours per week (i.e. < 1, 1-2, 3-4, 5-6, or > 6 hours), on average, they spent on these activities such as cleaning the house, working in the garden, or walking/biking to work.

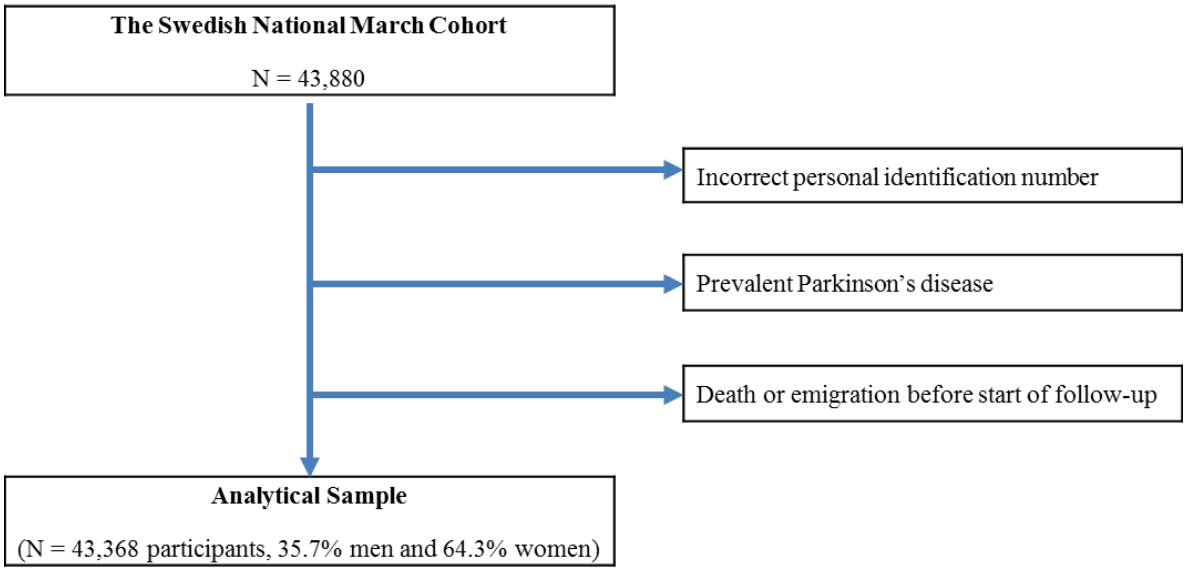

Figure 4.2 Flowchart describing the analytical sample of the Swedish National March Cohort

Regarding occupational physical activity, participants reported how physically demanding their daily job was during the last year (i.e. mostly sedentary, light but moving a little, rather strenuous, and very strenuous).

Regarding leisure time exercise, participants reported how many hours per week (i.e. never, 0-1, 2, 3, 4, or ≥ 5 hours), on average, they spent on exercise or outdoor activities. Leisure time exercise was assessed at three levels: light (e.g. casual walking), moderate (e.g. speedy walking, jogging or swimming), and heavy (e.g. vigorous training or competitive sporting) and was measured separately in summer and winter.

With regard to total daily physical activity, we assessed the activity level represented for a typical 24-hour day using a validated instrument (the Energy Expenditure Questionnaire, EEQ) with nine fixed scale steps.[154] Each scale step was exemplified by common activities. The nine scale steps together represented different estimates of activity in terms of MET-value, and were obtained from questions that were typically asked in epidemiological research. To calculate total physical activity level, hours engaged in each scale step were summed using

$$\sum_{i=1}^{9} t_i \times MET_i \quad (1)$$

As a result, mean intensity of total physical activity in a 24-hour day could be estimated for all activities in MET hours per day (MET-hours/day). Assessments for other physical

activities in the SNMC such as sedentary behavior have been described in detail previously but were not studied in the present study.[153]

Variables of physical activity were further processed before statistical analyses in **study I**. First, numbers of hours spent on household and commuting activity and leisure time exercise were quantified into MET-hours/day. The conversion was based on estimated oxygen consumption associated with these activities (e.g. 4.0 MET for household and commuting activity; 3.0, 6.0 and 10.0 METs for light, moderate and heavy leisure time exercise). In addition, we combined household and commuting activity and leisure time exercise into a variable termed "general physical activity". Scores of MET-hours/day of physical activity were then categorized into tertiles with separate cut-offs for all participants, men and women.

4.2.2 Swedish Collaboration on Health Effect of Snus Use (Study II)

4.2.2.1 Study participants

The Swedish Collaboration on Health Effects of Snus Use is a national consortium aimed to study the effect of snus use on human health. The collaboration pools almost all prospective cohort studies with information on the use of Swedish snus and tobacco smoking in the country. Study designs and measures for individual cohorts included in the collaboration were described in detail elsewhere.[81,85,150,153,155-158] At present, eight prospective cohort studies are included in the snus collaboration, of which individual participant data from seven cohorts were collected for **study II**.

Among individuals in the seven cohorts, only men (n=351,640) were included in **study II** due to extremely low prevalence of snus use in women at the time of study enrolment (Table 4.1). In addition, participants were excluded from analyses if they had a record of PD diagnosis (n=148) or if they were younger than 18 years (n=2891) before baseline, resulting in 348,601 participants. All participants in each cohort were followed from baseline until date of first-ever PD diagnosis, death, or end of follow-up, whichever came first.

Table 4.1 Cohorts included in the SNUS collaboration (men only)

	Number of participants (n)	Incident PD cases (n)	Never-tobacco smoker (%)	Ever-snus user (%)	Age at recruitment (mean±SD)	Follow-up period (mean±SD)
CWC	214,381	474	45.8	29.8	34.2±12.5	20.0±5.1
SPHC	39,212	89	55.3	27.8	50.2±16.9	4.8±3.1
MONICA	4,553	31	43.4	38.6	49.0±13.4	12.5±6.3
SNMC	15,012	165	50.2	23.1	53.4±17.1	12.6±2.5
SALT	18,316	117	45.5	23.9	56.6±8.0	10.0±1.6
VIP	49,940	291	48.2	42.2	47.2±9.2	12.4±6.3
WOLF	7,187	32	49.3	34.3	42.9±10.8	13.9±2.1
Scania	6,198	20	46.6	N/A	48.6±16.7	9.0±0.4

Scania was excluded from the study II because information on snus use status was only available for current or non-current snus use. SD = Standard deviation; PD = Parkinson's disease; CWC = Construction Workers Cohort; SPHC = Stockholm Public Health Cohort; MONICA = Multinational Monitoring of Trends and Determinants in Cardiovascular Disease; SNMC = Swedish National March Cohort; SALT = Screening Across the Lifespan Twin Study; VIP = Västerbotten Intervention Program; WOLF = Work, Lipids and Fibrinogen Study; Scania = Scania Public Health Cohort.

4.2.2.2 SALT and Q-73

In addition to twins from the Screening Across the Lifespan Twin Study (SALT) included in the snus collaboration, a nested case-control study design was further employed within **study II,** including additional twins born 1926-1958 who participated in a questionnaire sent in 1972-1973 (Q-73 cohort).[152,159] Purposes of this nested case-control study were to examine: 1) whether possible misclassifications or changes of exposure during the follow-up attenuated the association, and 2) whether there was a dose-response relationship between years of lifetime snus use and risk of PD.

4.2.2.3 Assessment of snus use

Snus use was assessed at baseline through a questionnaire in six cohorts and structured telephone interview in one cohort in **study II**. Information on tobacco smoking during lifetime was also collected. In addition, participants who were current users of snus at baseline provided their average amount (cans per week) and duration (years) for the snus use.

To investigate the effect of snus use on PD risk in **study II**, all participants in the seven cohorts were classified as "never user", "former user", or "current user" for their lifetime

regular use of snus and for any type of smoked tobacco products (e.g. cigarettes, pipes and cigars). In addition, amount of snus use for those who were current users at baseline was categorized as "light" (i.e. less than 2 cans per week) or "moderate-heavy" level (i.e. equal to or more than 2 cans per week), except for SNMC in which the "light level" equals to "less than or equal to 2 cans per week". Regarding duration of snus use for current users at baseline, we categorized the duration into "1-20 years" or "more than 20 years". Further, to examine the combined effect of snus use and tobacco smoking on future risk of PD, participants were cross-classified into different groups based on their answers on both snus use and tobacco smoking at baseline. Subsequently, those who "never used snus and never smoked tobacco products" were treated as the reference group.

4.2.3 Swedish Mammography Cohort and Cohort of Swedish Men (Study III)

4.2.3.1 Study participants

The Swedish Mammography Cohort (SMC) was first assembled in 1987 for the purposes of investigating the relationship of dietary and hormonal factors with risk of breast cancer. All women born 1914-1948 and living in Västmanland and Uppsala counties in central Sweden were invited to participate in the study. In 1997, those who were still alive and living in the area were contacted again to fill-out a 350-item questionnaire with detailed information on lifestyle and other risk factors for non-communicable diseases (70% response rate). In total 38,984 women were eligible for **study III**.

The Cohort of Swedish Men (COSM) was assembled in 1997. All men born 1918-1952 and living in Västmanland and Örebro counties in central Sweden were invited to answer a questionnaire identical to the one used in SMC, except for some sex-specific questions (49% response rate). In total 45,906 men were eligible for **study III**.

Participants in the SMC and COSM were excluded from analyses if they had a record of PD diagnosis in the NPR at study enrollment (n=116), resulting in 38,937 women and 45,837 men included in the **study III**. They were followed from baseline (September 15, 1997 for SMC and January 1, 1998 for COSM) until date of first-ever PD diagnosis, death, or end of follow-up on December 31, 2014, whichever came first.

4.2.3.2 Assessment of Dietary antioxidants

The questionnaires sent to participants in the SMC and COSM in 1997 contained a 96-item food-frequency questionnaire (FFQ) tailored to the Swedish diet. The FFQ assessed participants' dietary habits including exact consumption for each type of food component (per day or per week) during the past year.

To investigate associations between dietary antioxidants and risk of PD, intake of dietary antioxidants (including vitamins C and E, and beta-carotene) was calculated using food composition values from the Swedish Food Administration Database.[160]. An example of the calculation is illustrated below

$$\text{Vitamin C}_{\text{food item-1}} = \text{mean frenquency} \times \text{Vit C content in the database} \quad (2)$$

$$\text{Vitamin C}_{\text{total}} = \sum_{i=1}^{n} \text{Vitamin C}_{\text{food item}-n} \quad (3)$$

in which the intake of vitamin C is in unit of mg/day and values used for nutrient content for each food item are age- and sex-specific portion size. Similarly, dietary TAC was calculated using a formula

$$\text{TAC} = \sum_{i=1}^{n} \text{ORAC}_{\text{food item}-n} \quad (4)$$

Several validation studies of the FFQ used in **study III** have been performed previously. Self-reported micronutrient estimates were compared against 24-h recall interviews, showing a mean correlation coefficient of 0.62.[161,162] A moderate correlation (0.31) between self-reported TAC and plasma ORAC values was reported in another validation study.[163]

4.2.4 Swedish Census (Study IV)

4.2.4.1 Study participants

Statistics Sweden was responsible for population and housing censuses in Sweden between 1960 and 1990. In each decade, census questionnaires were sent to all Swedish residents who were older than 16 years old at the time of the census. Since it was mandatory to answer the census questionnaire, the response rate was extremely high at >99%.[164] In the census questionnaire, detailed information on demographic statistics were collected, including but not limited to housing, marital status, highest level of education, occupation, and social class. Individuals participated in the 1980 census were primary included in **study IV** (n=8,318,187). Data from the census questionnaires sent in 1960 and 1970 were also extracted.

Participants younger than 30 years of age at the time of the 1980 census were excluded (n=3,301,474), since previous studies reported a stable occupation and social class for individuals who were older than 30 years old.[141-143] Individuals with incorrect or missing Swedish registration number, date of death or migration, and SES information were also excluded (n=818). In addition, participants who had a record of PD diagnosis, death, and migration before start of follow-up on January 1, 1981 were excluded, resulting in 4,630,828 (48.5% for men and 51.5% for women) in **study IV**.

To study the association between SES and risk of PD, all participants were followed from baseline until date of PD diagnosis, emigration, death, or end of follow-up on December 31, 2010, whichever came first.

To study the association between SES and all-cause mortality while accounting for PD status, all participants were followed from baseline until date of death, emigration, or end of follow-up on December 31, 2010, whichever came first. The PD diagnosis was treated as a time-varying covariate that separates an individual's follow-up time into a PD-free period and a PD period.

4.2.4.2 Assessment of socioeconomic status

In the census questionnaires, questions about occupational history were asked. All participants in the 1980 census were categorized into 8 SES groups based on their answers identified with SEI (Table 4.2). The first four groups are considered ordered from the highest to the lowest SES, whereas the remaining groups may not be ordered.

A large group of individuals (i.e. group 7 and 8, n=1,173,184) had unclear SES in the 1980 census, of whom pensioners accounted 85% of the group. SES information for these persons was therefore collected from census questionnaires sent in 1970 or in 1960, wherever available.

Table 4.2 Classification of SES according to SEI in study IV

	Socioeconomic groups	N	SEI	Examples
1	Higher non-manual workers	862,724	46, 56	
2	Lower non-manual workers	557,817	33, 36	
3	Higher manual workers	575,829	21, 22	
4	Lower manual workers	1,294,068	11, 12	
5	Self-employed workers	273,965	60, 79	Company or farm owners
6	Farmers	187,690	89	
7	Pensioners	547,811	95	
8	Unclassifiable employed workers	330,924	91, 96, 97, 98, 99	House-workers and employed individuals that could not be assigned to an occupational class

SEI codes are not consecutive. N = number of participants included in the in study IV.

4.3 Outcome Ascertainment

Risk of PD was the primary outcome in **studies I - IV**. Incident PD cases were defined as individuals with a first-ever PD diagnosis in the NPR or the CDR, except for **study II**, in which PD cases were only identified from the NPR. The ICD codes used for PD were: 350 (ICD-7, 1964-68), 342 (ICD-8, 1969-86), 332.0 (ICD-9, 1987-96), and G20 (ICD-10, 1997-current). Date of first-ever PD diagnosis was defined as date of first hospital admission or outpatient contact, or date of death for cases only identified from the CDR. All-cause mortality was another outcome in **study IV** and it was defined as any death record in the CDR.

PD diagnoses in the NPR and the CDR have been validated against clinical diagnoses and showed a positive predictive value of 70.8% for the NPR and 66.7% for the CDR.[165] The sensitivity for PD diagnosis is 72.7% in NPR and 83.1% for the two registers combined.

4.4 Assessment of Covariates

Several covariates were considered as possible confounders based on subject knowledge, literature review, and causal diagrams known as directed acyclic graphs (DAG). Relevant information on those covariates was collected based on the availability in each study. Briefly, these covariates include cigarette smoking (**study I - IV**), alcohol intake (**study I - III**), coffee intake (**study I - III**), BMI (**study I, III**), educational level (**study I - III**), physical activity level (**study II**), total energy intake (**study III**), marital status (**study IV**), and area of living (**study IV**).

4.5 Statistical Methods

4.5.1 Cox Proportional Hazards Regression (Study I - IV)

Cox proportional hazards regression model with attained age as the underlying timescale was used as the primary statistical method (i.e. age-adjusted model) to estimate HRs with 95% CIs in all studies (**study I - IV**). Analyses were conducted for all participants (i.e. two genders combined) in **study I** and **IV** and stratified by gender in **study I - IV**. In addition to the age-adjusted model, we used a multivariable-adjusted model that adjusts for other possible confounders in each study.

In the multivariable-adjusted model for analyzing physical activity in **study I**, sex, cigarette smoking, alcohol and coffee intake, BMI, and educational level were included. In the analyses of snus use in **study II**, to avoid potential confounding effect from tobacco smoking on the association between snus use and PD risk, primary analyses were restricted to *never smokers*. Other covariates included in the multivariable-adjusted model were alcohol, education and physical activity in Stockholm Public Health Cohort; alcohol and physical activity in Multinational Monitoring of Trends and Determinants in Cardiovascular Disease Cohort; alcohol and physical activity in SNMC; alcohol, education, physical activity and coffee intake in SALT; education, alcohol and coffee intake in Västerbotten Intervention Program; alcohol, education and physical activity in Work, Lipids and Fibrinogen Study. Multivariable-adjusted model used in **study III** for analyses of dietary antioxidants was similar to the one used in **study I**, with additional adjustment for total energy intake. In **study**

IV, sex, marital status, and area of living were adjusted for in the multivariable-adjusted model.

Continuous exposures (e.g. MET-hours/day for physical activity in **study I** and mg/day for vitamin C intake in **study III**) were analyzed as both continuous and categorical variables, whenever possible. Trend analyses were performed for categorical variables, using the median of each exposure category as a single ordinal variable in the modelling (**study I - III**). The lowest category was used as the reference group in all analyses for categorical variables, except for **study IV** where participants with the highest SES (i.e. *higher non-manual workers*) was the reference group. Furthermore, possible effect modifiers (e.g. gender, age group, or PD status) on the association between exposures and PD risk were examined in **study I** and **IV**.

Proportional hazards assumption in the Cox model was examined in all studies, using scaled Schoenfeld residuals by plotting them against a function of the timescale and testing for linear trend.[166]

4.5.2 Meta-analysis (Study I & II)

To calculate the summarized effect of physical activity (**study I**) and snus use (**study II**) on the risk of PD, estimates of HRs and 95% CIs from each individual study or cohort were pooled using meta-analyses with a random-effects model.[167,168] Heterogeneity among studies included in the meta-analyses was examined using Q test and I^2 statistics.[169]

4.5.3 Conditional Logistic Regression (Study II)

Lifetime duration of snus use was calculated in years for all male twins nested in the Q-73 and SALT cohorts. PD cases were matched to five controls by age. A conditional logistic regression model was used to compare lifetime duration (years) of snus use in cases and controls.

4.5.4 Sensitivity analyses (Study I - IV)

A rich set of sensitivity analyses was performed for different purposes in each study. These analyses are introduced briefly below. Firstly, to exclude the possibility of reverse causality such that pre-clinical PD influences the exposures, lag-time analyses were conducted in **study I – III**. In **study I & II**, the first 8 years of follow-up were excluded, and in **study III**, the first 4 years were excluded. Secondly, to avoid potential confounding from tobacco smoking on the observed associations, analyses restricted to never smokers were performed in **study II & III**. Thirdly, the influence of primary or secondary PD diagnoses from the registers was also investigated by restricting the analyses to primary PD diagnosis as the definition of outcome (**study I**). Fourthly, multiple imputation was applied to assess the impact of missing values of covariates on the estimates of observed associations (**study I & II**). Further, to examine the effect of the cohort with most weight on the pooled estimates in the meta-analyses, the Construction Workers Cohort was excluded in **study II**. Finally, the effect of health seeking behavior on the association between SES and PD risk was examined in **study IV**.

5 MAIN RESULTS

5.1 Physical Activity and Parkinson's Disease (study I)

There was an inverse association between time spent on household and commuting activity and risk of PD. For example, there was an up to 43% risk reduction observed for participants who spent more than 6 hours on such activities (Table 5.1). A lower risk of PD was also found for participants with higher physically demanding levels of occupational activity, though this was not statistically significant.

Table 5.1 Risk of Parkinson's disease according to household and commuting activity and occupational activity for all participants in the Swedish National March Cohort

	n	**HR**	**95% CI**	***P*-trend**
Household and commuting activity				
<2 h/week	46	1	Reference	
3-4 h/week	78	0.68	0.46-1.00	
5-6 h/week	67	0.70	0.47-1.04	
>6 h/week	95	0.57	0.39-0.83	
				0.01
Physical demanding level of occupational activity				
Mostly sedentary	31	1	Reference	
Moving a little	209	0.91	0.60-1.38	
Strenuous	40	0.74	0.46-1.22	
				0.20

n = number of Parkinson's disease cases; HR = hazard ratio; CI = confidence interval.

When physical activities were analyzed in terms of MET-hours/day, there was no association between leisure time exercise and risk of PD (Figure 5.1). However, general physical activity (i.e. leisure time exercise and household and commuting activity combined) was associated with a lower risk of PD in men and in the full sample with both men and women. Notably, a 45% lower risk of PD (HR 0.55, 95% CI 0.35-0.87) was found for men with a medium level of total physical activity (mean = 39.1 MET-hours/day) compared to men with a low level of total physical activity (mean = 30.3 MET-hours/day).

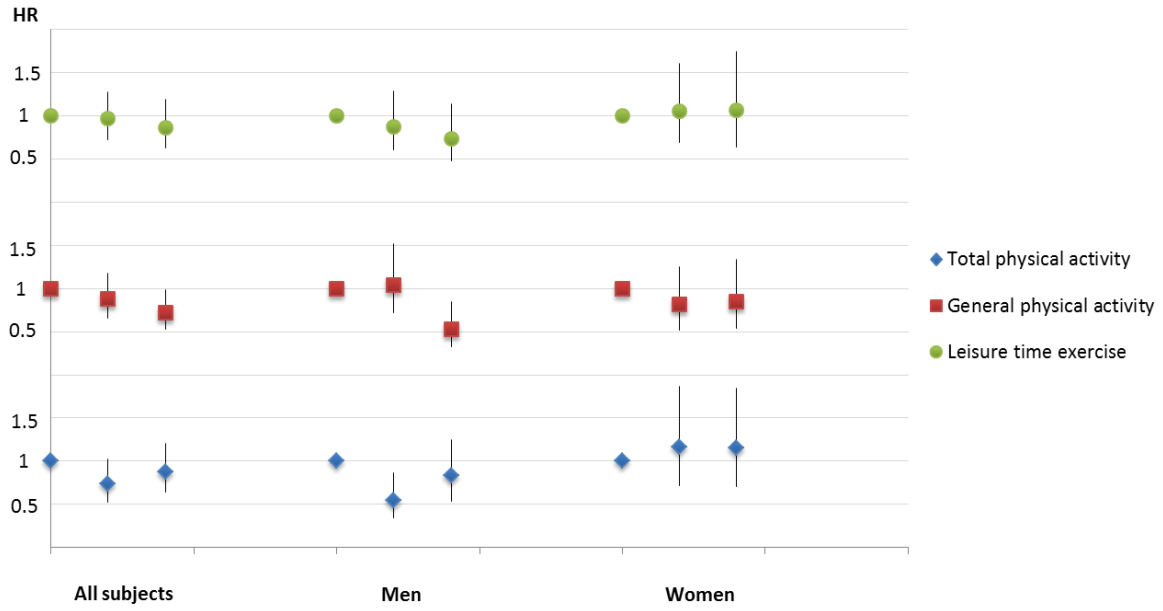

Figure 5.1 Associations between physical activity (MET-hours/day) and risk of Parkinson's disease

Physical activity was measured in MET-hours/day and categorized into tertiles. HR = hazard ratio.

We did not find an interaction effect with gender for any of the physical activity exposures mentioned above (*P* for interaction = 0.62, 0.62, 0.50, 0.15, 0.20 for household and commuting activity, occupational activity, leisure time exercise, general physical activity, and total physical activity, respectively).

When we pooled all previous cohort studies on physical activity in relation to risk of PD, four studies were included in the analyses stratified by gender.[52-55] We found that higher levels of physical activity were associated with lower risk of PD in both men and women (Figure 5.2).

Lastly, all sensitivity analyses showed similar results compared to the main analyses. For example, when only primary diagnosis of PD was used to define PD cases, multivariable-adjusted HR for the medium level of total physical activity was 0.52 (95% CI 0.31-0.90).

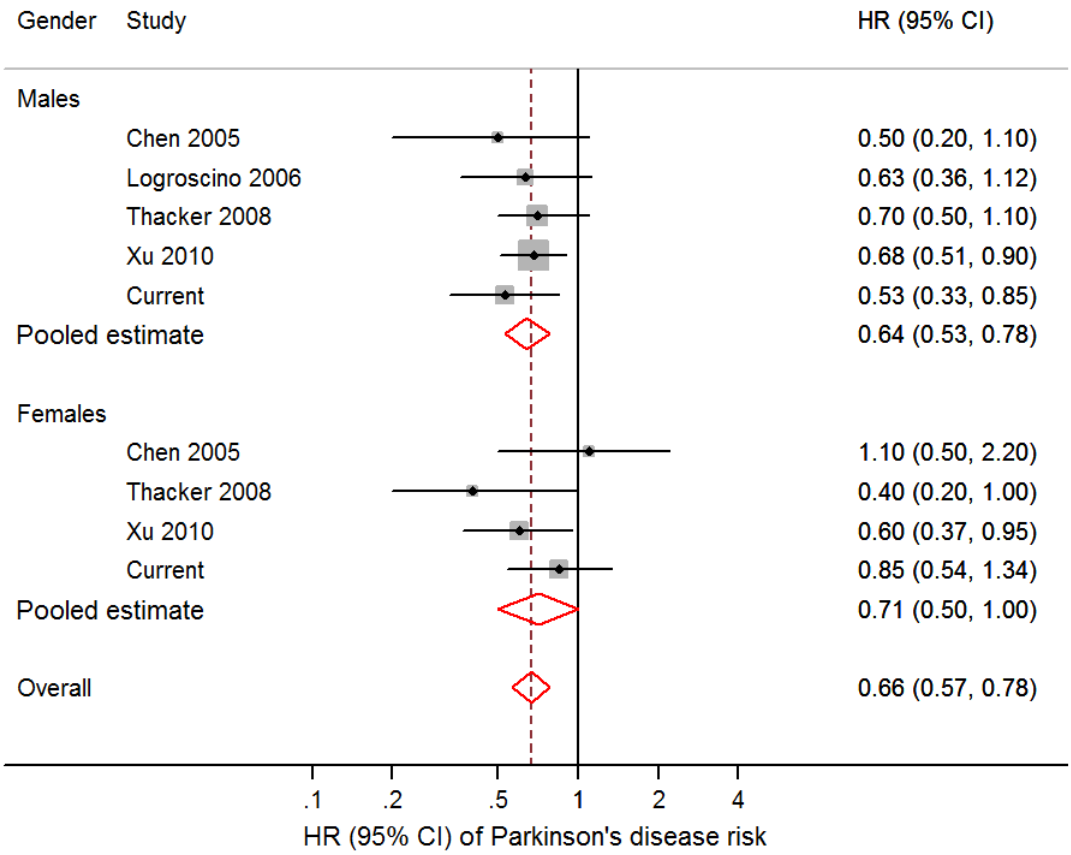

Figure 5.2 Forest plot - Pooled estimates of physical activity for risk of Parkinson's disease from all previous cohort studies (Reprinted from Yang F et al.,[170] with permission from Oxford University Press)

HRs were pooled using meta-analysis with a random-effects model. The size of each square is proportional to the weight of that study in the meta-analysis. HR = hazard ratio; CI = confidence interval.

5.2 Snus and Parkinson's Disease (Study II)

There was an inverse association between snus use and risk of PD in the entire study population (Figure 5.3). When analyses were restricted to never-tobacco smokers, men who ever used snus had a substantially lower risk of PD compared to never-snus users. The inverse association between snus use and PD risk was even greater for current-snus users, whereas an inverse but non-significant association between former-snus users and PD risk was observed (Table 5.2).

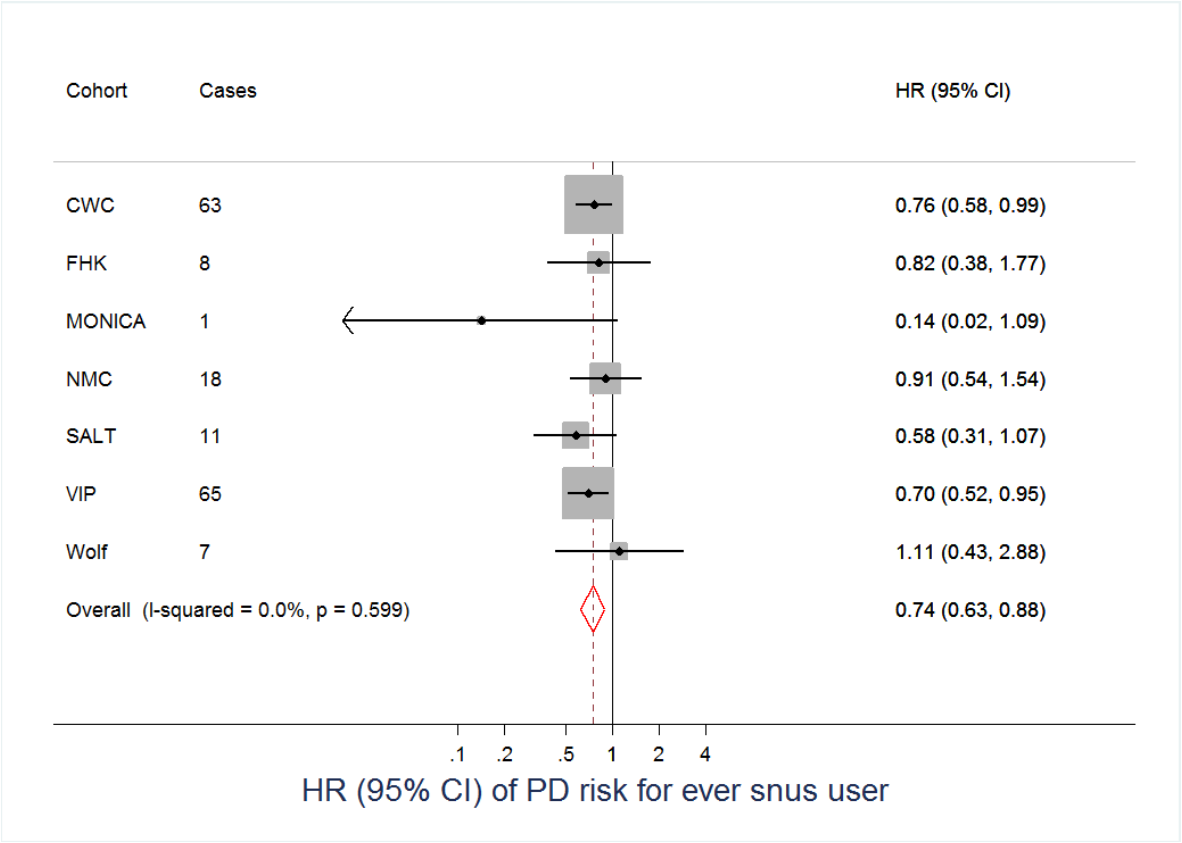

Figure 5.3 Forest plot - Pooled HR of snus use for risk of Parkinson's disease in seven Swedish prospective cohort studies

The size of each square is proportional to the weight of that study in the meta-analysis. Number of cases is for the exposed group (i.e. ever-snus users). Reference group is never-snus users. HR = hazard ratio, CI = confidence interval.

Table 5.2 Pooled HR of snus use for risk of Parkinson's disease in seven Swedish prospective cohort studies among never-tobacco smokers

	n	HR	95% CI	Heterogeneity	
				I^2	*P*-value
Never-snus users	531	1	Reference		
Ever-snus users	27	0.41	0.28 – 0.61	0%	0.93
Former-snus users	10	0.68	0.36 - 1.28	0%	0.91
Current-snus users	17	0.38	0.23 - 0.63	0%	0.73

n = number of cases; HR = hazard ratio; CI = confidence interval.

In addition, non-smoking men with higher snus consumption and longer period of snus use had lower risk of PD (Figure 5.4 & 5.5). Moreover, when analyzing snus use and tobacco smoking together, more prominent risk reductions were found with current use regardless of tobacco type (Figure 5.6). Furthermore, results from several sensitivity analyses were comparable to that in the main analyses.

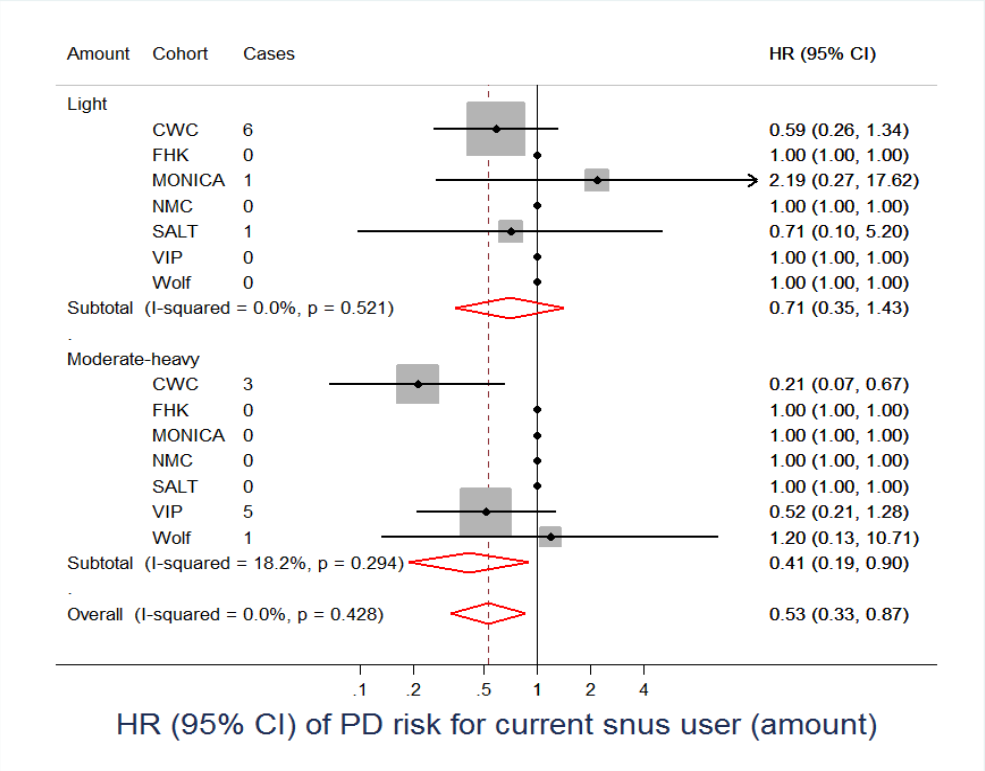

Figure 5.4 Forest plot - Pooled HR of amount of snus use for risk of Parkinson's disease

The size of each square is proportional to the weight of that study in the meta-analysis. Analyses were conducted among current-snus users. Reference group is never-snus users. Number of cases is for the exposed group. HR = hazard ratio, CI = confidence interval.

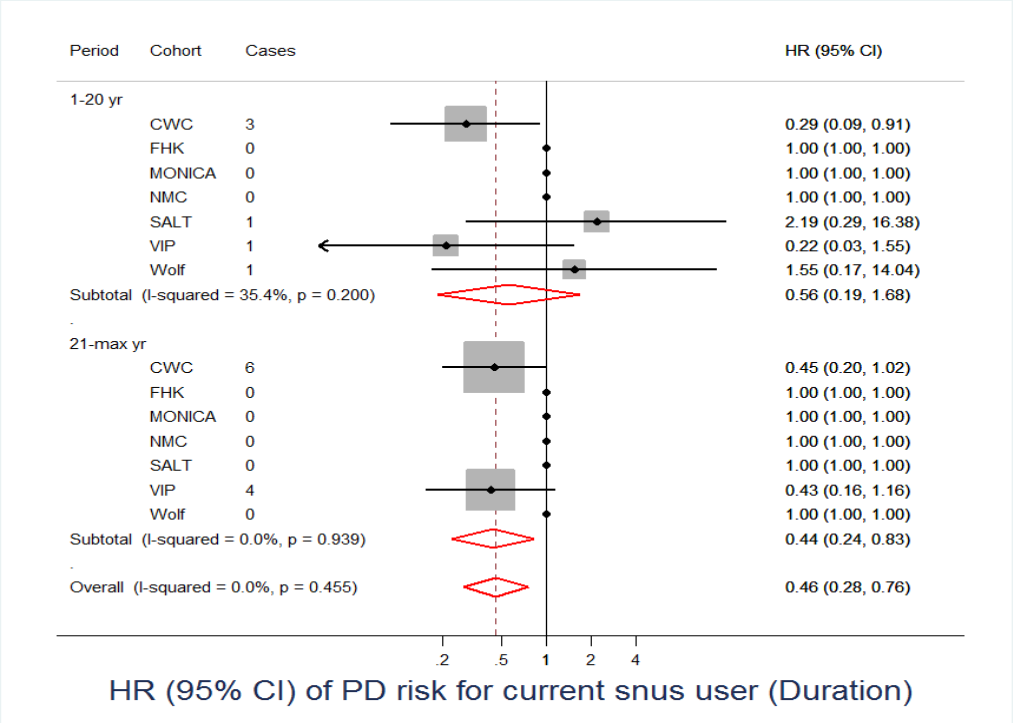

Figure 5.5 Forest plot - Pooled HR of duration of snus use for risk of Parkinson's disease

The size of each square is proportional to the weight of that study in the meta-analysis. Analyses were conducted among current-snus users. Reference group is never-snus users. Number of cases is for the exposed group. HR = hazard ratio, CI = confidence interval.

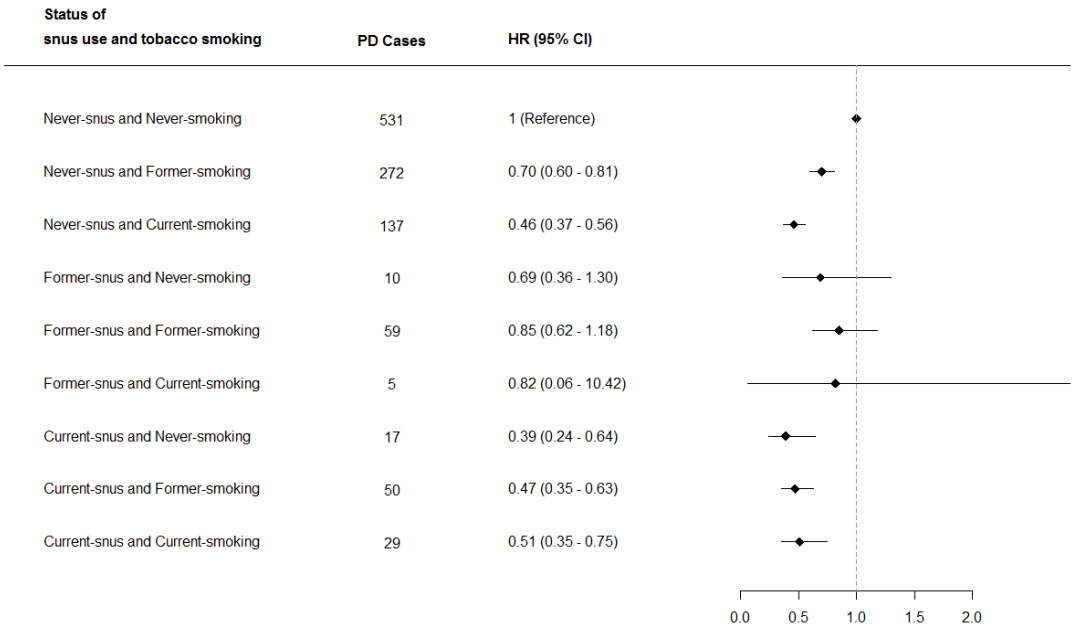

Figure 5.6 Combined effect of snus use and tobacco smoking on risk of Parkinson's disease (Reprinted from Yang F et al.,[171] with permission from Oxford University Press)

HR = hazard ratio, CI = confidence interval.

Lastly, when using SALT and Q-73 data and considering snus use, smoking, and alcohol drinking as time-varying exposures, PD cases had shorter duration of lifetime snus use than that in the controls (Table 5.3).

Table 5.3 Association between years of lifetime snus use and Parkinson's disease risk

	Never-smokers			
	n	**OR**	**95% CI**	***P*-value**
Per year of life-time snus use	115	0.96	0.93-1.00	0.04

Analyses were performed using conditional logistic regressions with cases and matched controls (n=5) by age, adjusting for smoking (never, former, and current users), alcohol intake, educational level, and birth cohort. n = number of Parkinson's disease cases; OR = odds ratio; CI = confidence interval.

5.3 Diet and Parkinson's Disease (Study III)

In the multivariable-adjusted model, we found a lower risk of PD with higher intake of dietary β-carotene in both genders, and with higher intake of vitamin E in women (Figure 5.7). Results remained significant in the 4-years lag-time analyses. In addition, the association of PD risk with higher intake of dietary vitamin C for women was borderline significant (HR 0.91, 95% CI 0.83-1.00 for an increase per 50 mg/day; *P*-trend for categorical analysis = 0.04). However, the inverse association for vitamin C intake in women was no longer significant in the 4-years lag-time analyses. There was no association between dietary TAC and PD risk.

When analyses were conducted among never smokers, there was an inverse association between intake of dietary vitamin E and risk of PD in women (HR 0.65, 95% CI 0.46-0.90 for highest quartile vs. lowest quartile; *P*-trend = 0.02). No other associations were observed.

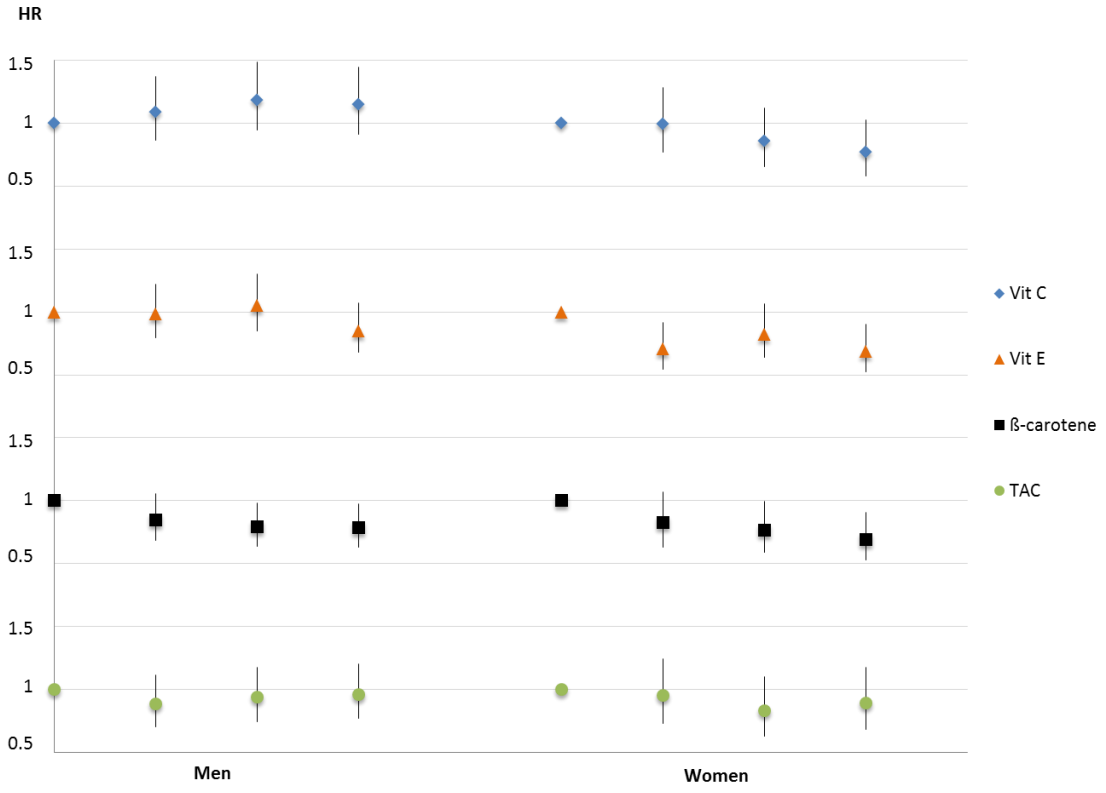

Figure 5.7 Associations between daily intake of dietary antioxidants and risk of Parkinson's disease

Intake of antioxidant vitamins was measured in mg/day. TAC was measured in micromole Trolox equivalents. All measures were categorized into quartiles with the lowest category as the reference group. HR = hazard ratio, TAC = total antioxidant capacity.

5.4 Socioeconomic Status and Parkinson's Disease (Study IV)

In general, individuals with lower SES were associated with lower risk of PD, but no clear dose-response relationship was found (Table 5.4). The associations were similar for all participants and for men and women separately. In addition, an analysis adjusting for hospital visits as a proxy of health seeking behavior did not change the association. Age group was an effect modifier for the association between SES and PD risk (*P*-value for interaction < 0.05), with weaker associations for low non-manual and high manual works in older ages.

Table 5.4 Associations between socioeconomic status and risk of Parkinson's disease for all participants

	n	**HR**	**95% CI**
Socioeconomic status			
High non-manual workers	12,435	1	Reference
Low non-manual workers	7,121	0.92	0.90 - 0.95
High manual workers	8,224	0.88	0.86 - 0.91
Low manual workers	18,537	0.93	0.91 - 0.95
Self-employed workers	4,311	0.96	0.93 - 0.99
Farmers	3,825	1.01	0.97 - 1.05
Pensioners	8,575	1.08	1.05 - 1.11
Unclassifiable	3,304	0.90	0.86 - 0.94

n = number of incident Parkinson's disease cases; HR = hazard ratio; CI = confidence interval.

When all-cause mortality was considered as the outcome, standardized mortality rate was on average 3-fold higher in PD patients compared to PD-free individuals. Importantly, there was an inverse association between SES and all-cause mortality for both PD-free individuals and PD patients (Figure 5.8). In addition, we found that PD diagnosis was an effect modifier for the association between SES and all-cause mortality (*P*-value for interaction < 0.01), with a somewhat attenuated inverse association among PD patients. The associations were similar for all participants and for men and women separately.

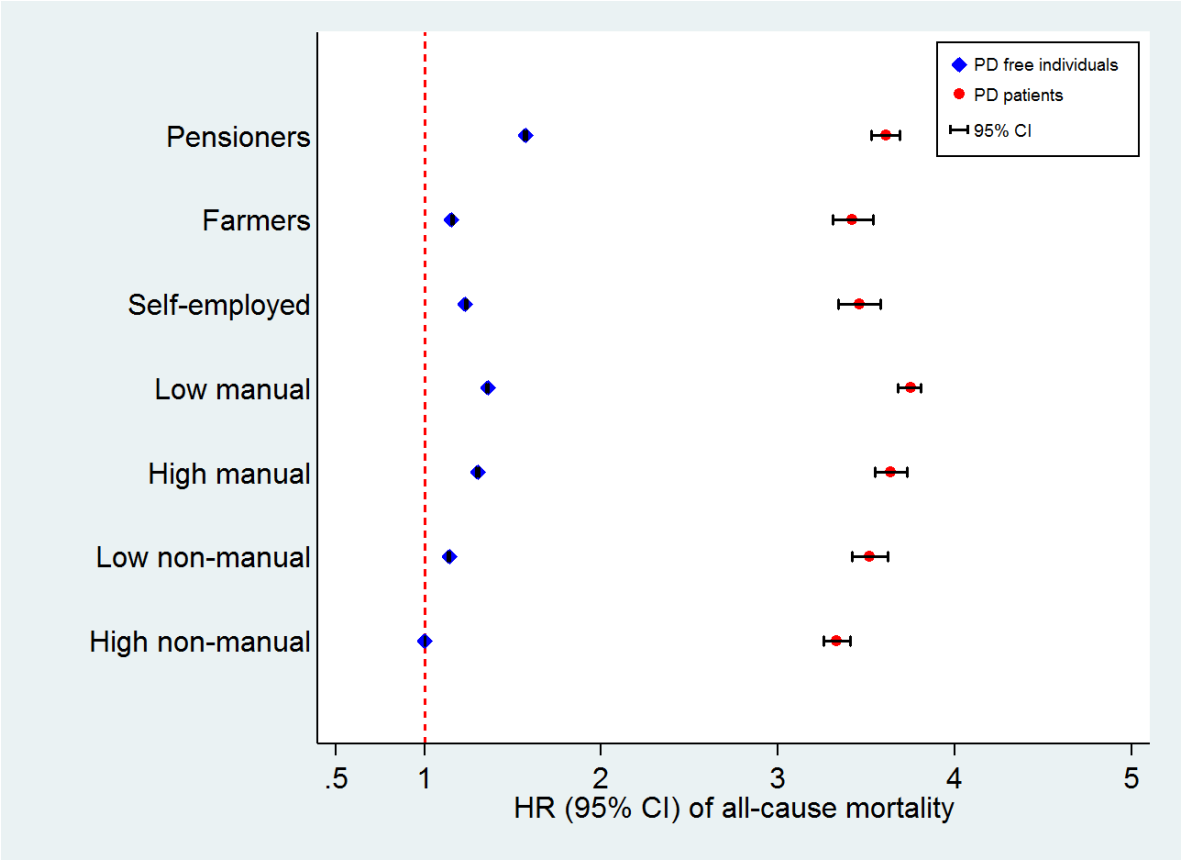

Figure 5.8 Associations between SES and all-cause mortality in PD-free individuals and patients with PD.

High non-manual workers were the reference group in this figure. SES = socioeconomic status, PD = Parkinson's disease, HR = hazard ratio, CI = confidence interval.

6 DISCUSSION AND METHODOLOGICAL CONSIDERATIONS

6.1 Association or Causation?

It is crucial to distinguish between the concepts of association and causation. Although results on lifestyle factors and risk of PD from this thesis are compelling, it is not possible to conclude that these lifestyle factors truly prevent against the disease. Rather, it has to be acknowledged that an association does not mean a causal relationship. However, several arguments should be highlighted, which have been used to counteract the concerns arising when implying a causal effect from observational results.

6.1.1 Study design

In epidemiology, a cohort study is when "*all subjects in a source population are classified according to their exposure status and followed over time to ascertain disease incidence*".[139] Under such a design with prospective data collection, exposures are measured at baseline (also known as the start of study) and during the follow up, an outcome (i.e. a disease) occurs later. Then, we can estimate the causal effect of the exposure on the outcome in some form, such as the ratio of the risk of the outcome in the exposed group compared to the risk among the unexposed group. Therefore, such an observed exposure-outcome relationship by nature implies causal inference, meaning that the exposure causes the outcome. In a retrospective cohort design, the study is initiated after the follow-up time has already passed and historical data are used. Nevertheless, if the exposure information was collected before the start of the study, the data collection is still considered prospective.

A randomized controlled trial (RCT) is a type of prospective cohort study in which the exposure (i.e. experiment or intervention) has been randomly assigned within the cohort. RCTs have been considered the "gold standard" in making causal inference in clinical research, but they may face unethical, unaffordable or unfeasible issues to apply the treatment or intervention on human subjects. Observational (also known as nonexperimental) studies, on the other hand, have several advantages in investigating exposure-outcome relationships, such as ethical acceptability, cost efficiency and real-life applicability.[172]

Compared with cross-sectional and retrospective case-control studies which are unable to conclude causality or prone to reverse-causality, an observational prospective cohort study is probably better at mimicking an RCT for testing a causal hypothesis. Other advantages of a prospective design include the possibilities to study a rare exposure or multiple outcomes, assess the time between exposure and outcome, and estimate prevalence, incidence and relative risks. Conducting a prospective cohort study, however, is typically very expensive and may need a long follow-up time for certain diseases to be developed. Another important issue that should be considered is the potential large proportion of drop-out during the long follow-up period.

In the present thesis, all studies (**study I-IV**) were conducted using a prospective cohort study design. As a result, all the exposure information were collected at the beginning of follow up and the first occurrence of PD diagnosis (**study I-IV**) or all-cause mortality (**study IV**) was the outcome. Therefore, potential problems arising from retrospective study designs such as recall bias were avoided. In addition, a nested case-control design was used in **study II**. A typical case-control study has advantages in studying rare diseases or diseases with a long latency period. It also allows studying multiple exposures in a single study. A nested case-control design is simply a case-control study nested in a cohort with a dynamic sampling method. Compared with the classical case-control design, the nested case-control design provides further opportunities to study time-varying exposures or outcomes, or in a twin or sibling design to adjust for unmeasured familial confounding. Furthermore, estimates of odds ratios from the nested case-control design may be closer to the HRs estimated from the cohort design.[173]

6.1.2 Prevalent cases versus incident cases

Dopaminergic neuron degeneration is gradual. A PD patient is diagnosed based on cardinal signs of movement dysfunctions but by that moment, at least 80% of neurons are already lost. Many symptoms have been proposed as prodromal symptoms of PD (i.e. symptoms are present but not sufficient to diagnose the disease), yet the exact onset of PD is largely uncertain.[174] In an epidemiological study, once the disease has developed, systematic measurement error of exposure will occur (i.e. differential misclassification) and the disease may alter the exposure (i.e. reverse causation), leading to biased results. Therefore, for the

purpose of risk factor identification, it is of utmost importance, to exclude prevalent cases at the beginning of follow-up and to study the effect of exposure on incident cases only.

For the studies presented in this thesis, information on diagnosis of PD was collected from population-based registers, in which the date of first inpatient admission and/or outpatient contact was defined as the index date of diagnosis for an incident PD case. The index date is probably the closest one to the date of disease onset when such information is not available. However, a previous study reported an average time period of 7.5 years between clinical onset and register record of PD diagnosis in the NPR.[165] Given this information, we performed lag-time analyses by excluding the first few years of follow up in **studies I - III**, and the conclusions remained. Regarding the possible long preclinical period (i.e. in which neurodegenerative processes start but no evident symptoms or signs are present), it was not possible to examine this issue since the data on prodromal symptoms such as olfactory loss or sleep disorder were not available.[175]

6.1.3 Aspects that may violate validity

To make a causal inference from an observed association, the validity of the results of a study is of similar importance as the aspects mentioned above. The validity of results can be summarized into three major components: internal validity, external validity, and statistical conclusion validity. Internal validity can be classified into three categories: information bias, selection bias, and confounding. External validity, on the contrary, refers to generalizability. Statistical conclusion validity is the degree to which conclusions can be drawn from the data unless specified otherwise.

6.1.3.1 Internal validity

Internal validity of epidemiological results is subject to validity of measurement (of exposure and outcome) that depends on both validity (i.e. absence of *systematic error* - also known as *bias*) and precision (i.e. absence of *random error*). Systematic error should be taken into consideration in both study design and data analysis, and results need to be interpreted with caution. Random error, however, can be dealt relatively easily by increasing the size of the study population.

Information bias (misclassification)

In epidemiological research, misclassification of exposure or outcome is a type of information bias arising from measurement error that can be further divided into differential and non-differential misclassification.[139]

Differential misclassification means that the degree of misclassification depends on other variables (status of exposure and/or outcome). As a result, this type of misclassification can be problematic and biases (either overestimates or underestimates) the true estimate for an observed association, perhaps in case-control and cross-sectional studies especially.[176]. On the other hand, a non-differential misclassification means that the probability of being misclassified among all study subjects is the same (i.e. *at random* and not related to other variables). It happens, for example, in a case-control study when participants (both cases and controls) do not remember their exposure correctly (i.e. recall bias), or in a cohort study when the exposure is measured at baseline and independent of the outcome (e.g. risk of disease).

In **study I** and **III**, physical activity level and dietary antioxidant vitamins were measured on an individual level through validated questionnaires with high quality, which makes misclassification of exposure less likely. In **study II**, tobacco use (both snus and cigarettes) was self-reported. It is known that people may underreport such behavior or completely deny using tobacco. There may also be a trend that cigarette smokers will quit smoking but initiate snus use as a smoking cessation method during a long follow-up period. In **study IV**, we used census data to assess SES, which may have low specificity. Another limitation is that the SES was characterized according to the SEI that is mainly classified by occupation, which may not represent SES correctly and further reduces specificity. In addition, a large proportion of participants were missing information on SEI in the 1980 census. Although we tried to trace SES information from the 1970 and 1960 censuses, this may introduce unforeseen misclassification. Thus, misclassification of exposure is more likely a concern in **study II** and **IV** compared to **study I** and **III**. However, in observational cohort studies using register-based data, such misclassification of exposure is independent of disease (i.e. non-differential) since the exposure information is collected prospectively, which may in theory lead to a dilution of the effect.

One should note the potential misclassification of outcome in register-based research. In this thesis, information on PD diagnosis was extracted from Swedish national health registers. Postmortem neuropathology information was not available. Neither clinical diagnosis nor

medical record was available to the researchers. Hospital discharge diagnoses of PD have been validated against rigorous clinical workup, showing a relatively high positive predictive value and sensitivity.[165] Although misclassifications of diagnoses between PD and other parkinsonian disorders are common, such misclassification between PD and other parkinsonian disorders goes both ways.[165]

In **study I**, we restricted the definition of PD to include only primary diagnoses in the registers for higher accuracy (i.e. positive predictive value). Although sensitivity decreases, results were very similar since specificity is almost perfect for PD diagnosis in the registers (range between 98.5-100%).[165] In addition, the positive predictive value for inpatient PD diagnosis was found to decline with age, indicating more misclassification in older age groups,[165] whereas studies in this thesis were population-based with no particular age restriction. Moreover, Swedish universal healthcare coverage ensures equal access to inpatient hospital care and analysis adjusting for hospital visits in **study IV** showed similar results. On the other hand, the validation study by Feldman et al. showed that PD diagnoses were less reliable in the CDR,[165] which could be a consequence of underreporting PD as a cause of death.[177-179] PD diagnoses from outpatient visits were not validated, yet it is reasonable to assume a relatively high accuracy for the diagnoses given from specialists. Therefore, misclassification of outcome, although likely present, is usually non-differential in prospective cohort studies. As a result, the true effect of the exposure on the outcome is expected to be stronger than the observed effect.

Selection bias

Selection bias may occur in the process of selecting participants caused by, for example, diagnostic and self-selection biases and differential loss to follow up.[139] It is one of the major concerns in observational studies. However, register-based observational research, to a large extent, eliminates such concern, since 1) no recruitment processes are needed, and 2) inclusion and exclusion criteria are usually independent of other factors and well predefined. Nevertheless, some issues should be noted.

In all studies presented in this thesis, incident PD was defined by at least one record of PD diagnosis in the NPR and CDR. Patients with PD identified from primary health care visits were not included. In addition, data from outpatient visits did not cover the entire spectrum of study period. Therefore, the incidence of PD could be underestimated, especially when

outpatient data were not available. If such bias exists, the estimates would be biased towards the null.

Looking at the different studies separately, the individuals included in **study I** were probably physically healthier and more health conscious than those who were not included, since they participated in the National March event and mailed questionnaires back to researchers. In addition, those with missing information on several components of total physical activity level were excluded from the study due to difficulty of calculating their daily total MET-hours. Similarly, in **study III**, individuals who actively finished FFQs from SMC and COSM were included. Study participants who joined the study but had missing information for several items on FFQs were lost from analyses. Another limitation of **study III** is the differential response rates between respondents and non-respondents for these two cohorts (70% for SMC and 49% for COSM). Selection bias may exist when differential disease status and SES cause such differences in response rates in **study I** and **III**. However, it is unclear whether all these factors are related to PD or not.

A special note should be given to a hypothesis suggested by Ritz and colleagues, which highlights a possibility that PD patients may be less nicotine dependent and consequently quit smoking more easily than non-PD individuals.[180] Accordingly, in **study II**, preclinical PD patients would quit snus more easily compared to non-PD individuals, and possibly even early in their lives, i.e. during adolescence. Then, a pseudo situation (i.e. observed) would happen with fewer snus users among preclinical PD patients. As a result, the observed inverse association between snus use and PD risk will be a consequence of selection bias due to reverse causation. Unfortunately, this hypothesis of nicotine dependence could not be tested with the data available in **study II**, nor could it be examined easily in other prospective studies since it is unclear whether the less nicotine-reward mechanism is truly associated with PD development.

In **study IV**, almost every Swedish inhabitant was initially included in the study but those below age 30 were excluded from the analysis, based on the assumption that SES was unstable before 30 years of age. However, there could be a proportion of individuals with a stable SES during follow up in that age group. In addition, SES was assessed based on labor market status and individuals with chronically disabling diseases (e.g. patients with early diagnosed PD) may therefore be excluded from the study since they tend to be outside the labor force. Furthermore, individuals with a high SES might have been more health conscious

and actively visited clinics, and were consequently more likely to be given a diagnosis of PD. As a result, they would have a higher observed incidence rate of PD compared to those with a low SES. On the other hand, individuals with a low SES may have higher morbidity and therefore visit healthcare more often than those with high SES. These individuals with low SES could also have higher chances to be given a diagnosis of PD. This assumption was tested by further adjusting for the number of times visiting hospitals for reasons other than PD in a sensitivity analysis. Results were largely comparable to the main analysis, suggesting that those individuals were not self-selected. Taken together, selection bias due to differences in health seeking behavior by SES is unlikely. Nevertheless, the models in **study I - IV** were all adjusted for educational level as a proxy of SES or health seeking behavior to account for such potential bias.

Confounding

An observed association between two factors can be causal such that one causes another, or they share a common cause as a confounding factor (i.e. confounder), or both. A confounder should be associated both with the exposure and outcome, but not as an effect of the exposure in the causal path way between the exposure and outcome (i.e. mediator), nor as the effect of both the exposure and outcome (i.e. collider).[139] Confounding occurs when exposed and unexposed groups differ in this third factor. Without proper adjustment for confounding, the association between an exposure and an outcome can be underestimated, overestimated, or the association may even be reversed. In epidemiology, causal diagrams (also known as Directed Acyclic Graphs, DAGs) have been widely used to identify potential confounders for data analysis and to establish a causal relationship visually (Figure 6.1).[181,182]

At the design stage of a study, unlike RCTs where participants and confounders are randomly assigned to both exposed and unexposed groups, observational studies are particularly prone to confounding. We can, however, avoid or adjust for confounding by applying *stratification* on confounders. In **study I, III**, and **IV**, men and women were analyzed separately. In **study II**, there were few snus users among women and the biggest confounder considered was tobacco products other than snus. Therefore, primary analyses were restricted to never smokers of tobacco products.

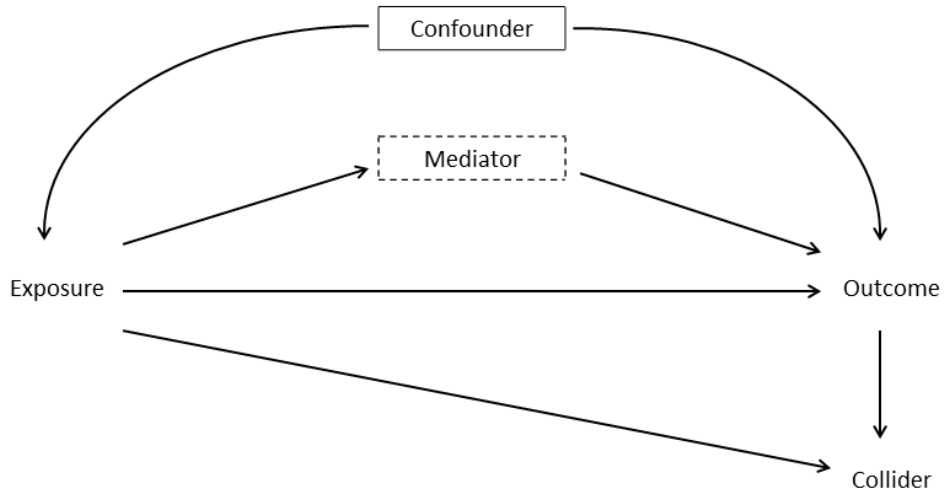

Figure 6.1 An illustration of concepts in causal inference from a Directed Acyclic Graph

We identified confounders based on DAGs and subject knowledge in all studies in this thesis. At the analysis stage where confounding information was available, confounding was examined through adjustment in regression models. Unmeasured or residual confounding, however, can still exist in these studies. For example, although cigarette smoking was controlled for through adjustment in regression models in **study I** and **III**, it was crudely classified into never, current and former smokers. In **study IV**, smoking information was not available but considered as potential explanation for the observed inverse association between SES and risk of PD. In addition, genetic confounding of the association between snus use and risk of PD was not taken into consideration in **study II**.

Mediation is another issue worth noting. For example, in **study I**, BMI was adjusted for as a confounder. However, it may also be considered as a mediator between physical activity and risk of PD. Adjustment for a mediator changes the interpretation of the results regarding how much of the association between physical activity and PD risk can be attributed to the specific

covariate. Unfortunately, with available data, we were unable to assess the exact role of BMI in this association.

6.1.3.2 External validity

External validity (also known as generalizability) means whether results from a particular study can be generalized to individuals in another population. All studies in this thesis were based on data from Sweden. Our results may be generalizable to another Scandinavian country since they may share similar population characteristics, but it may not be applicable to other countries. With regards to **study I**, given the specific characteristics of the event (i.e. National March for cancer prevention fundraising), it is reasonable to believe that participants are generally more physically active than others. Therefore, results may not apply to less physically active individuals. **Study II** was restricted to men and results cannot be generalized to women. Individuals in **study III** were in middle to older age groups; therefore the results cannot be generalized to younger cohorts.

6.1.3.3 Statistical conclusion validity

A causal effect of statistical inference is conceptually different from causal inference. Statistical inference derives from an association being correctly estimated using a counterfactual approach (i.e. proper statistical methods), whereas, for example, DAGs conceptualize the problem graphically and infer causality. We used mainly the Cox regression model in studies included in the present thesis to estimate HR as RR of the outcome between exposed and unexposed groups. A few other statistical methods were used in different studies for correct estimates of statistical inference and are worth discussing.

Interaction

Interaction or effect modification can be classified into biological (i.e. causal) and statistical interactions. A biological interaction exists when the effect of one factor (i.e. exposure) on the outcome is dependent on the presence of another factor (i.e. effect modifier) and when some specific mechanism or hypothesis is available. A statistical interaction is an effect-measure modification that measures changes of effect of the exposure over values of the effect modifier under the assumption of no bias, for example, when physical activity has differential effects (e.g. positive vs. negative or additive vs. multiplicative) on PD risk in men and women. If a biological or social hypothesis is available, statistical interaction should be

assessed. In **study I**, gender was hypothesized as an effect modifier and examined in the analyses. In **study IV**, PD diagnosis was considered an effect modifier and time-varying covariate for the association between SES and all-cause mortality, and stratified analyses were consequently performed.

Multiple testing

Multiple testing or multiple comparisons is an issue when many statistical tests are performed simultaneously or repeatedly in a set. The probability to reject the null hypothesis incorrectly or to observe a rare event by chance will increase, resulting in a false positive or Type I error. Multiple testing is common in genetic studies where numerous genetic variants are tested for a single trait. Several statistical techniques have been developed to address this issue in order to get an unbiased statistical inference. The *Bonferroni* correction is the most acknowledged method requiring a higher significance threshold to reject the null hypothesis.[183] Although correction for multiple testing was performed in **study I** and the conclusion remained the same, in conventional epidemiological research (e.g. studies included in this thesis), adjustment for multiple comparisons is usually not necessary. The primary reasons are that the number of statistical analyses is limited and they are only performed when there is a strong hypothesis.[184]

Relatedness between study subjects

Another issue one should note is the relatedness between study subjects (e.g. twins in a twin pair) in the study. Such dependence between individuals would affect the estimate of the standard error. Under the hypothesis that monozygotic twin pairs share 100% and dizygotic twin pairs share 50% of their genetic background, using twin data in epidemiological studies requires particular attention regarding the statistical analysis. In **study II** where twin data were used, we applied a robust sandwich estimator that takes into account the relatedness in data to estimate the '*correct'* standard errors. In fact, estimates of standard errors with and without robust sandwich estimator were similar in this study, owning to a very large sample size.

Meta-analysis

A meta-analysis systematically reviews and summarizes all pertinent evidence on a particular research question. In **study I**, HRs from the study and previous prospective cohort studies on physical activity and risk of PD were meta-analyzed with a two-stage method that uses extracted aggregate data from each study. On the one hand, this approach provides an overall

interpretation and evidence synthesis of all the studies included in the analysis. On the other hand, although results were compelling, this method may be subject to different biases.[185] In **study II**, we performed meta-analysis utilizing individual participant data, where definitions of exposures, outcomes and covariates, and study-specific analyses used the same protocol. This individual participant data meta-analysis offers several advantages compared to conventional aggregate data meta-analysis, and is considered the gold standard approach to evidence synthesis.[186,187]

6.2 Ethical Considerations

In medical research, it is very important to ensure the study participant's privacy and individual integrity. This brings trust to the public and reinforces a good relationship between the scientific community and public society. In principle, individuals participating in any medical research should be contacted and well-informed about the purpose and potential consequences of the study. They need to agree and give researchers their consent every time their sensitive personal information is collected and used for a specific purpose. According to the Swedish law (Lag om etikprövning 2003:460), personal information such as medical history is sensitive and all studies based on sensitive data must be approved by an ethical committee before the study starts. This ethical committee will judge whether the potential benefit of the study outweighs the potential harm to the participants.

It should be noted that all studies conducted in this thesis were register-based and data were pseudonymous. All studies were approved by the Ethical Review Board in Stockholm. In addition, study participants' sensitive data were provided by the governmental authority (i.e. Statistics Sweden and National Board of Health and Welfare) and reverse identification is strictly prohibited and unlikely.

In my opinion, it would be impractical or economically unfeasible to acquire informed consent from every single individual, as thousands and even millions of participants were included in these studies. Although some findings in this thesis can be potentially sensitive to a certain group of participants, I believe the risk of violating participants' integrity is minimal. However, I think the ways of handling sensitive information such as transforming data and distributing or discussing findings are more important. This applies, in particular, to future epidemiological studies where "big data" have the ability to identify an individual from

even anonymous information. Then, data safety will rely more on researcher's scientific quality and ethic judgment for human dignity. Nevertheless, we as researchers should do our best to protect the integrity of all individuals in the study.

7 INTERPRETATION AND CONCLUSION

It is never too ambitious to state that I want to defeat PD and I very much hope my work may contribute to the prevention of the disease. However, every scientific work, including studies in this thesis, can only solve a small piece of the puzzle. It may also be too early to judge the true impact of findings yielded from this thesis, but I embrace them simply because they answered at least one important research question in each study. Taking into consideration ongoing development in the field, I believe the work described in this thesis provide a significant fragment of the big picture.

Physical activity may influence inflammation and oxidative stress, which in turn are considered putative mechanisms in PD pathogenesis. In the large prospective cohort study with nearly 13 years of follow up (**study I**), people with more than 6 hours per week of moderate level of physical activity had a substantially lower risk of PD. Compared with previous studies with similar findings, our results illustrate that all physical activity, not just physical exercise, can reduce PD risk. This message is particularly important to the public both for PD prevention and for the general health benefits of physical activity. Moreover, physical activity strategy can be relatively easy and cheap to implement in practice. Furthermore, physical activity has been demonstrated to maintain and improve function of patients already diagnosed with PD.

A well-known but intriguing finding in PD research is the inverse association between cigarette smoking and risk of PD. We used a novel method to examine the potential effect of nicotine from smokeless tobacco on PD risk (**study II**). Compared to the possible beneficial effect of smoking, using snus had an even stronger protective effect on risk of PD. Importantly, our primary results were restricted to never smokers, largely alleviating any potential confounding effect from smoking. However, we need particular caution to interpret these results. First, smoking has many adverse health effects and needs to be prohibited. Second, although scientific evidence on snus use and health consequences is limited, nicotine in itself is highly addictive and advocating snus use is unethical. Moreover, the hypothesis about a less nicotine-mediated reward among preclinical PD patients suggests a reverse causation bias for the neuroprotective effect of nicotine in PD. Nevertheless, nicotine or other components of tobacco leaves may indeed play an important role in PD pathogenesis. Therefore, further research is warranted before promoting nicotine as a candidate treatment in a PD prevention clinical trial.

Intake of dietary vitamin E and β-carotene was associated with a lower risk of PD in two large population-based cohorts (**study III**). Although diet and dietary factors in relation to disease risk are of high interest in the society, many findings in this field are not reproducible. This is primarily due to the difficulty of diet assessment and the heterogeneity of diet. Nevertheless, taking together all the evidence on diet and PD, we conclude that dietary antioxidants such as vitamin E and β-carotene are associated with lower risk of PD. To explore whether or not they can prevent PD, further studies must move to the molecular level and discover the underlying mechanisms.

In the register-based study of almost the entire Swedish population (**study IV**), we found that lower SES was associated with lower risk of PD and with higher all-cause mortality. In addition, the impact of SES on all-cause mortality was weaker in PD patients than non-PD individuals. The purpose of this study was not only disease risk identification but also to generate new hypotheses about risk factors. At this point, with limited information on the individual level from the registers, the role of SES in PD prevention remains inconclusive. However, the three-fold higher all-cause mortality for PD patients compared to the general population needs further attention.

In conclusion, I think that an easily applicable program for PD prevention is not available based on current epidemiological evidence on lifestyle factors and PD. The results in this thesis, however, provide fruitful insights into a promising way of preventing PD by advocating a better lifestyle such as an increase in physical activity and a healthy dietary pattern.

8 FUTURE PERSPECTIVES

PD as a chronic aging related disease will pose more social and economic challenges to both patients and society as life expectancy is increasing worldwide. Since the essay written by James Parkinson exactly 200 years ago, numerous investigations have tried to uncover the truth of the disease. If we knew the cause of PD, it would be much simpler. We could either focus on drug development or eliminate factors that lead to the disease, or both. Unfortunately, we do not know why people develop PD.

Our understanding of PD has increased through scientific and technological development. Laboratory data are important for establishing causal pathways between biological candidates and the disease, but using lab techniques to identify risk factors for a disease is just like a Chinese old proverb "大海捞针", meaning "searching a needle in the ocean". Genetic studies have provided significant insights into the disease pathogenesis and etiology in the last decade. However, until now, identifiable genetic causes are only available for a small proportion of PD, while the majority of PD patients are sporadic cases. It may seem frustrating at this moment but epidemiological (i.e. observational) evidence, with several longitudinal studies, have elucidated various risk factors as well as protective factors for the disease.

Identifying risk factors is useful for pinpointing potential biological targets and developing targets for disease modifying therapies. Identification of genetic and non-genetic factors, and most likely, the gene-environment interplay will be the mainstay to determine the exact causality underlying selective dopamine neuron death in PD. Then, clinically definitive diagnosis can be made at the true onset of the disease. Eventually, detection of modifiable risk factors for PD will enable prevention of the disease.

- THE END –

9 ACKNOWLEDGEMENTS

It has been a long journey for me as a PhD student at the Department of Medical Epidemiology and Biostatistics (MEB). I feel very fortunate to be a MEBer working with everyone here. I want to express my sincere gratitude to all of you who have shared a moment with me at MEB.

Karin Wirdefeldt – My dearest main supervisor, I would like to thank you first. I will not forget the moment when we first met. I told you that I needed to go for an internship before joining the group and you said that this position will certainly wait for me. I am really grateful for the opportunity that you gave me. I thank you deeply for the time you spent on my project and for guiding me thoroughly in epidemiology, medicine and scientific writing. I think we may have different personalities but we are both quite stubborn with our scientific opinions and the way we expressing them. I absolutely appreciate all the in-depth discussions. Without them, I couldn't have learned about the creativity in science.

Nancy Pedersen – How can I express my thanks to you! You are always kind, nice and willing to share your knowledge and experience with me. I cannot imagine that I will ever be as wisdom as you. I am very honored to have you as my co-supervisor. The research environment you created for our research group is fantastic and it made me feel at home all the time. I really thank you for all the guidance in science and my future career development.

Fang Fang – I am very glad that you agreed to be my co-supervisor. You are always there to support and encourage me. Thank you for your valuable inputs on my manuscripts. It has been really fun to work with you!

Asgeir R. Helgason – Thank you for being the supervisor of my master thesis and scientific mentor during my PhD study.

To other co-authors **Ylva Trolle Lagerros**, **Rino Bellocco**, **Hans-Olov Adami**, **Weimin Ye** , **Zhiwei Liu**, **Margareta Norberg**, **Lars Forsgren**, **Lars Alfredsson**, **Anders Knutsson**, **Jan-Håkan Jansson**, **Patrik Wennberg**, **Rosaria Galanti**, **Anton Lager**, **Marzieh Araghi**, **Michael Lundberg**, **Cecilia Magnusson**, **Alicja Wolk**, **Niclas Håkansson**, **Anna Johansson**, **Margaret Gatz**, thank you all for sharing your knowledge with me. Our collaborations made the studies in this thesis great.

To other (former) research group members **Adina Feldman**, **Andrea Foebel**, **Anna Dahl Aslan**, **Bojing Liu**, **Ci Song**, **Catalina Zavala**, **Dylan Williams**, **Ida Karlsson**, **Johanna Sieurin**, **Juulia Jylhävä**, **Kelli Lehto**, **Kathleen Bokenberger**, **Lotte Gerritsen**, **Miriam Mosing**, **Malin Ericsson**, **Robert Miller**, **Sara Hägg**, **Xu Chen**, **Yiqiang Zhan**, **Yunzhang Wang**, **Erika Nordenhagen**, and **Michaela Rydén**, I enjoyed every moment with you guys and I wish you all the best in the future.

Bojing Liu and **Johanna Sieurin**, thank you for sharing the time being a PhD student together with me in Karin's group. I will remember our Berlin trip forever. Johanna, thank you for sharing the office with me and I enjoy all our chats and occasional Swedish lessons.

Special thanks to **Ida Karlsson** – I will remember you love beers and you may not forget how much I dislike walking. It is really great to have you as my friend. Thank you for all the support and help during my PhD period. By the way, your dissertation is also close and I know you will nail it!

Special thanks to **Sofi (Karla Gutierrez Gonzalez)** and **Abi (Abhinav Gupta)** – Our time together was short, but one day, I will visit you two in Costa Rica and United State. And to **Sue Hui Tan** and **Vincent Fremont**, our friendships will stay no matter when and where.

Anna Plym, **Daniela Mariosa**, **Eva Herweijer**, **Johanna Holm**, **Linda Abrahamsson**, thank you for being my friends. I will remember all the lovely activities we did together. **Barbro Sandin**, thank you very much for the cover photo and all the exciting discussions on SAS coding. **Karin Larsson**, thank you for taking me to the internship.

To current MEB colleagues **Hannah Bower**, **Isabell Brikell**, **Carolyn Cesta**, **Elisabeth Dahlqwist**, **Bénédicte Delcoigne**, **Isabella Ekheden**, **Laura Ghirardi**, **Nelson Ndegwa Gichora**, **Alessandra Grotta**, **Bronwyn Brew**, **Anna Hedman**, **Gabriel Isheden**, **Andreas Jangmo**, **Anna Johansson**, **Robert Karlsson**, **Andreas Karlsson**, **Ralf Kuja-Halkola**, **Anna Kähler**, **Favelle Lamb**, **Mariam Lashkariani**, **Frida Lundberg**, **Cecilia Lundholm**, **Patrik Magnusson**, **Henrik Olsson**, **Frida Palmér Thisell**, **Yudi Pawitan**, **Erik Pettersson**, **Marie Reilly**, **Mina Rydell**, **Amir Sariaslan**, **Arvid Sjölander**, **Camilla Sjörs f.d. Gard**, **Peter Ström**, **Annika Tillander**, **Emilio Ugalde Morales**, **Vihelmina Ullemar**, **Thomas Whitington**, **Camilla Wiklund**, **Johan Zetterqvist**, and former MEBers **Amy**

Leval, **Anne Örtqvist**, **Andrea Ganna**, **Lisa Möller**, **Stephanie Bonn**, **Tracy Peters**, **Iffat Rahman**, **Miriam Elfström**, **Myeongjee Lee**, **Jiaqi Huang**, **Zhiwei Liu**, **Xingdong Chen**, **Gao He**, **Haiyun Wang**, **Tong Gong**, **Chen Suo**, **Shu Mei Teo**, thank you all for the happy talks during lunch time. I will certainly meet you in one way or another, either in science-related or other fun activities. I also would like to thank **Rikard Öberg**, **Vivekananda Lanka**, **Frida Lundberg**, and **Patrik Magnusson** for the interesting talks about motorcycles.

And to Chinese MEBers, **Zheng Chang**, **Qi Chen**, **Ruoqing Chen**, **Wei He**, **Tingting Huang**, **Jiayao Lei**, **Xingrong Liu**, **Yi Lu**, **Donghao Lu**, **Zheng Ning**, **Xia Shen**, **Qing Shen**, **Jie Song**, **Huan Song**, **Shihua Sun**, **Mei Wang**, **Jiangrong Wang**, **Haomin Yang**, **Shuyang Yao**, **Li Yin**, and **Jianwei Zhu**, thank you all for creating an Chinese environment at MEB. It is so great to have you guys around to make me not that homesick.

Lastly, I would like to thank **Camilla Ahlqvist**, **Gunilla Nilsson Roos**, **Gunilla Sonnebring**, and **Marie Jansson** for the administrative support and **MEB-IT group** for their continuous efforts to make MEB an attractive, enjoyable and fun place to work.

我的家人啊，我是如此的深爱着你们。老妈，从小你就告诉我要追逐梦想。为了不当一个坏小子，你从初中开始就把我送离了家，并且你好像一点都不担心我样的。其实，我很感谢你当初做的决定，并且我也证明了你这么多年的支持是对的。儿子不会让你失望的。老爸，尽管你从来都不说，但我知道你心中一定很牵挂着我。好男儿志在四方，儿子会一直努力的。啰啰，我应该是上辈子积德了才有你当我的兄弟。我时不时会有点后悔自己不能当一个负责的哥哥在你身边保护你。但是你现在已经成长为一个男人了，我想你会知道我希望自己能在你长大的时候陪伴你更多的时间。不管咋样，我们是兄弟，并且我会无条件的永远爱护你。

傻宝，如果不学医，如果不来瑞典，如果不读博士，我可能不会遇见你。但是我做的这么多决定最后让我遇到了你，这在统计学里会是多小的机率啊。很感谢你陪伴着我的这么多年，但是我，傻蛋和蛋卷会陪伴你更久的。Och till Minmin och Benben, tack så mycket för att ni accepterar mig som en av familjen.

10 REFERENCES

1. Parkinson J. Essay on the shaking palsy. *Whittingham and Roland* 1817;for Sherwood, Neely, and Jones
2. Kempster PA, Hurwitz B, Lees AJ. A new look at James Parkinson's Essay on the Shaking Palsy. *Neurology.* Jul 31 2007;69(5):482-485.
3. Louis ED. The shaking palsy, the first forty-five years: a journey through the British literature. *Mov Disord.* Nov 1997;12(6):1068-1072.
4. Goetz CG. Charcot on Parkinson's disease. *Mov Disord.* 1986;1(1):27-32.
5. Lill CM, Roehr JT, McQueen MB, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet.* 2012;8(3):e1002548.
6. Kalia LV, Lang AE. Parkinson's disease. *Lancet.* Aug 29 2015;386(9996):896-912.
7. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine.* Jan 2012;2(1):a008888.
8. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat Disord.* Feb 2016;23:1-9.
9. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet neurology.* Jun 2006;5(6):525-535.
10. Nussbaum RL, Ellis CE. Alzheimer's disease and Parkinson's disease. *N Engl J Med.* Apr 03 2003;348(14):1356-1364.
11. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. *Nat Rev Neurol.* Jan 2013;9(1):13-24.
12. Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging.* Mar-Apr 2003;24(2):197-211.
13. Hughes AJ, Ben-Shlomo Y, Daniel SE, Lees AJ. What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study. *Neurology.* Jun 1992;42(6):1142-1146.
14. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. *Arch Neurol.* Jan 1999;56(1):33-39.
15. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry.* Jun 1988;51(6):745-752.
16. Ward CD, Gibb WR. Research diagnostic criteria for Parkinson's disease. *Adv Neurol.* 1990;53:245-249.
17. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry.* Mar 1992;55(3):181-184.
18. Berg D, Lang AE, Postuma RB, et al. Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. *Lancet Neurol.* May 2013;12(5):514-524.
19. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* Oct 2015;30(12):1591-1601.

20. Berg D, Postuma RB, Bloem B, et al. Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease. *Mov Disord.* Apr 2014;29(4):454-462.
21. Li J, Jin M, Wang L, Qin B, Wang K. MDS clinical diagnostic criteria for Parkinson's disease in China. *J Neurol.* Dec 26 2016.
22. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord.* Nov 2014;29(13):1583-1590.
23. Van Den Eeden SK, Tanner CM, Bernstein AL, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol.* Jun 1 2003;157(11):1015-1022.
24. Wirdefeldt K, Adami HO, Cole P, Trichopoulos D, Mandel J. Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol.* Jun 2011;26 Suppl 1:S1-58.
25. Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Influence of strict, intermediate, and broad diagnostic criteria on the age- and sex-specific incidence of Parkinson's disease. *Mov Disord.* Sep 2000;15(5):819-825.
26. Twelves D, Perkins KS, Counsell C. Systematic review of incidence studies of Parkinson's disease. *Mov Disord.* Jan 2003;18(1):19-31.
27. Hirtz D, Thurman DJ, Gwinn-Hardy K, Mohamed M, Chaudhuri AR, Zalutsky R. How common are the "common" neurologic disorders? *Neurology.* Jan 30 2007;68(5):326-337.
28. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. *Lancet Neurol.* Nov 2016;15(12):1257-1272.
29. Taylor KS, Cook JA, Counsell CE. Heterogeneity in male to female risk for Parkinson's disease. *J Neurol Neurosurg Psychiatry.* Aug 2007;78(8):905-906.
30. Elbaz A, Bower JH, Maraganore DM, et al. Risk tables for parkinsonism and Parkinson's disease. *J Clin Epidemiol.* Jan 2002;55(1):25-31.
31. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science.* Jun 27 1997;276(5321):2045-2047.
32. Leroy E, Boyer R, Auburger G, et al. The ubiquitin pathway in Parkinson's disease. *Nature.* 1998;395(6701):451-452.
33. Lesage S, Brice A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet.* Apr 15 2009;18(R1):R48-59.
34. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet.* Sep 2014;46(9):989-993.
35. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging.* Oct 2011;32(10):1923 e1921-1928.
36. Marttila RJ, Kaprio J, Koskenvuo M, Rinne UK. Parkinson's disease in a nationwide twin cohort. *Neurology.* 1988;38(8):1217-1219.
37. Vieregge P, Schiffke KA, Friedrich HJ, Muller B, Ludin HP. Parkinson's disease in twins. *Neurology.* 1992;42(8):1453-1461.
38. Vieregge P, Hagenah J, Heberlein I, Klein C, Ludin HP. Parkinson's disease in twins: a follow-up study. *Neurology.* 1999;53(3):566-572.

39. Tanner CM, Ottman R, Goldman SM, et al. Parkinson disease in twins: an etiologic study. *Jama.* 1999;281(4):341-346.
40. Wirdefeldt K, Gatz M, Schalling M, Pedersen NL. No evidence for heritability of Parkinson disease in Swedish twins. *Neurology.* Jul 27 2004;63(2):305-311.
41. Thacker EL, Ascherio A. Familial aggregation of Parkinson's disease: a meta-analysis. *Mov Disord.* Jun 15 2008;23(8):1174-1183.
42. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. Meta-analysis of early nonmotor features and risk factors for Parkinson disease. *Ann Neurol.* Dec 2012;72(6):893-901.
43. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* Mar-Apr 1985;100(2):126-131.
44. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Human nutrition. Clinical nutrition.* 1985;39 Suppl 1:5-41.
45. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc.* Jan 1993;25(1):71-80.
46. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc.* Sep 2000;32(9 Suppl):S498-504.
47. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc.* Aug 2011;43(8):1575-1581.
48. Sasco AJ, Paffenbarger RS, Jr., Gendre I, Wing AL. The role of physical exercise in the occurrence of Parkinson's disease. *Arch Neurol-Chicago.* Apr 1992;49(4):360-365.
49. Kuopio AM, Marttila RJ, Helenius H, Rinne UK. Environmental risk factors in Parkinson's disease. *Mov Disord.* Nov 1999;14(6):928-939.
50. Fertl E, Doppelbauer A, Auff E. Physical activity and sports in patients suffering from Parkinson's disease in comparison with healthy seniors. *J Neural Transm Park Dis Dement Sect.* 1993;5(2):157-161.
51. Shih IF, Liew Z, Krause N, Ritz B. Lifetime occupational and leisure time physical activity and risk of Parkinson's disease. *Parkinsonism Relat Disord.* Jul 2016;28:112-117.
52. Logroscino G, Sesso HD, Paffenbarger RS, Jr., Lee IM. Physical activity and risk of Parkinson's disease: a prospective cohort study. *J Neurol Neurosurg Psychiatry.* Dec 2006;77(12):1318-1322.
53. Thacker EL, Chen H, Patel AV, et al. Recreational physical activity and risk of Parkinson's disease. *Mov Disord.* Jan 2008;23(1):69-74.
54. Chen H, Zhang SM, Schwarzschild MA, Hernan MA, Ascherio A. Physical activity and the risk of Parkinson disease. *Neurology.* Feb 22 2005;64(4):664-669.
55. Xu Q, Park Y, Huang X, et al. Physical activities and future risk of Parkinson disease. *Neurology.* Jul 27 2010;75(4):341-348.
56. Saaksjarvi K, Knekt P, Mannisto S, et al. Reduced risk of Parkinson's disease associated with lower body mass index and heavy leisure-time physical activity. *Eur J Epidemiol.* Apr 2014;29(4):285-292.

57. Dorn HF. Tobacco consumption and mortality from cancer and other diseases. *Public Health Rep.* Jul 1959;74(7):581-593.
58. Hernan MA, Takkouche B, Caamano-Isorna F, Gestal-Otero JJ. A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Ann Neurol.* 2002;52(3):276-284.
59. Morens DM, Grandinetti A, Reed D, White LR, Ross GW. Cigarette smoking and protection from Parkinson's disease: false association or etiologic clue? *Neurology.* 1995;45(6):1041-1051.
60. Fratiglioni L, Wang HX. Smoking and Parkinson's and Alzheimer's disease: review of the epidemiological studies. *Behav Brain Res.* 2000;113(1-2):117-120.
61. Hernan MA, Zhang SM, Rueda-deCastro AM, Colditz GA, Speizer FE, Ascherio A. Cigarette smoking and the incidence of Parkinson's disease in two prospective studies. *Ann Neurol.* Dec 2001;50(6):780-786.
62. Sasco AJ, Paffenbarger RS, Jr. Smoking and Parkinson's disease. *Epidemiology.* Nov 1990;1(6):460-465.
63. Paganini-Hill A. Risk factors for parkinson's disease: the leisure world cohort study. *Neuroepidemiology.* May 2001;20(2):118-124.
64. Wirdefeldt K, Gatz M, Pawitan Y, Pedersen NL. Risk and protective factors for Parkinson's disease: a study in Swedish twins. *Ann Neurol.* Jan 2005;57(1):27-33.
65. Grandinetti A, Morens DM, Reed D, MacEachern D. Prospective study of cigarette smoking and the risk of developing idiopathic Parkinson's disease. *Am J Epidemiol.* Jun 15 1994;139(12):1129-1138.
66. Thacker EL, O'Reilly EJ, Weisskopf MG, et al. Temporal relationship between cigarette smoking and risk of Parkinson disease. *Neurology.* Mar 6 2007;68(10):764-768.
67. Tan LC, Koh WP, Yuan JM, et al. Differential effects of black versus green tea on risk of Parkinson's disease in the Singapore Chinese Health Study. *Am J Epidemiol.* Mar 1 2008;167(5):553-560.
68. Saaksjarvi K, Knekt P, Rissanen H, Laaksonen MA, Reunanen A, Mannisto S. Prospective study of coffee consumption and risk of Parkinson's disease. *Eur J Clin Nutr.* Jul 2008;62(7):908-915.
69. Chen H, Huang X, Guo X, et al. Smoking duration, intensity, and risk of Parkinson disease. *Neurology.* Mar 16 2010;74(11):878-884.
70. Morens DM, Grandinetti A, Davis JW, Ross GW, White LR, Reed D. Evidence against the operation of selective mortality in explaining the association between cigarette smoking and reduced occurrence of idiopathic Parkinson disease. *Am J Epidemiol.* Aug 15 1996;144(4):400-404.
71. Ritz B, Ascherio A, Checkoway H, et al. Pooled analysis of tobacco use and risk of Parkinson disease. *Arch Neurol.* Jul 2007;64(7):990-997.
72. Quik M. Smoking, nicotine and Parkinson's disease. *Trends in neurosciences.* Sep 2004;27(9):561-568.
73. Quik M, Perez XA, Bordia T. Nicotine as a potential neuroprotective agent for Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society.* Jul 2012;27(8):947-957.

74. Benedetti MD, Bower JH, Maraganore DM, et al. Smoking, alcohol, and coffee consumption preceding Parkinson's disease: a case-control study. *Neurology.* Nov 14 2000;55(9):1350-1358.
75. O'Reilly EJ, McCullough ML, Chao A, et al. Smokeless tobacco use and the risk of Parkinson's disease mortality. *Mov Disord.* Oct 2005;20(10):1383-1384.
76. SwedishMatch. http://snus.swedishmatch.com/en/Snus-Academy/The-History-of-Snus/.
77. Lundqvist G, Sandstrom H, Ohman A, Weinehall L. Patterns of tobacco use: a 10-year follow-up study of smoking and snus habits in a middle-aged Swedish population. *Scand J Public Health.* Mar 2009;37(2):161-167.
78. Folkhälsomyndigheten. Snusbruk och hälsorisker. 2017; https://www.folkhalsomyndigheten.se/livsvillkor-levnadsvanor/alkohol-narkotika-dopning-tobak-och-spel-andts/tobak/snusbruk-och-halsorisker/.
79. Foulds J, Ramstrom L, Burke M, Fagerstrom K. Effect of smokeless tobacco (snus) on smoking and public health in Sweden. *Tob Control.* Dec 2003;12(4):349-359.
80. Wilson KM, Markt SC, Fang F, et al. Snus use, smoking and survival among prostate cancer patients. *Int J Cancer.* Dec 15 2016;139(12):2753-2759.
81. Hergens MP, Alfredsson L, Bolinder G, Lambe M, Pershagen G, Ye W. Long-term use of Swedish moist snuff and the risk of myocardial infarction amongst men. *J Intern Med.* Sep 2007;262(3):351-359.
82. Boffetta P, Aagnes B, Weiderpass E, Andersen A. Smokeless tobacco use and risk of cancer of the pancreas and other organs. *Int J Cancer.* May 10 2005;114(6):992-995.
83. Luo J, Ye W, Zendehdel K, et al. Oral use of Swedish moist snuff (snus) and risk for cancer of the mouth, lung, and pancreas in male construction workers: a retrospective cohort study. *Lancet.* Jun 16 2007;369(9578):2015-2020.
84. Gartner CE, Hall WD, Vos T, Bertram MY, Wallace AL, Lim SS. Assessment of Swedish snus for tobacco harm reduction: an epidemiological modelling study. *Lancet.* Jun 16 2007;369(9578):2010-2014.
85. Hansson J, Galanti MR, Hergens MP, et al. Use of snus and acute myocardial infarction: pooled analysis of eight prospective observational studies. *Eur J Epidemiol.* Oct 2012;27(10):771-779.
86. Gartner CE, Hall WD, Chapman S, Freeman B. Should the health community promote smokeless tobacco (snus) as a harm reduction measure? *Plos Med.* Jul 2007;4(7):e185.
87. Shiffman S, Gorsline J, Gorodetzky CW. Efficacy of over-the-counter nicotine patch. *Nicotine Tob Res.* Nov 2002;4(4):477-483.
88. Gray N. Mixed feelings on snus. *Lancet.* Sep 17-23 2005;366(9490):966-967.
89. Ferrence RS, J.; Room, R.; Pope, M. *Nicotine and public health.* Washington: American Public Health Association; 2000.
90. Rodu B. Snus and the risk of cancer of the mouth, lung, and pancreas. *Lancet.* Oct 06 2007;370(9594):1207-1208; author reply 1208.
91. Nyren O, Luo JH, Ye WM, Adami HO, Boffetta P. Snus and the risk of cancer of the mouth, lung, and pancreas - Reply. *Lancet.* Oct 6 2007;370(9594):1208-1208.
92. Rodu B, Jansson C. Smokeless tobacco and oral cancer: a review of the risks and determinants. *Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists.* Sep 01 2004;15(5):252-263.

93. Foulds J, Kozlowski L. Snus - what should the public-health response be? *Lancet.* Jun 16 2007;369(9578):1976-1978.
94. McKee M, Gilmore A. Swedish snus for tobacco harm reduction. *Lancet.* Oct 06 2007;370(9594):1206; author reply 1206-1207.
95. Olanow CW. A radical hypothesis for neurodegeneration. *Trends Neurosci.* Nov 1993;16(11):439-444.
96. Blesa J, Trigo-Damas I, Quiroga-Varela A, Jackson-Lewis VR. Oxidative stress and Parkinson's disease. *Frontiers in neuroanatomy.* 2015;9:91.
97. Beal MF. Mitochondria, oxidative damage, and inflammation in Parkinson's disease. *Ann N Y Acad Sci.* Jun 2003;991:120-131.
98. Greenamyre JT, Hastings TG. Biomedicine. Parkinson's--divergent causes, convergent mechanisms. *Science.* May 21 2004;304(5674):1120-1122.
99. Sies H, Stahl W, Sundquist AR. Antioxidant functions of vitamins. Vitamins E and C, beta-carotene, and other carotenoids. *Ann N Y Acad Sci.* Sep 30 1992;669:7-20.
100. Prasad KN, Cole WC, Hovland AR, et al. Multiple antioxidants in the prevention and treatment of neurodegenerative disease: analysis of biologic rationale. *Curr Opin Neurol.* Dec 1999;12(6):761-770.
101. Casetta I, Govoni V, Granieri E. Oxidative stress, antioxidants and neurodegenerative diseases. *Curr Pharm Des.* 2005;11(16):2033-2052.
102. Ishihara L, Brayne C. A systematic review of nutritional risk factors of Parkinson's disease. *Nutr Res Rev.* Dec 2005;18(2):259-282.
103. Hellenbrand W, Boeing H, Robra BP, et al. Diet and Parkinson's disease. II: A possible role for the past intake of specific nutrients. Results from a self-administered food-frequency questionnaire in a case-control study. *Neurology.* Sep 1996;47(3):644-650.
104. Morens DM, Grandinetti A, Waslien CI, Park CB, Ross GW, White LR. Case-control study of idiopathic Parkinson's disease and dietary vitamin E intake. *Neurology.* May 1996;46(5):1270-1274.
105. de Rijk MC, Breteler MM, den Breeijen JH, et al. Dietary antioxidants and Parkinson disease. The Rotterdam Study. *Arch Neurol.* Jun 1997;54(6):762-765.
106. Anderson C, Checkoway H, Franklin GM, Beresford S, Smith-Weller T, Swanson PD. Dietary factors in Parkinson's disease: the role of food groups and specific foods. *Mov Disord.* Jan 1999;14(1):21-27.
107. Miyake Y, Fukushima W, Tanaka K, et al. Dietary intake of antioxidant vitamins and risk of Parkinson's disease: a case-control study in Japan. *Eur J Neurol.* Jan 2011;18(1):106-113.
108. Johnson CC, Gorell JM, Rybicki BA, Sanders K, Peterson EL. Adult nutrient intake as a risk factor for Parkinson's disease. *Int J Epidemiol.* Dec 1999;28(6):1102-1109.
109. Scheider WL, Hershey LA, Vena JE, Holmlund T, Marshall JR, Freudenheim. Dietary antioxidants and other dietary factors in the etiology of Parkinson's disease. *Mov Disord.* Mar 1997;12(2):190-196.
110. Powers KM, Smith-Weller T, Franklin GM, Longstreth WT, Jr., Swanson PD, Checkoway H. Parkinson's disease risks associated with dietary iron, manganese, and other nutrient intakes. *Neurology.* Jun 10 2003;60(11):1761-1766.

111. Logroscino G, Marder K, Cote L, Tang MX, Shea S, Mayeux R. Dietary lipids and antioxidants in Parkinson's disease: a population-based, case-control study. *Ann Neurol.* Jan 1996;39(1):89-94.
112. Murakami K, Miyake Y, Sasaki S, et al. Dietary intake of folate, vitamin B6, vitamin B12 and riboflavin and risk of Parkinson's disease: a case-control study in Japan. *Br J Nutr.* Sep 2010;104(5):757-764.
113. Gao X, Chen H, Choi HK, Curhan G, Schwarzschild MA, Ascherio A. Diet, urate, and Parkinson's disease risk in men. *Am J Epidemiol.* Apr 1 2008;167(7):831-838.
114. Zhang SM, Hernan MA, Chen H, Spiegelman D, Willett WC, Ascherio A. Intakes of vitamins E and C, carotenoids, vitamin supplements, and PD risk. *Neurology.* Oct 22 2002;59(8):1161-1169.
115. Chen H, Zhang SM, Schwarzschild MA, et al. Folate intake and risk of Parkinson's disease. *Am J Epidemiol.* Aug 15 2004;160(4):368-375.
116. de Lau LM, Koudstaal PJ, Witteman JC, Hofman A, Breteler MM. Dietary folate, vitamin B12, and vitamin B6 and the risk of Parkinson disease. *Neurology.* Jul 25 2006;67(2):315-318.
117. Hughes KC, Gao X, Kim IY, et al. Intake of antioxidant vitamins and risk of Parkinson's disease. *Mov Disord.* Dec 2016;31(12):1909-1914.
118. Etminan M, Gill SS, Samii A. Intake of vitamin E, vitamin C, and carotenoids and the risk of Parkinson's disease: a meta-analysis. *Lancet Neurol.* Jun 2005;4(6):362-365.
119. Takeda A, Nyssen OP, Syed A, Jansen E, Bueno-de-Mesquita B, Gallo V. Vitamin A and carotenoids and the risk of Parkinson's disease: a systematic review and meta-analysis. *Neuroepidemiology.* 2014;42(1):25-38.
120. Wu X, Beecher GR, Holden JM, Haytowitz DB, Gebhardt SE, Prior RL. Lipophilic and hydrophilic antioxidant capacities of common foods in the United States. *J Agric Food Chem.* Jun 16 2004;52(12):4026-4037.
121. Prior RL, Hoang H, Gu L, et al. Assays for hydrophilic and lipophilic antioxidant capacity (oxygen radical absorbance capacity (ORAC(FL))) of plasma and other biological and food samples. *J Agric Food Chem.* May 21 2003;51(11):3273-3279.
122. Sanchez-Moreno C, Cao G, Ou B, Prior RL. Anthocyanin and proanthocyanidin content in selected white and red wines. Oxygen radical absorbance capacity comparison with nontraditional wines obtained from highbush blueberry. *J Agric Food Chem.* Aug 13 2003;51(17):4889-4896.
123. Rautiainen S, Levitan EB, Orsini N, et al. Total antioxidant capacity from diet and risk of myocardial infarction: a prospective cohort of women. *Am J Med.* Oct 2012;125(10):974-980.
124. Rautiainen S, Lindblad BE, Morgenstern R, Wolk A. Total antioxidant capacity of the diet and risk of age-related cataract: a population-based prospective cohort of women. *JAMA Ophthalmol.* Mar 2014;132(3):247-252.
125. Rautiainen S, Larsson S, Virtamo J, Wolk A. Total antioxidant capacity of diet and risk of stroke: a population-based prospective cohort of women. *Stroke.* Feb 2012;43(2):335-340.
126. Nascimento-Souza MA, Paiva PG, Martino HS, Ribeiro AQ. Dietary total antioxidant capacity as a tool in health outcomes in middle-aged and older adults: a systematic review. *Crit Rev Food Sci Nutr.* Sep 19 2016:0.

127. Rautiainen S, Levitan EB, Mittleman MA, Wolk A. Total antioxidant capacity of diet and risk of heart failure: a population-based prospective cohort of women. *Am J Med.* Jun 2013;126(6):494-500.
128. Berkman LF, Kawachi I. *Social Epidemiology.* USA: Oxford University Press; 2000.
129. Breen R, Rottman D. Class Analysis and Class Theory. *Sociology.* Aug 1995;29(3):453-473.
130. Marmot M. *Status Syndrome: How Your Social Standing Directly Affects Your Health and Life Expectancy.* London: Bloomsbury Publishing; 2004.
131. Geyer S, Hemstrom O, Peter R, Vagero D. Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *J Epidemiol Community Health.* Sep 2006;60(9):804-810.
132. Pressley JC, Tang MX, Marder K, Cote LJ, Mayeux R. Disparities in the recording of Parkinson's disease on death certificates. *Movement Disord.* Mar 2005;20(3):315-321.
133. Kotagal V, Bohnen NI, Muller ML, et al. Educational attainment and motor burden in Parkinson's disease. *Mov Disord.* Jul 2015;30(8):1143-1147.
134. Haaxma CA, Borm GF, van der Linden D, Kappelle AC, Bloem BR. Artistic occupations are associated with a reduced risk of Parkinson's disease. *J Neurol.* Sep 2015;262(9):2171-2176.
135. Schernhammer ES, Lassen CF, Kenborg L, Ritz B, Olsen JH, Hansen J. Occupational history of night shift work and Parkinson's disease in Denmark. *Scand J Work Environ Health.* Jul 2015;41(4):377-383.
136. Rostila M, Toivanen S. *Den orättvisa hälsan : om socioekonomiska skillnader i hälsa och livslängd.* Stockholm: Liber; 2012.
137. StatisticsSweden. *Socioekonomisk indelning: SEI = Swedish socioeconomic classification.* Stockholm: Statistics Sweden 1983.
138. StatisticsSweden. Folk-och bostadsräkningen 1990 = population and housing census 1990. 1992; http://www.scb.se/folkochbostadsrakningen1990_sos_/.
139. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
140. Horsfall L, Petersen I, Walters K, Schrag A. Time trends in incidence of Parkinson's disease diagnosis in UK primary care. *J Neurol.* May 2013;260(5):1351-1357.
141. Breen R, Jonsson JO. Explaining change in social fluidity: Educational equalization and educational expansion in twentieth-century Sweden. *Am J Sociol.* May 2007;112(6):1775-1810.
142. Bohlmark A, Lindquist MJ. Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden. *J Labor Econ.* Oct 2006;24(4):879-896.
143. Harkonen J, Bihagen E. Occupational Attainment and Career Progression in Sweden. *Eur Soc.* 2011;13(3):451-479.
144. Karp A, Kareholt I, Qiu C, Bellander T, Winblad B, Fratiglioni L. Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *Am J Epidemiol.* Jan 15 2004;159(2):175-183.
145. StatisticsSweden. Statistics Sweden's History. 2017; http://www.scb.se/en_/about-us/main-activity/statistics-swedens-history/.
146. Socialstyrelsen. The National Board of Health and Welfare - Our history. 2017; http://www.socialstyrelsen.se/english/aboutus/ourhistory.

147. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol.* 2009;24(11):659-667.
148. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health.* 2011;11:450.
149. Socialstyrelsen. *Dödsorsaksstatistik.* Stockholm: Socialstyrelsen;2010.
150. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the third millennium. *Twin Res.* Oct 2002;5(5):427-432.
151. Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res Hum Genet.* Feb 2013;16(1):317-329.
152. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med.* Sep 2002;252(3):184-205.
153. Lagerros YT, Bellocco R, Adami HO, Nyren O. Measures of physical activity and their correlates: the Swedish National March Cohort. *Eur J Epidemiol.* 2009;24(4):161-169.
154. Lagerros YT, Mucci LA, Bellocco R, Nyren O, Balter O, Balter KA. Validity and reliability of self-reported total energy expenditure using a novel instrument. *Eur J Epidemiol.* 2006;21(3):227-236.
155. Stegmayr B, Lundberg V, Asplund K. The events registration and survey procedures in the Northern Sweden MONICA Project. *Scand J Public Health Suppl.* 2003;61:9-17.
156. Svensson AC, Fredlund P, Laflamme L, et al. Cohort profile: The Stockholm Public Health Cohort. *Int J Epidemiol.* Oct 2013;42(5):1263-1272.
157. Alfredsson L, Hammar N, Fransson E, et al. Job strain and major risk factors for coronary heart disease among employed males and females in a Swedish study on work, lipids and fibrinogen. *Scand J Work Environ Health.* Aug 2002;28(4):238-248.
158. Norberg M, Wall S, Boman K, Weinehall L. The Vasterbotten Intervention Programme: background, design and implications. *Glob Health Action.* 2010;3.
159. Medlund P, Cederlof R, Floderus-Myrhed B, Friberg L, Sorensen S. A new Swedish twin registry containing environmental and medical base line data from about 14,000 same-sexed pairs born 1926-58. *Acta Med Scand Suppl.* 1976;600:1-111.
160. Bergström L, Kylberg E, Hagman U, Erikson H, Bruce Å. The food composition database KOST: the National Administration's information system for nutritive values of food. *Vår Föda.* 1991;43:439-447.
161. Khani BR, Ye WM, Terry P, Wolk A. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *Journal of Nutrition.* Jun 2004;134(6):1541-1545.
162. Messerer M, Johansson SE, Wolk A. The validity of questionnaire-based micronutrient intake estimates is increased by including dietary supplement use in Swedish men. *Journal of Nutrition.* Jul 2004;134(7):1800-1805.
163. Rautiainen S, Serafini M, Morgenstern R, Prior RL, Wolk A. The validity and reproducibility of food-frequency questionnaire-based total antioxidant capacity estimates in Swedish women. *Am J Clin Nutr.* May 2008;87(5):1247-1253.

164. StatisticsSweden. Folk-och bostadsräkningen 1980 = population and housing census 1980. 1981; http://www.scb.se/folkochbostadsrakningen1980_sos_/.
165. Feldman AL, Johansson AL, Gatz M, et al. Accuracy and sensitivity of Parkinsonian disorder diagnoses in two Swedish national health registers. *Neuroepidemiology.* 2012;38(3):186-193.
166. Therneau TM, Grambsch PM. *Modelling Survival Data: Extending the Cox Model.* New York: Springer; 2000.
167. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* Sep 1986;7(3):177-188.
168. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemporary clinical trials.* Feb 2007;28(2):105-114.
169. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* Jun 15 2002;21(11):1539-1558.
170. Yang F, Trolle Lagerros Y, Bellocco R, et al. Physical activity and risk of Parkinson's disease in the Swedish National March Cohort. *Brain.* Feb 2015;138(Pt 2):269-275.
171. Yang F, Pedersen NL, Ye W, et al. Moist smokeless tobacco (Snus) use and risk of Parkinson's disease. *Int J Epidemiol.* Dec 10 2016.
172. Grossman J, Mackenzie FJ. The randomized controlled trial: gold standard, or merely standard? *Perspectives in biology and medicine.* Autumn 2005;48(4):516-534.
173. Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol.* Oct 2012;41(5):1480-1489.
174. Berg D, Postuma RB, Adler CH, et al. MDS research criteria for prodromal Parkinson's disease. *Mov Disord.* Oct 2015;30(12):1600-1611.
175. Stern MB, Lang A, Poewe W. Toward a redefinition of Parkinson's disease. *Mov Disord.* Jan 2012;27(1):54-60.
176. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* May 1977;105(5):488-495.
177. Beyer MK, Herlofson K, Arsland D, Larsen JP. Causes of death in a community-based study of Parkinson's disease. *Acta Neurol Scand.* Jan 2001;103(1):7-11.
178. Fall PA, Saleh A, Fredrickson M, Olsson JE, Granerus AK. Survival time, mortality, and cause of death in elderly patients with Parkinson's disease: a 9-year follow-up. *Mov Disord.* Nov 2003;18(11):1312-1316.
179. Paulson GW, Gill WM. Are death certificates reliable to estimate the incidence of Parkinson's disease? *Mov Disord.* Sep 1995;10(5):678.
180. Ritz B, Lee PC, Lassen CF, Arah OA. Parkinson disease and smoking revisited: ease of quitting is an early sign of the disease. *Neurology.* Oct 14 2014;83(16):1396-1402.
181. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* Jan 1999;10(1):37-48.
182. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology.* May 2001;12(3):313-320.
183. Bender R, Lange S. Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology.* Apr 2001;54(4):343-349.
184. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology.* Jan 1990;1(1):43-46.

185. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ.* Jul 14 2001;323(7304):101-105.
186. Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PLoS One.* 2012;7(10):e46042.
187. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* Feb 05 2010;340:c221.

11 APPENDIX

Appendix 1 – Swedish socio-economic classification (SEI)[1]

http://www.scb.se/Grupp/Hitta_statistik/Forsta_Statistik/Klassifikationer/_Dokument/SEI-AGG_Eng.pdf

Manual workers Occupations normally organized by LO (The Swedish Trade Union Confederation)

11 Unskilled employees in goods production
Less than 2 years of post-comprehensive school education

12 Unskilled employees in service production
Less than 2 years of post-comprehensive school education

21 Skilled employees in goods production
2 years or more of post-comprehensive school education

22 Skilled employees in service production
2 years or more of post-comprehensive school education

Non-manual employees Occupations normally organized by TCO (The Swedish Confederation of Professional Employees) or SACO (The Swedish Confederation of Professional Associations)

33 Assistant non-manual employees, lower level
Less than 2 years of post-comprehensive school education

36 Assistant non-manual employees, higher level
Less than 2 but not 3 years of post-comprehensive school education

46 Intermediate non-manual employees
Less than 3 but not 6 years of post-comprehensive school education

56 Professionals and other higher non-manual employees
At least 6 years of post-comprehensive school education

57 Upper-level executives
Upper-level executives in private enterprises or organizations with at least 100 employees or in public service

Employers

60 Self-employed professionals
At least 6 years of post-comprehensive school education

79 Self-employed other than professionals and farmers

89 Farmers

[1] Published in Reports on Statistical Co-ordination 1982:4, Statistics Sweden

Appendix 2 – Occupational and SEI variables in the censuses

SEI FoB60

01 Företagare inom jordbruk, skogsbruk m m
02 Arbetare inom jordbruk, skogsbruk m m
03 Företagare inom industri-, handels-, transport- och serviceyrken
04 Företagare inom fria yrken (läkare, advokater m fl)
05 Företagsledare (anställda)
06 Tjänstemän (arbetsledare, tekniker, kontors- och handelspersonal m fl)
07 Arbetare andra än i grupp 2
08 Anställda inom serviceyrken
09 Militärer
10 Personer med ej identifierbara yrken
11 Studerande (ej förvärvsarb)
12 Övriga ej förvärvsarbetande eller studerande

SEI FoB70

01 Företagare inom jordbruk, skogsbruk m m
02 Anställda inom jordbruk, skogsbruk m m
03 Företagare inom industri-, handels-, transport- och serviceyrken
04 Företagare inom fria yrken (läkare, advokater m fl)
05 Företagsledare (anställda)
06 Anställda inom tekniska, humanistiska, kontorstekniska och kommersiella yrken
07 Anställda inom huvudsakligen industri- och transportyrken
08 Anställda inom vissa serviceyrken
09 Militärer
10 Personer med ej identifierbara yrken

SEI FoB80

11 Ej facklärd i varuproduktion
12 Ej facklärd i tjänsteproduktion
21 Facklärd i varuproduktion
22 Facklärd i tjänsteproduktion
33 Lägre tjänsteman I
36 Lägre tjänsteman II
46 Tjänsteman på mellannivå
56 Högre tjänsteman/ledande befattningar
60 Fria yrkesutövare med akademiska yrken
79 Företagare (exkl. lantbrukare)
89 Lantbrukare
91 Oklassificerade anställda
95 Pensionär
96 Hemarbete
97 Studerande

98 Deltidsarbete
99 Uppgift saknas

SEI FoB90

11 Ej facklärd i varuproduktion
12 Ej facklärd i tjänsteproduktion
21 Facklärd i varuproduktion
22 Facklärd i tjänsteproduktion
33 Lägre tjänsteman I
36 Lägre tjänsteman II
46 Tjänsteman på mellannivå
56 Högre tjänsteman
57 Ledande befattningar
60 Fria yrkesutövare/akademiska yrken
79 Företagare (exkl. lantbrukare)
89 Lantbrukare
91 Oklassificerade anställda
99 Uppgift saknas