

From the **Department of Medical Epidemiology and
Biostatistics**
Karolinska Institutet, Stockholm, Sweden

IF I ONLY HAD A BRAIN

- Epidemiological Studies of
Parkinson's disease

Adina Leah Feldman



**Karolinska
Institutet**

Stockholm 2013

Front and back cover illustration by Rebecca Litzell based on an idea by the author and a sketch by Vilhelmina Ullemar. The silver ribbon is a symbol of Parkinson's disease awareness and the purple ribbon is a symbol of Alzheimer's disease awareness.

Photograph on the back cover by Noemi Naszarkowski.

The comic "How do I love you, thesis?" as well as the drawing on page 95 printed with kind permission from Jorge Cham of "Piled Higher and Deeper" (www.phdcomics.com).

Fig 10.1 (page 82) is printed with permission from www.gapminder.org (free material).

All other diagrams and pictures are by the author, unless otherwise specified.

All previously published papers were reproduced with permission from the publishers.

Published by Karolinska Institutet. Printed by Åtta.45 Tryckeri AB.

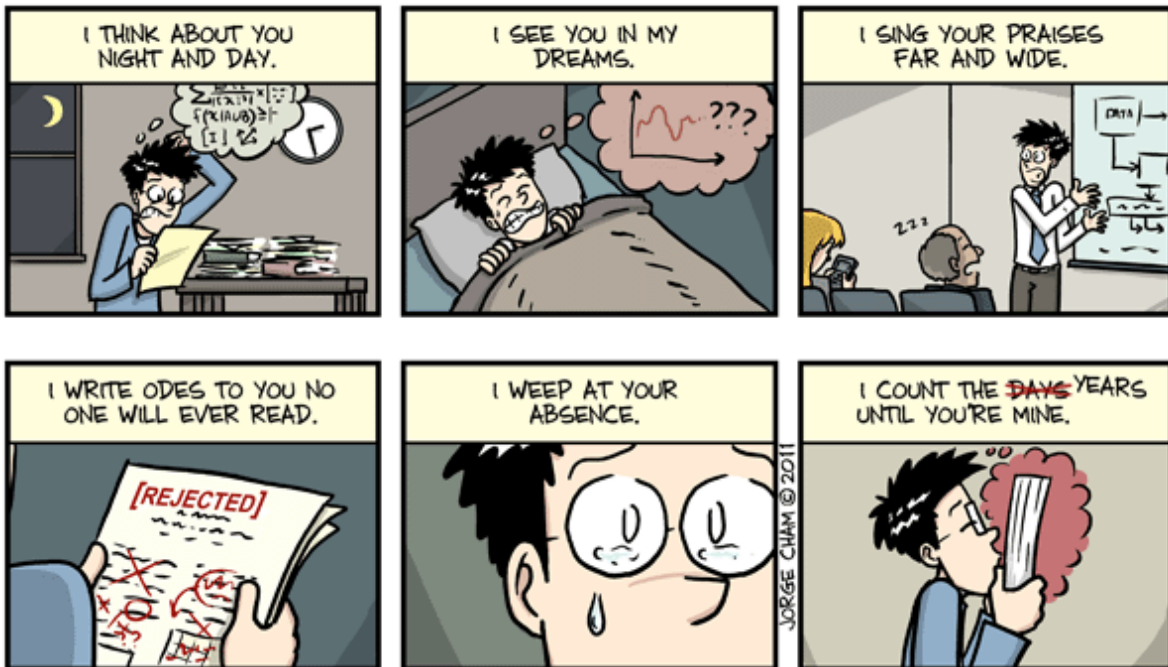
© Adina Leiah Feldman, 2013

ISBN 978-91-7549-396-1

“ Luminous being are we,
not this crude matter ”

-Yoda, *The Empire Strikes Back* (1980)

HOW DO I LOVE YOU, THESIS? LET ME COUNT THE WAYS...



WWW.PHDCOMICS.COM

Source: "Piled Higher and Deeper" by Jorge Cham (www.phdcomics.com).

FOREWORD

“I'd unravel any riddle, For any individ'le, In trouble or in pain (...)

I would dance and be merry, Life would be a ding-a-derry, If I only had a brain”

- From “If I only had a brain” by Harold Arlen & E.Y. Harburg (*The Wizard of Oz*, 1939)

Usually, when the brain is depicted by artists, it's in association with feelings, thinking, dreaming, intelligence and other psychological features. Consider, for instance, the poem “The brain is wider than the sky” by Emily Dickinson or the cartoon “Pinky and the Brain” from the 1990's. The body's movements are perhaps more associated with features like strength and organs like the skeleton and muscles. Of course, the brain's role in movement is just as important. An artistic exception is the song “If I only had a brain” from the Wizard of Oz. It's sung by the Scarecrow who Dorothy meets on the way to Oz, and he's imagining all the things he would do if he had a brain, including not only solving riddles and thinking thoughts, but also dancing – clearly a movement that you need the brain to perform. Perhaps the song writers intended dancing to just be an expression of joy and not a task that you need the brain for, but it doesn't matter.

One needs the brain to move and to think, but having a brain means that you are also vulnerable to diseases of the brain. This thesis is about trying to understand some of the causes of Parkinson's disease, a movement disorder, but it's also about dementia, a disorder of cognition. My reason for wanting to study these diseases is because they are devastatingly debilitating to the patients but also to the families of those affected and there are so many things we don't yet understand about them.

A few words about the structure of the thesis; it is meant to be a complement to the published constituent papers which are reprinted in the back and I have thus attempted to repeat as little as possible of the published texts. If it feels like a daunting task to read the whole thing, don't worry; researchers rarely read entire articles not to mention whole theses. There are lots of diagrams and pictures, I won't mind if you just look at them. I'm just happy that you reached the end of this foreword.

Sine Metu,

A handwritten signature in black ink that reads "Adine L. Feldman". The script is cursive and fluid, with the first letters of each name being capitalized and prominent.

AD 2013
AM 5774
AKI 202
SAS 19693

SAMMANFATTNING PÅ SVENSKA

Kroppens rörelser, som det mina fingrar gör när de skriver den här texten på datorns tangentbord, utförs av muskler styrda av impulser från centrala nervsystemet via perifera nervsystemet. Centrala och perifera nervsystemen utgörs av enorma nätverk av nervceller, där varje enskild cell består av en cellkropp med korta utskott (dendriter) och långa utskott (axon). Dessa celler kommunicerar med varandra genom att olika signalmolekyler (neurotransmittorer) utsöndras från en nervcells utskott och reagerar med specialiserade mottagare (receptorer) på en annan nervcells yta. Högst upp i hjärnstammen, i mellanhjärnan, sitter substantia nigra där specialiserade nervceller producerar neurotransmittorn dopamin, vilken behövs bl.a. för att initiera och styra rörelser i centrala nervsystemet. I Parkinsons sjukdom dör (degenererar) dessa nervceller vilket leder till lokal brist på dopamin med påverkan på rörelseförmågan som en konsekvens. Patienter som drabbas av Parkinsons sjukdom lider bl.a. av skakningar, stelhet och ovanligt långsamma rörelser. Symptomen börjar ofta på en sida av kroppen och förvärras successivt tills det blir ett allvarligt handikappande tillstånd. Förekomsten av Parkinsons sjukdom är vanligast över 65 års ålder och upp till ca. 1% av populationen är drabbad. Man har hittat flera gener som medför ökad risk för Parkinsons sjukdom, men det är bara en mycket liten andel av patienterna vars sjukdom kan förklaras helt av ärftliga faktorer. De flesta fallen orsakas av ett samspel mellan olika ärftliga och miljömässiga riskfaktorer, av vilka de flesta än så länge är okända. Det starkaste dokumenterade sambandet mellan en riskfaktor och Parkinsons sjukdom är en skyddande effekt av rökning, dvs. att de som röker har lägre risk att drabbas jämfört med de som inte röker. Andra riskfaktorer som har studerats och föreslagits ha ett samband med ökad risk av Parkinsons sjukdom är bl.a. bekämpningsmedel inom jordbruk, mjölkprodukter och metaller.

Förutom nedsatt rörelseförmåga utvecklar patienter med Parkinsons sjukdom, oftare än friska personer, även demens, dvs. nedsatt kognitiv förmåga som karakteriseras av problem med framför allt minne, koncentration och uppfattningsförmåga. Det finns två troliga hypoteser som kan förklara varför Parkinsonpatienter drabbas av demens: 1) den skada som drabbar dopaminproducerande nervceller i mellanhjärnan "sprider sig" till andra delar av hjärnan som styr kognitiv funktion, och 2) samma ärftliga och/eller miljömässiga riskfaktorer som orsakar neurodegenerationen i mellanhjärnan leder också till neurodegenerationen i de andra delarna av hjärnan som har att göra med kognitiv funktion.

Det övergripande målet med den här avhandlingen var att öka förståelsen för vad som orsakar Parkinsons sjukdom och varför patienter med Parkinsons sjukdom drabbas av demens.

Mycket epidemiologisk forskning i Sverige och i andra länder bedrivs med data insamlade för andra ändamål än forskning, inklusive administrativa register inom sjukvården. I studie I undersöktes tillförlitligheten av Parkinsondiagnoser från svenska

Patientregistret och Dödsorsaksregistret i epidemiologiska studier. Som "gold standard", dvs. den bästa möjliga tillgängliga diagnosen, användes uppgifter från en storskalig undersökning av över 35000 tvillingar som var 50 år eller äldre under åren 1998-2004. Resultatet visade att bland de som hade Parkinsons sjukdom som huvuddiagnos från minst en sjukhusvistelse registrerad i Patientregistret, kunde 83.0% av fallen bekräftas av gold standard (det positiva prediktiva värdet). Motsvarande siffra för dödsorsaksregistret var 80.0%. De flesta fallen som feldiagnosticerades i registren led av sjukdomar som är snarlika Parkinsons sjukdom och som också drabbar rörelseapparaten (parkinsonismsjukdomar). Av alla Parkinsonfall detekterades 50.0% i patientregistret någon gång under sitt liv (sensitiviteten).

I studie II undersöktes sambandet mellan risken för Parkinsons sjukdom och 14 olika yrkesmässiga exponeringar bland över 14000 manliga tvillingar som deltog i enkätstudier under 1960- och 1970-talet, och sedan följdes upp i svenska hälsoregister under sammanlagt 43 år. Huruvida männen var exponerade mättes genom en matris som uppskattade sannolikheten för att vara utsatt för olika kemiska ämnen i sitt yrke. Resultaten visade en 63% ökad risk (hazardkvot 1.63 (95% konfidensintervall: 1.09-2.44)) för Parkinsons sjukdom bland de som exponerats för ickeorganiskt damm i vilket en stor grupp ämnen ingår, bl.a. vägdamm från trafikerade stadsgator. Inget samband sågs mellan Parkinsons sjukdom och de 13 andra yrkesmässiga exponeringarna som undersöktes, där bl.a. bekämpningsmedel inom jordbruk ingick.

Studie III och IV syftade båda till att undersöka huruvida det är troligt att det finns gemensamma ärftliga och/eller miljömässiga faktorer som delas inom familjer (p.g.a. delad livsstil, etc.) som kan orsaka både Parkinsons sjukdom och demens. Ett sätt att angripa den här forskningsfrågan är att undersöka om dessa sjukdomar ansamlas i familjer, dvs. om de förekommer tillsammans bland förstegradssläktingar (helsyskon eller föräldrar och barn) oftare än vad som skulle vara väntat. Studie III är en systematisk översiktsartikel (review) där resultat från 16 publicerade studier om sambandet mellan risk för demens och att ha familjemedlemmar med Parkinsons sjukdom, eller risk för Parkinsons sjukdom och att ha familjemedlemmar med demens, sammanställdes i en metaanalys. I studie IV undersöktes samma samband bland över 2 miljoner svenska invånare med information om släktskap från flergenerationsregistret och sjukdomsstatus från hälsoregister. Resultaten från studie III visade en 18% högre risk att få demens bland förstegradssläktingar till personer med Parkinsons sjukdom jämfört med de som inte har Parkinsons sjukdom i familjen (hazardkvot 1.18 (95% konfidensintervall: 1.00-1.39)). Resultaten från studie IV visade att bland syskon till personer med Parkinsons sjukdom är det motsvarande 20% högre risk att drabbas av demens (hazardkvot 1.20 (95% konfidensintervall: 1.02-1.41)). Slutsatsen är därmed att det finns en liten ansamling av dessa två sjukdomar i familjer vilket innebär att Parkinsons sjukdom och demens sannolikt delar vissa riskfaktorer inom familjen, men att dessa delade riskfaktorer (dvs. hypotes 2 ovan) inte helt kan förklara varför patienter med Parkinsons sjukdom i så stor utsträckning drabbas av demens själva.

LIST OF PUBLICATIONS

- I. Feldman AL, Johansson ALV, Gatz M, Flensburg M, Petzinger GM, Widner H, Lew MF, Pedersen NL, Wirdefeldt K.
Accuracy and Sensitivity of Parkinsonian Disorder Diagnoses in Two Swedish National Health Registers.
Neuroepidemiology. 2012;38:186-93.
- II. Feldman AL, Johansson ALV, Nise G, Gatz M, Pedersen NL, Wirdefeldt K.
Occupational Exposure in Parkinsonian Disorders: a 43-year Prospective Cohort Study in Men.
Parkinsonism and Related Disorders. 2011;17:677-82.
- III. Feldman AL, Johansson ALV, Lambert PC, Sieurin J, Yang F, Pedersen NL, Wirdefeldt K.
Familial Coaggregation of Alzheimer's Disease and Parkinson's Disease: Systematic Review and Meta-Analysis.
Neuroepidemiology. 2014;42:69-80.
- IV. Feldman AL, Wirdefeldt K, Johansson ALV, Gatz M, Pedersen NL.
Evidence for Modest Familial Co-aggregation Between Dementia and Parkinsonism.
European Journal of Epidemiology. 2013 [Epub ahead of print].

ASSOCIATED PUBLICATION (NOT INCLUDED IN THESIS)

Fang F, Chen H, Feldman AL, Kamel F, Ye W, Wirdefeldt K.

Head Injury and Parkinson's Disease: A Population-based Study.

Movement Disorders. 2012;27:1632-5.

CONTENTS

1	General Introduction to Epidemiology	15
1.1	Causal inference	16
2	Background.....	18
2.1	The Brain	18
2.2	Parkinson's disease	19
2.2.1	Clinical symptoms and diagnosis	19
2.2.2	Neuropathology	20
2.2.3	Parkinsonian disorders.....	21
2.2.4	Epidemiology.....	22
2.2.5	Comorbidity between parkinsonism and dementia.....	23
2.2.6	Familial coaggregation of Parkinson's disease and dementia.....	25
2.3	Familial aggregation and coaggregation of diseases	25
2.3.1	Epidemiological definition.....	26
3	Aims	30
4	Data.....	31
4.1	Register-based data	31
4.1.1	Personal identification numbers.....	32
4.1.2	National Patient Register	32
4.1.3	Cause of Death Register	33
4.1.4	Multigeneration register	34
4.2	The Swedish Twin Registry	37
4.2.1	Questionnaire data	37
4.2.2	Zygoty determination	37
4.2.3	Screening Across the Lifespan of Twins Study	37
4.2.4	SALT/PARKIN	38
5	Methods	39
5.1	The Validation Study (Study I).....	39
5.1.1	In brief.....	39
5.1.2	Study population	39
5.1.3	Gold Standard.....	39
5.1.4	Validation procedure	41
5.2	The Occupational Exposure Study (Study II)	44
5.2.1	In brief.....	44
5.2.2	Participants	44
5.2.3	Outcome	44
5.2.4	Exposure	45
5.2.5	Statistical methods	47
5.3	The Systematic Review and Meta-analysis (Study III)	48
5.3.1	In brief.....	48

5.3.2	Inclusion of studies.....	48
5.3.3	Statistical analysis	52
5.3.4	Additional studies	54
5.4	The Family-based Study (Study IV)	54
5.4.1	In brief	54
5.4.2	Participants.....	54
5.4.3	Outcomes	56
5.4.4	Exposures	56
5.4.5	Statistical methods.....	58
5.4.6	The Twin Study (Study IVtw).....	61
6	Ethical Considerations	62
7	Main Results and Discussion	63
7.1	Validity of register-based Parkinson’s disease and parkinsonian disorder cases (Study I).....	63
7.1.1	Methodological considerations.....	66
7.2	Association between occupational exposure and Parkinson’s disease (Study II)	66
7.2.1	Methodological considerations.....	68
7.3	Familial aggregation of Parkinson’s disease (Study IV, IVtw)	68
7.4	Co-occurrence of parkinsonism and dementia (Study IV, IVtw)	69
7.5	Familial coaggregation of Parkinson’s disease and dementia (study III, IV, IVtw).....	71
7.5.1	Methodological considerations.....	76
8	Conclusions	78
9	Streszczenie po polsku	79
10	Afterword.....	82
10.1	Errata	82
11	Funding sources.....	83
12	Acknowledgements	84
13	References	87

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
CB	Corticobasal degeneration
CDR	Cause of Death Register
CI	Confidence interval
CvP	Cerebrovascular parkinsonism
DAG	Directed acyclical graph
DZ	Dizygotic
ET	Essential tremor
FDR	First-degree relatives
FH	Family history
FN	False negative
FP	False positive
FTLD	Fronto-temporal lobar degeneration
HR	Hazard ratio
ICD	International Classification of Diseases
IR	Incidence rate
IRR	Incidence rate ratio
JEM	Job exposure matrix
LBD	Lewy body dementia
MGR	Multigeneration register
MSA	Multiple systems atrophy
MZ	Monozygotic
NPR	National Patient Register
NPV	Negative predictive value
OPR	Outpatient Register
OR	Odds ratio
PD	Parkinson's disease
PDD	Parkinson's disease dementia
PDS	Parkinsonian disorders
PM	Particulate matter
PPV	Positive predictive value
PSP	Progressive supranuclear palsy
PYR	Person-years
SD	Standard deviation
SE	Standard error
SIR	Standardized incidence ratio
SNCA	α -synuclein gene
STR	Swedish Twin Registry
TN	True negative
TP	True positive
TPR	Total Population Register
VaD	Vascular dementia

1 GENERAL INTRODUCTION TO EPIDEMIOLOGY

There is unfortunately a very common misunderstanding about epidemiology; (Medical) epidemiology *is* the study of diseases or other medical or health-related outcomes on the population level, it is *not* the study of epidemics. In other words, epidemiology includes, but is not limited to, the study of communicable disease that may cause epidemics. Non-medical epidemiology also exists, for instance social epidemiology, which often employs the same methods as medical epidemiology but may study other types of outcomes and exposures. One way of defining epidemiological studies is like this: Studies where the unit of analysis is always a human being (not a mouse or a cell sample, *etc.*) and where the research question concerns something on the population level (not case studies, *etc.*). Note that epidemiology is a quantitative science as opposed to a qualitative science, and that this broad definition includes many studies where the investigators themselves may not consider the study to be an epidemiological study, but perhaps in many cases they should.

When we talk about the epidemiology of an outcome, we usually talk about two main aspects; the distribution of the outcome in the population (descriptive epidemiology) and the relation between the outcome and potential risk factors (analytical epidemiology). Descriptive epidemiology concerns the occurrence of an outcome (*e.g.* a disease), usually in terms of prevalence (the proportion affected in a population at a given time), incidence (the rate of emergence of new cases over time), absolute numbers of affected persons, and distribution of affected persons over population characteristics such as age, sex and race. Analytical epidemiology concerns the association between an outcome and other variables (*e.g.* risk factors), referred to as exposures. In the epidemiological sense, an exposure can be a behavior or characteristic which may seem elusive, such as educational level, neuroticism or a genotype, or it can be a potentially modifiable environmental factor, such as smoking or physical activity. Exposures may also be both causative and protective in relation to the outcome; it all depends on how they are defined. Analytical epidemiology is most often boiled down to some kind of an approximation of a relative risk of outcome in exposed compared to non-exposed persons. However, relative risks do not tell us anything about the overall population distribution of diseases or exposures and therefore both descriptive and analytical epidemiology is necessary in understanding the public health importance of risk factors. The majority of diseases have complex etiologies, meaning that several separate risk factors are needed to cause the disease, and these risk factors may vary between affected individuals. Even in diseases with known necessary risk factors, such as in the case of human papilloma virus infection in cervical cancer, there may still be contributory risk factors important to the etiology of the disease which influence who becomes affected or not, such as concurrent infection or cervical screening participation.

Another important distinction among analytical epidemiological studies is that between observational and interventional (experimental studies). Whether a study is observational or experimental refers to how the distribution of exposure in the study population is determined, whether it's assigned by the study investigators, such as a disease treatment in a randomized control study (interventional), or self-selected, such as smoking (observational), or inherent, such as a genotype (also observational), *etc.* Some exposures are possible to study in both observational and experimental settings, such as physical activity, but many exposures cannot be assigned by investigators to study participants for ethical or practical reasons, such as socio-economic status or head trauma. A third type of study is ecological studies, generally defined as studies that do not include person-level data (micro data), only aggregated data on populations. Finally, a word on retrospective vs. cross-sectional vs. prospective studies. These terms all refer to the timing of exposure in relation to the outcome. In a retrospective study, an exposure that occurred before the outcome is ascertained after the incidence of the outcome, whereas in a prospective study the exposure is ascertained before the incidence of the outcome. In a cross-sectional study the outcome and a current exposure are ascertained at the same time.

1.1 CAUSAL INFERENCE

In experimental studies the investigators can control the parameters of the study, including importantly the timing of the exposure in relation to the outcome. Thus, the investigators can hopefully reasonably well assume that any observed association between the exposure and outcome was caused by their intervention (discounting placebo effects, *etc.*, which lie outside of the scope of this introduction). However, in observational studies there are often issues of uncertainty about the validity of the results due to, among other reasons, the fact that virtually all environmental exposures are non-randomly distributed in the population. Thus, when an association is observed in an observational study we must always ask ourselves: Is it causal? Is it plausible that if we had been able to, for instance, remove the exposure in the exposed group, had the risk of the outcome in that group then changed to become that of the non-exposed group? To complicate issues further, sometimes observed associations are partly or completely spurious, that is, artificially induced by the study design. To be able to tackle and communicate these issues in a systematic way we use the framework of causal inference and its visual representation: directed acyclical graphs (DAG, fig. 1.1).

The generic DAG in fig. 1.1 could be applied to various analyses, for instance, a study of the association between coffee drinking (exposure) and speed of thesis writing (outcome). An arrow between two factors indicates that these two factors are potentially associated and it gives us the assumed direction of association, but does not say anything about whether it is a positive or a negative association or the strength of the association. A confounder is a factor that causes the exposure and the outcome. For instance, time of day or the number of hours of sleep the previous night may cause both coffee drinking and thesis writing speed. Familial confounding is a term used for all factors that can be shared between family members, including both genetic and environmental exposures. Conditioning on a variable (by, for instance, statistical adjustment or stratification/exclusion) which closes the causal path through that factor is represented by drawing a solid box around the variable. A mediator is a factor that lies in the causal path between exposure and outcome. There are practically always potential mediators between the exposure and outcome, but that does not mean that the exposure is not causal. For instance, coffee drinking causes a chain of physiological reactions (mediators) that in turn affect writing speed. A collider is something that is caused by the exposure and the outcome. For instance, stomach aches may be caused by both coffee drinking and thesis writing. Paths through colliders are closed, unless the collider is adjusted for in which case the path is opened, which may cause spurious associations. In every model there are potential unmeasured confounders of importance. In particular, you should always consider potential confounders between a mediator and the outcome if you intend to adjust for that mediator. The reason is that the previously closed path through that unmeasured confounder may then be opened and that could lead to a spurious association. Finally, often we cannot measure a true outcome precisely but rely on some kind of proxy measure that may have some degree of misclassification (the same is true for all other variables in the model). It's important to consider what the impact of an imprecise measure is on the association and whether there may be differential misclassification with regards to the other variables in the model.

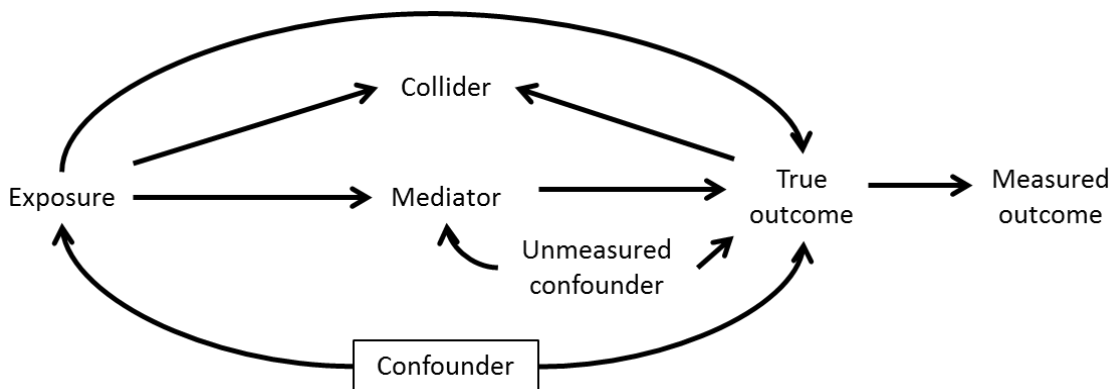


Fig. 1.1. Illustration of concepts in causal inference and its visual representation Directed Acyclical Graphs (DAG).

2 BACKGROUND

2.1 THE BRAIN

Together with the spinal cord, the brain makes up the central nervous system in humans. The brain is divided into two symmetrical hemispheres, the cerebellum and the brain stem (fig 2.1). The cerebral cortex of the hemispheres is further divided into four lobes, each with several specific functional regions. For instance, the primary motor cortex from which movement is controlled is located in the pre-central section of the frontal lobe. Memory is stored and processed in the temporal lobe. Visual input is processed in the occipital lobe and language and sensory input is processed in the parietal lobe.¹ The brain stem connects the spinal cord with the brain and it is divided into the medulla, pons and mecencephalon (midbrain). Throughout the brain and brain stem are several specialized nuclei such as the highly connected basal ganglia which include the subthalamic nucleus, globus pallidus, nucleus accumbens, substantia nigra, and the striatum comprised of the caudate nucleus and putamen. Surrounding the brain and spinal cord and every blood vessel in the central nervous system is a layer of specialized endothelial cells which separate the normal circulation from the cerebrospinal fluid and neuronal tissue. This is called the blood-brain barrier and it protects the central nervous system from microorganisms and molecules that could be harmful. Neurons in the central and peripheral nervous systems have short and/or long outgrowths called dendrites and axons, allowing them to form pathways and networks and communicate across sometimes long distances. The interface between two neurons is called a synapse; there the pre-synaptic neuron releases signal molecules called neurotransmitters which react with receptors on the surface of the post-synaptic neuron causing an action potential which propagates the signal further.

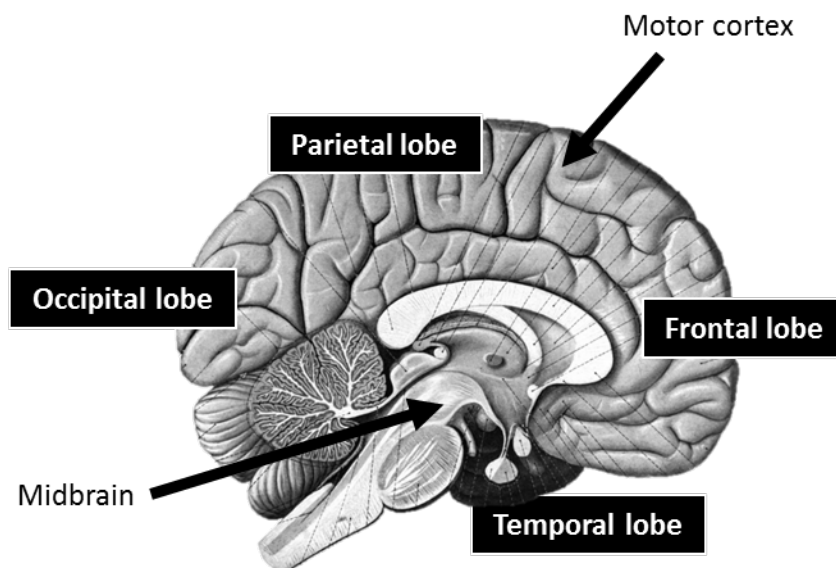


Fig. 2.1. Illustration of a sagittal brain section. Left hemisphere pictured. Original illustration of brain by Johannes Sobotta from Sobotta's Atlas of Human Anatomy (1908).

2.2 PARKINSON'S DISEASE

2.2.1 Clinical symptoms and diagnosis

Parkinson's disease (PD) is a movement disorder named after the James Parkinson who wrote "An Essay on the Shaking Palsy" in 1817, in which he described a group of cases with several distinct symptoms.^{2,3} Today three cardinal features together referred to as parkinsonism are recognized in PD⁴:

1. *Resting tremor*

Otherwise known as involuntary shaking or oscillations of the limbs or other parts of the body when in a resting position, such as sitting down with the hands resting on the lap. This is different from action tremor, which is involuntary shaking during different types of movement.

2. *Rigidity*

Hypertonia or stiffness due to too much muscle tone which makes the joints of the arms and legs harder to move.

3. *Bradykinesia*

Unusually slow movements during, for instance, walking or using the hands and arms.

The onset of these symptoms is typically asymmetrical in PD patients, *i.e.*, one side of the body is affected first. Other symptoms include micrographia (unusually small handwriting), hypophonia (low voice) and less expressive facial features sometimes known as masking. Patients are often divided into two groups based on whether the most dominant symptom is either tremor or bradykinesia. PD is a chronic progressive disorder meaning that symptoms worsen over time often leading to a severely debilitating condition.

The diagnosis of PD is divided into three levels of certainty⁴:

1. *Possible PD*

Presence of tremor or bradykinesia plus rigidity or asymmetrical onset of symptoms. In addition, the patient does not have any features that suggest an alternative cause of the symptoms such as the use of neuroleptic drugs, presence of dementia within one year of onset of motor symptoms or hallucinations or specific atypical motor impairment symptoms such as prominent postural instability or "freezing" early in the disease course.

2. *Probable PD*

After at least three years observation, if at least three of the four cardinal features (including asymmetric onset) are still present and there is a substantial and sustained response to PD medication (*e.g.* Levo-dopa and dopamine agonists).

3. *Definite PD*

If a post-mortem neuropathological examination of a patient with probable PD shows presence of neurodegeneration in the substantia nigra and no signs of any other probable cause of parkinsonism.

2.2.2 Neuropathology

One of the ways that movement is controlled in the central nervous system is through the so called nigro-striatal pathway. The motor cortex sends signals to dopaminergic neurons in the substantia nigra in the midbrain which project into the striatum where the neurotransmitter dopamine causes an either excitatory or inhibitory signal depending on which type of dopamine receptor it binds to. The pathway continues via other brain nuclei back to the motor cortex and to the peripheral nervous system which is in contact with the muscles of the body.

Neurodegeneration is the process of progressive loss of neurons in the central nervous system due to apoptosis. Apoptosis is controlled cell death which has an important function in physiology, but in neurodegeneration it is triggered abnormally. Depending on which part of central nervous system is affected, different functional domains are impaired. Neurodegeneration is coupled with the observation of pathological protein aggregates in or around the apoptotic and pre-apoptotic neurons. Thus, neurodegenerative diseases are sometimes referred to as proteinopathies (protein misfolding diseases) when this particular aspect is studied.

The motor symptoms in PD are caused by the selective degeneration of the dopaminergic neurons in the substantia nigra leading to a depletion of dopamine in the striatum. In PD, the normally pre-synaptic soluble protein α -synuclein aggregates together with other proteins intra-cellularly into characteristic insoluble spherical and thread-like shapes called Lewy bodies and Lewy neurites after their discoverer Fritz Heinrich Lewy.⁵ In PD, Lewy bodies are found extensively in the dopaminergic neurons of the substantia nigra pars compacta, but also in many other parts of the central nervous system. According to the neuropathological staging of PD by Braak *et al.*,⁶ the disease process can be summarized in six stages, where each stage represents further spreading of Lewy body pathology in the central nervous system:

1. Lesions found in the dorsal motor nucleus of the glossopharyngeal and vagal nerves (the medulla of the brain stem)
2. Coeruleus-subcoeruleus complex (midbrain), and commonly also the anterior olfactory nucleus
3. Substantia nigra (midbrain)
4. Anteromedial temporal mesocortex (temporal lobes)
5. High order sensory association areas and prefrontal fields (frontal and parietal lobes)
6. First order sensory association areas, premotor areas, as well as primary sensory and motor fields (frontal and parietal lobes)

PD does not become symptomatic until the pathology reaches stage 3 or 4 (or higher), when the majority of the dopaminergic neurons in the nigro-striatal pathway have been lost.

2.2.3 Parkinsonian disorders

There are several neurological disorders with parkinsonism besides PD. Together with PD they are referred to as parkinsonian disorders and the differential diagnosis of these disorders is not always clear-cut.⁷ Parkinsonian disorders include:

- *Multiple Systems Atrophy (MSA)*⁸
Patients have parkinsonism with often prominent postural instability early in the disease course, cerebellar ataxia which is problems coordinating movements and autonomic problems such as bladder and erectile dysfunction. As in PD, Lewy bodies are found in the substantia nigra making MSA a synucleinopathy like PD. Apart from neurodegeneration of the nigro-striatal pathway there is also degeneration of olivopontocerebellar structures in the pons of the brain stem and cerebellum in MSA.⁹
- *Progressive Supranuclear Palsy (PSP)*¹⁰
The prominent features of PSP is parkinsonism with early onset of postural instability and falls, vertical gaze palsy which is difficulty moving the eyes up and down and dysphagia or problems with swallowing. The prominent protein aggregation in PSP is due to hyperphosphorylated tau protein forming intracellular neurofibrillary tangles, making PSP a tauopathy.
- *Corticobasal degeneration (CB)*¹¹
CB or corticobasal syndrome is characterized by parkinsonism with specific cognitive and behavior disturbances, and it is sometimes grouped together with PSP and fronto-temporal lobar degeneration (FTLD) as the three conditions share features and can be difficult to differentiate. CB is a tauopathy like PSP.
- *Cerebrovascular parkinsonism (CvP)*¹²
CVP or just vascular parkinsonism is not primarily a neurodegenerative disorder but rather parkinsonism caused by white matter lesions due to small vessel disease and/or lacunar infarctions in the brain. Clinically CvP patients often present with parkinsonian symptoms in the lower body first.

Lewy body dementia (LBD) and Parkinson's disease dementia (PDD) are also parkinsonian disorders and these are described in section 2.2.5.

There are also several other movement disorder which are not usually included in the parkinsonian syndrome, such as essential tremor (ET) and Huntington's disease.

2.2.4 Epidemiology

2.2.4.1 Occurrence

The overall prevalence of PD in all age-groups is estimated at 0.3%.¹³ The incidence of PD is very rare before age 50 after which it increases sharply with each year of life, reaching an estimated prevalence of up to 4% over age 80.¹³ The incidence of PD is estimated at about 100/100,000 person-years (PYR) at 60-70 years and up to 400/100,000 PYR over age 80.¹³ Based on a study by Elbaz *et al.*,¹⁴ the lifetime risk of developing PD is estimated at 2.0% and 1.3% for men and women, respectively. The lifetime risk of developing any parkinsonism is slightly higher, estimated at 4.4% for men and 3.7% for women.¹⁴

2.2.4.2 Genetic risk factors

Traditionally, there have been two main ways of quantitatively investigating if a disease with complex etiology has genetic risk factors: through the study of heritability (proportional contribution of heritable factors to the variance of a phenotype) based on differences between concordance in monozygotic (MZ) and dizygotic (DZ) twin pairs and through the study of familial aggregation. Due to the low prevalence of PD, it was for a long time difficult to study heritability as large twin populations are needed. Early studies did not detect any significant heritability leading to a conclusion that heritability for PD was low.¹⁵ However, a more powerful recent twin study in the Swedish Twin Registry (STR) was able to estimate heritability for PD at 34% and heritability for any parkinsonian disorder at 40%.¹⁶ Familial aggregation, defined as relative risk of disease associated with having affected first-degree relatives, has been studied extensively for PD and a meta-analysis based on the 6 most methodologically rigorous published studies estimated this relative risk at 2.9 (95% Confidence Interval (CI): 2.2-3.8).¹⁷ Although familial aggregation of this magnitude is a strong indicator of the presence of heritable genetic risk factors for PD, it could be explained in part by environmental factors shared within families.

A recent review of genome-wide association data for PD reported that 11 loci have so far been linked with PD risk, including loci in the genes encoding α -synuclein (SNCA) and tau protein.¹⁸ In addition, a small proportion of PD patients have a disease caused by autosomal dominant or recessive genetic variants of which some are high penetrant mutations. One such autosomal dominant variant is a duplication or triplication of the SNCA gene causing overabundance of the protein in the brain.¹⁹

2.2.4.3 Environmental and lifestyle risk factors

The best established environmental risk factor for PD is smoking which has a strong protective effect. A meta-analysis of 48 studies estimated the relative risk of PD for ever vs. never smokers at 0.59 (95% CI: 0.54-0.63).²⁰ This effect is consistent in both case-control and prospective cohort studies and there appears to also be evidence for

a dose-response relationship between increasing pack-years of smoking and decreased PD risk, which is one of several indications that the effect of smoking on PD is a true causal effect. There are also plausible biological mechanisms as nicotine has a neuroprotective and stimulating effect on the dopaminergic system.²¹ Caffeine intake has similarly been associated with a reduced risk of PD consistently in both case-control and prospective studies with a meta-estimate based on 13 studies of 0.69 (95% CI: 0.59-0.80) for PD risk among coffee drinkers vs. nondrinkers. An association between PD risk and head trauma was seen in several case-control studies, but was not confirmed in prospective cohort studies indicating that the observed association may have been due to reverse causation.^{22,23} With the possible exception of dairy intake which may be a risk factor for PD,²⁴ no other dietary, lifestyle or environmental risk factors have been consistently associated with PD.

2.2.4.4 Socioeconomic and occupational risk factors

Although the evidence is not consistent, there are indications that higher education may be associated with increased PD risk.²⁵⁻²⁷ For specific occupations the evidence is likewise inconsistent, but nevertheless two types of occupations may be associated with PD risk: occupations associated with high education and high socioeconomic status, *e.g.*, teachers and physicians,^{25,26,28} and occupations that are at high risk of exposure to pesticides, *e.g.* farmers.^{29,30} Specific pesticides have also been linked to PD in both epidemiological and animal studies, in particular the insecticide Rotenone and the herbicide Paraquat.³¹ In animal studies exposure to these pesticides causes selective degeneration of dopaminergic pathways and parkinsonism-like symptoms.³² Both of these pesticides can cross the blood-brain barrier. Manganese exposure which is common in welders can lead to a condition called manganism which shares similarities with parkinsonism. However, there is no evidence that manganese exposure is associated with PD risk.³³

2.2.5 Comorbidity between parkinsonism and dementia

Dementia is a syndrome of cognitive decline that can be caused by several different conditions, of which Alzheimer's disease (AD) is the most common. Alzheimer's disease is named after Alois Alzheimer who published a report about the neuropathology of AD in 1907.³⁴ The features of dementia include decline of cognitive function, *i.e.* memory impairment, difficulty in learning new information, visio-spatial problems and attention problems, and sometimes psychiatric symptoms such as hallucinations. As in PD, the diagnosis of AD is based on clinical symptoms and absence of cause for exclusion such as movement impairment early in the disease course or atypical sudden onset of symptoms.³⁵ AD is a neurodegenerative disease with wide-spread involvement in the central nervous system. The neuropathological process in AD appears to be initiated in and around the hippocampus in the temporal lobe and the primary protein aggregates are extracellular amyloid- β plaques and intra-

cellular neurofibrillar tau tangles. In late stages of AD, macroscopic brain atrophy (loss of cortical matter) is present in addition to the wide-spread abnormal protein aggregation in the brain.³⁶ Other types of non-AD dementia include vascular dementia (VaD), LBD and FTLD. Dementia can also occur secondary to PD as PDD or due to *e.g.* drug or alcohol abuse or a brain tumor. Overall, the lifetime risk of dementia has been suggested to be 10-12%,³⁷ whereas the lifetime risk after age 65 is estimated at about 20%.^{38,39} Based on twin studies, the heritability of AD is estimated at 79%.⁴⁰ Most studies of familial aggregation of AD among first-degree relatives have been case-control studies conducted more than 20 years ago and a meta-estimate of 7 such studies showed a relative risk of 3.5 (95% CI: 2.6-4.6) for AD associated with first-degree family history of AD.⁴¹

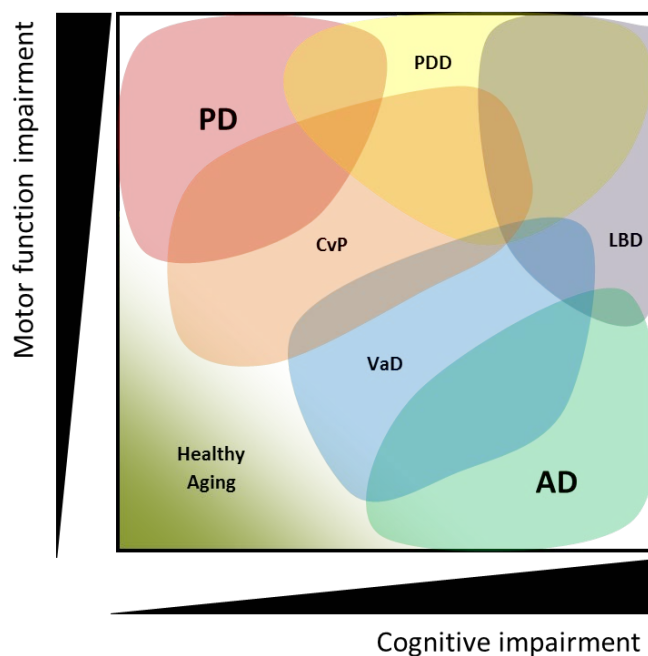


Fig. 2.2. Illustration of clinical overlap between parkinsonian disorders and dementia. AD: Alzheimer’s disease, CvP: Cerebrovascular parkinsonism, LBD: Lewy body dementia, PD: Parkinson’s disease, PDD: Parkinson’s disease dementia.

As described above, dementia is an exclusion criterion for PD and parkinsonism is an exclusion criterion for AD.^{4,35} When the onset of dementia is either before or within one year of parkinsonism, the diagnosis is LBD (sometimes called dementia with Lewy bodies).⁴² When the onset of dementia comes after a diagnosis of PD is set, the diagnosis is PDD. There is an ongoing debate on whether LBD and PDD are the same or two different conditions.⁴³ Neuropathologically, LBD and PDD patients often exhibit characteristics common to both AD and PD such as Lewy bodies and neuritic plaques, although LBD patients may have more white matter atrophy than PDD patients and

some different clinical symptoms such as more often hallucinations.^{44,45} An illustration of the overlap between parkinsonian disorders and dementia is shown in fig. 2.2. For PD patients, the risk of developing dementia (*i.e.*, progressing to PDD) is consistently found to be many times higher than for PD-free persons; the relative risk has been estimated at between 1.7 and 5.9⁴⁶ and 24.5% of PD patients are estimated to have dementia.⁴⁷

2.2.6 Familial coaggregation of Parkinson's disease and dementia

As described above, both PD and AD aggregate in families and commonly co-occur in individuals as comorbidity between parkinsonism and dementia. Thus, it has been hypothesized that these diseases also coaggregate in families, which would indicate that a cause of the co-occurrence of dementia and parkinsonism in individuals could be due to shared familial risk. The hypothesis is that familial (correlated) genetic or environmental risk factors may contribute to the etiology of both PD and dementia. Epidemiological studies that have investigated this hypothesis have been mainly case-control^{48,49} or reconstructed cohort studies⁵⁰⁻⁵⁴ studying the association between AD or dementia and family history of PD and results have been largely inconclusive. In 1991, Van Duijn *et al.* performed a meta-analysis of two early such case-control studies^{48,55} and they found evidence for familial coaggregation of these diseases; pooled odds ratio (OR) 2.4, 95% CI: 1.0-5.8 for risk of AD associated with family history of PD.⁴¹

2.3 FAMILIAL AGGREGATION AND COAGGREGATION OF DISEASES

Why do we study familial aggregation and coaggregation of disease at all? Originally such studies were performed to understand complex diseases, *i.e.* diseases without clear etiology such as mendelian inheritance (based on pedigrees) or infectious agents. Familial aggregation studies share an aim with heritability studies; to understand the general source of causes for complex diseases (or other phenotypes), and these two designs can be seen as complimentary. Generally, we expect complex diseases to aggregate in families to some extent, even if they have low or no heritability, as familial aggregation can be due to other reasons than shared genetic factors. However, the magnitude of aggregation is interesting and just as we study differences in concordance between relatives with varying degree of shared genes, we are interested in the difference in magnitude of aggregation between different types of biological relatives, such as siblings, parents and offspring, MZ and DZ twins and non-biological relatives such as spouses and adopted siblings. An advantage of familial aggregation studies over heritability studies is that the interpretation of risks is somewhat easier to understand and communicate to patients and their families.

The earliest familial aggregation studies appear to have been concerning different types of cancer. Below is a quote from a study about familial aggregation of breast cancer by Penrose *et al.*⁵⁶ published in 1948:

“In spite of evidence brought forward by a great many observers that mammary cancer can frequently be found affecting several of the female members of a family group, doubt still remains as to whether heredity plays any significant aetiological part. This doubt arises because of the great prevalence of the disease in the general population, where it is responsible for nearly 3 percent of all female deaths. With so common a condition, it is difficult to demonstrate convincingly that the occasional familial concentration is not merely the result of a random distribution of cases.”

Today we know that breast cancer is caused by both genetic and environmental factors, partly thanks to these early familial aggregation studies that showed that breast cancer was more common among women with family history of the disease than what would be expected by chance.

Even if two diseases aggregate separately in families and co-occur in individuals to a large degree, we don't necessarily expect them to show familial coaggregation too. There are examples from disease groups where co-occurrence and separate familial aggregation has corresponded to familial coaggregation of diseases, such as in bipolar disorder and schizophrenia,⁵⁷ but also examples of disease groups where that was not case, such as in cryptorchidism, hypospadias and testicular germ cell cancer.⁵⁸ When two complex diseases are found to coaggregate in families this indicates that they share genetic and/or environmental causes, and if the etiology of one of the diseases is better understood than the other than the presence of familial coaggregation may provide clues to the etiology of the other disease. However, if there is considerable co-occurrence of two disease but no or low familial coaggregation, this is an indication that the co-occurrence may be due to a direct causal effect of one disease on the other.

2.3.1 Epidemiological definition

Familial aggregation and coaggregation studies are studies of disease risk associated with family history of the same or another disease. The outcome or family history exposure does not strictly speaking have to be a disease; it can also be a physiological measurement, a behavior, a biomarker, *etc.* However, in this thesis I will discuss familial aggregation and coaggregation studies in the context of chronic diseases that are measured as binary variables. Typically outcomes and exposures in epidemiological studies are ascertained in the same person and thus familial aggregation and coaggregation studies pose a problem since the outcome and exposure must be ascertained in at least two related persons. Firstly, let's define the different types of epidemiological designs possible when studying familial aggregation and coaggregation of diseases:

- *Case-control*
Index cases with a certain outcome are compared for presence of the family history exposure to matched or otherwise sampled (index) controls. The outcome is prevalent or incident, depending on the method of case recruitment. The exposure is defined as family history of disease which can be either prevalent (presence of affected living relatives at the time) or, more commonly, prevalent/past (ever presence of any affected living or dead relatives). Sometimes different exposure levels are ascertained, such as family history in siblings, parents/offspring, second-degree relatives, *etc.*, or none/one/two/more than two affected relatives.
- *Reconstructed cohort*
Relatives of index cases (exposed to family history of the cases' disease) are compared for the presence of prevalent or incident disease to relatives (unexposed to family history) of (index) controls matched to the index cases or otherwise sampled. This is called a reconstructed cohort design because an exposed cohort is compared to an unexposed cohort, but the study persons (the relatives) are not themselves participants in the study. In fact, relatives who are dead at the time of the study may also be included as study persons. The outcome can be defined as prevalent (all relatives affected at the time of the study), but most often the outcome is retrospectively attempted to be ascertained as if it was incident and the cohorts are analyzed in a longitudinal manner (time-to-event analyses). The exposure is prevalent or incident disease, depending on the method of recruitment of the index cases.
- *Matched cohort*
Persons with family history of a disease (exposed cohort) are compared for presence or incidence of the outcome to persons without family history of that disease (unexposed cohort) who are matched to exposed relatives or otherwise sampled. The difference between this and a reconstructed cohort design is that here the relatives who experience the outcome are the participating study persons themselves. This study design does not involve index cases and controls and it's practically a cohort study. Definition of outcome depends on whether the cohort is cross-sectional or followed-up prospectively. As in a case-control study, the exposure can be defined as either prevalent (all living relatives affected at the time) or, more commonly, prevalent/past (all living relatives affected at the time and all dead relatives affected at the time of their death).

- *Cohort*
Study persons in a cohort are each ascertained for presence or absence of family history of disease (the exposure). The relatives that provide the family history exposure can also themselves be included in the cohort as study persons. Outcome and exposure defined as in a matched cohort study. (This is the design in study IV and IVtw).

In a register-based setting, familial aggregation and coaggregation studies can be designed that share characteristics with both case-control and matched cohort studies. These are described further in the section on register-based research in section 4.1.4.1

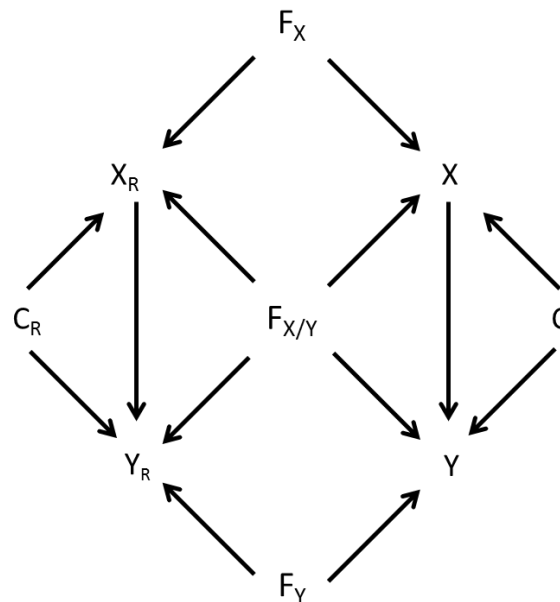


Fig. 2.3. Directed acyclical graph of causal pathways in familial aggregation and coaggregation studies. Subscript *R*'s indicate outcomes in relatives (on the left). Based on model by Hudson, et al.⁵⁹

In fig. 2.3 a DAG of a possible model of the causal pathways in familial aggregation and coaggregation of two diseases is shown. This DAG is based on the model suggested by Hudson *et al.*⁵⁹ This model assumes that the association between exposure disease in relatives (X_R or Y_R) and outcome disease in study persons (X and Y) is caused by both familial factors unique to X and Y (F_X and F_Y) and familial factors shared between the diseases ($F_{X/Y}$). In addition, the study person and relatives have non-shared familial factors that may cause both diseases (C/C_R). One way of looking at familial aggregation studies is that they are the search for F_X or F_Y and familial coaggregation studies are the search for $F_{X/Y}$.

There are three primary problems that may affect familial aggregation and coaggregation studies:

1. *Multiple time-lines*

Regardless of which study design is employed, at least two time-lines are relevant for each observation; one for the outcome and one or several for the exposure. A “regular” epidemiological cohort study with one time-line for the exposure and one for the outcome would be called a time-varying exposure study and would be solved by having study persons potentially contribute both exposed and unexposed time. While this is theoretically possible and perhaps the most epidemiologically sound solution in longitudinal familial aggregation or coaggregation studies too, it’s problematic. Firstly, it is only possible if each study person only has one relative (see section on relative dyads below). Secondly, it can be seen as introducing an additional assumption to the causal model, namely the direct effect of a relative’s disease on the outcome; in fig. 2.3, imagine an additional direct arrow from X_R to Y . In some instances this is a sound assumption, for instance in the study of familial coaggregation of suicidal behavior and family history of psychiatric disease because we expect that having a relative with a psychiatric condition may have a direct effect on one’s own suicidal behavior. However, for neurological diseases this assumption is not supported.

2. *Information bias*

In most studies where the exposure is family history of a disease, ascertainment of either the exposure (case-control, matched cohort and cohort studies) or the outcome (reconstructed cohort studies) is dependent on information from study persons or index cases/controls or their proxies, which are most often spouses or biological relatives. Thus, there is often cause to suspect that reporting of disease history in relatives may be associated with disease status in the study persons.⁶⁰ This source of bias is difficult to handle completely but it may be reduced by, for instance, studying the medical records of all living and deceased relatives or by confirming diagnoses in those that were reported as affected.

3. *Dependency of data*

In all study designs where multiple family members from the same families are included the data can be considered to be non-independent. This fact may bias the standard errors to appear smaller than they are. Luckily, there are several ways to statistically adjust standard errors in clustered data and thus reduce this source of bias.

3 AIMS

The overarching aim of this thesis is to increase the understanding of the etiology of Parkinson's disease and the causes of comorbidity of parkinsonism and dementia.

Specific aims of each study:

- I. To study the validity of register-based Parkinson's disease diagnoses in the Swedish National Patient and Cause of Death Registers with the aim to find the most optimal register-based case definition for use in epidemiological studies.
- II. To explore the association between Parkinson's disease risk and occupational exposure to metal dust, wood dust, animal handling, pesticides, stone and concrete dust, inorganic dust, chrome and nickel dust, quartz dust, organic dust, welding smoke, oil, asbestos, organic solvents and irritating gas.

Papers III, IV and IVtw are studies of the familial coaggregation of Parkinson's disease and dementia that aimed to elucidate whether there is shared familial risk for parkinsonism and dementia, which could be one of the mechanisms explaining comorbidity of the disorders.

The specific aims of each study were:

- III. To systematically review and synthesize published studies on familial coaggregation of Parkinson's disease and dementia.
- IV. To investigate familial coaggregation of Parkinson's disease and dementia in first-degree relatives. To investigate familial aggregation of Parkinson's disease, parkinsonian disorders, Alzheimer's disease and dementia in first-degree relatives.

In addition, results are presented from a study where the aim was to investigate familial aggregation and coaggregation of Parkinson's disease and dementia in monozygotic and dizygotic twins. This study will be referred to as Study IVtw in order to distinguish it from Study IV.

4 DATA

4.1 REGISTER-BASED DATA

In this thesis, register-based epidemiology and register-based studies refer to epidemiological studies that use data from registers at any point in the research process. In Sweden, that definition most likely includes the vast majority of epidemiological research, including studies where the researchers themselves do not consider the study to be register-based.

What is meant by “register”? Usually it refers to administrative registers; that is, depositories of data for some kind of unit (persons, regions, businesses, *etc.*) collected for administrative purposes (including economic or statistical evaluations). These data may be public, meaning that anyone has the right to access them, such as personal tax records in Sweden, or they may be protected, such as hospital discharge records. But the key is that the data are collected primarily for non-research purposes. This type of data is also called “secondary data” or “non-primary data”. For example, when Swedes apply to university, it’s not necessary to state grades from upper secondary school on the university application as all grades are deposited into a central register and then retrieved automatically when the application is processed. Consequently, there is a register of educational achievement available to researchers. However, “register” can also refer to a depository of data collected for research purposes, but not for the purpose of one particular study. For example, the STR collects various data on twins, or the so called national quality registers for health and medical services which collect data for monitoring purposes, but also for research. Register-based studies can be divided into three general categories, although these are not mutually exclusive;

1. *Registers used as sampling frame for primary data collection*

E.g., a case-control study may use the National Patient Register (NPR) to identify cases and the Total Population Register (TPR) to sample controls. After sampling, the cases and controls are contacted by the researchers and primary data is collected.

2. *Registers used for follow-up or for additional data to primary data collection*

E.g., participants of a study may be recruited in the community through ads or personal contacts and then after primary data collection, register-based data may be added through record linkage (study I, II, IVtw).

3. *Exclusively data from registers used without primary data collection*

E.g., several registers are linked together in a relational database and the entire study is conducted on the register-based data without ever contacting the participants or collecting any primary research data (study IV).

4.1.1 Personal identification numbers

In 1947 Sweden established personal records administered by the local parishes for all residents.⁶¹ On the personal record, which stated information such as date and place of birth, sex and address, a personal identification number was assigned. For children aged 15 or younger (*i.e.* born 1932 or later), names and personal identification numbers of living parents were also recorded (this enabled the multigeneration register (MGR) to be established in the 2000's). The personal identification number consists of a six-digit birth date and a four-digit identification number which indicates sex (if the digit in the third position is even (female) or odd (male)) and the county of first residence for persons born until 1990.⁶² The last digit of the four-digit identification number is a check digit derived from the 9 other digits and its calculation is based on the modulus 10-method.

Immigrants to Sweden are assigned a personal identification number if they become permanent residents or intend to stay for more than one year. In special circumstances immigrants may receive a personal identification number that has been previously assigned to a deceased person. This is most likely to happen if date of birth is uncertain in which case January 1st or July 1st is most often chosen, leading to a shortage of unique personal numbers for these birthdates in certain years. At present about 15,000 personal identification numbers have been reused.⁶² A list of reused numbers is kept by Statistics Sweden and data linked to these numbers is usually deleted in population-based register studies to avoid problems related to data from two persons being linked to the same observation.

The personal identification numbers are the basis for all public administration in Sweden and thus all administrative (and other) registers are indexed by this number. This enables registers to be linked to each other and to primary research data.

4.1.2 National Patient Register

Sweden has universal health insurance covering all residents. Patients pay minor out-of-pocket fees for health care visits and hospital stays, but the majority of the cost is carried by the public insurance system which is financed by taxes levied by counties via each resident's home municipality. Although the National Board of Health and Welfare provides national guidelines and policy, health care in Sweden is decentralized; each of the presently 21 counties has wide-ranging autonomy with regards to health care organization and public health policy.

In 1964 an administrative register of hospital discharge records was established at the National Board of Health and Welfare (then *Medicinalstyrelsen*), the NPR.^{63,64} Reported data include personal identification numbers, dates of admission and discharge, type of hospital department, whether the hospital stay was scheduled or not (*i.e.* an emergency), and most importantly, cause of hospitalization operationalized as a primary diagnosis and contributory diagnoses coded according to

the current version at time of discharge of the Swedish translation of the International Classification of Diseases (ICD) as determined by the World Health Organization.⁶⁵ This data is entered more or less raw into the NPR, meaning that excerpts available to researchers frequently include spelling mistakes, date irregularities and missing variables.

At first, 6 counties in and around the Uppsala area in Sweden were included in the NPR.⁶⁴ Between 1964 and 1983, there was a gradual expansion of number of included counties until the register covered about 85% of all discharges. In 1984 it was decided that the register should cover all counties, but plans to expand the register were halted due to concerns about personal integrity and security of the data from the Data Inspection Board. Until the legal and ethical issues could be resolved, hospital discharge data continued to be collected without personal identification numbers. When the legality of the register was finally settled in 1993, most counties were able to retroactively update the register with personal identification numbers, but for some regions of Sweden data from the beginning of the period (1984-1986) was lost. Thus, the register has complete national coverage only from January 1st 1987. In 2001 the next major change of the NPR occurred when outpatient visits from hospital-based clinics began to be reported as well. These data are integrated into the NPR, but are sometimes referred to separately as data from the Outpatient Register (OPR). Primary care data is not included in NPR.

4.1.3 Cause of Death Register

There has been a legal framework to report deaths along with the causes of death in Sweden in various forms since at least 1749.⁶⁶ The history of the organization of Swedish death data reporting is fascinating, but unfortunately not within the scope of this thesis. The Cause of Death Register (CDR),⁶⁷ in the form available for epidemiological research, includes deaths for all Swedish residents, regardless of whether they died in Sweden or abroad, and it is complete since January 1st 1961 (there is also some data digitized for the period 1952-1960). While the NPR contains raw data and diagnoses coded according to the Swedish versions of the ICD, the CDR is cleaned and causes of death are coded according to the international versions of ICD. For researchers who use both registers, the difference in data quality is tangible between NPR and CDR with regards to completeness, spelling mistakes and date errors. The most important variable for the administrative cause of death statistics is the underlying cause of death, defined as the disease or injury that initiated the chain of events that led to death, or the circumstances of the accident or violent act that led to death. The underlying and contributing causes of death are assigned by a physician who may use autopsy data, hospital or primary care records, reports from family members or a visual inspection of the body as basis for the determination. The CDR does contain a variable indicating if an autopsy was performed, but it's not indicated if neuropathological data was collected at autopsy.

4.1.4 Multigeneration register

Information on parents and children from the personal records for Swedish residents that have existed in different forms since 1947, the Population and Housing censuses 1960-1990 and the TPR were developed into the MGR by Statistics Sweden during 2000-2005 and is since then continuously updated.⁶¹ The core of the register contains 5 variables: Index person's personal identification number and the personal identification numbers of their biological mother, biological father, adopted mother and adopted father. If one of the parents is unknown or does not have a personal identification number or if the index person was not adopted, the concerned variable is left blank.

Included index persons are those born in Sweden since 1932 who were residents at any point since 1961. Linkage between index persons and parents is possible if the parents were alive and residents in 1947 or after. Immigrated index persons are linked to parents if they immigrated with their parents before age 18. Thus, even if full siblings immigrated together, they can only be identified as siblings in the MGR if both immigrated before age 18 together with their parents. Paternity is established for men married to the biological mother at time of birth or by acknowledgement.

4.1.4.1 Studies of familial aggregation and coaggregation in MGR

Table 4.1. shows a fictional excerpt of three observations from the MGR. This data can be restructured in many different ways to be used in familial aggregation and coaggregation analyses, for instance in the form of relative dyads; relative pairs with informative order (each pair gives rise to two dyads). Table 4.2 lists all possible relative dyads that can be found in the data from table 4.1., with first-degree relative dyads marked by a star. In this example, the information of three index persons constructed 26 relative dyads, of which 12 were for first-degree relatives. In its basic form, the MGR does not contain data on sex and birthdates, but if we were to merge on data from TPR we can further classify two dyads, Luke-Leia and Leia-Luka as DZ opposite sexed twins.

Table 4.1. Fictional excerpt from the multigeneration register. For illustration purposes the names of included persons are listed, normally excerpts used by researchers would only contain coded index numbers.

Index person	Biological mother	Biological father	Adopted mother	Adopted father
Anakin	Shmi			
Luke	Padme	Anakin	Beru	Owen
Leia	Padme	Anakin	Breha	Bail

Ref: http://starwars.wikia.com/wiki/Skywalker_family

Table 4.2. Fictional example of relative dyads. *indicates first-degree relatives.
Spouses defined as biological parents of at least one child.

Index person	Relative	Relation type
Anakin	Shmi	Offspring-parent*
Anakin	Padme	Spouse-spouse
Anakin	Luke	Parent-offspring*
Anakin	Leia	Parent-offspring*
Shmi	Anakin	Parent-offspring*
Shmi	Luke	Grandparent-grandchild
Shmi	Leia	Grandparent-grandchild
Padme	Anakin	Spouse-spouse
Padme	Luke	Parent-offspring*
Padme	Leia	Parent-offspring*
Luke	Anakin	Offspring-parent*
Luke	Padme	Offspring-parent*
Luke	Shmi	Grandchild-grandparent
Luke	Leia	Full sibling-full sibling (DZ twin-DZ twin)*
Luke	Beru	Adopted child-adopted parent
Luke	Owen	Adopted child-adopted parent
Leia	Anakin	Offspring-parent*
Leia	Padme	Offspring-parent*
Leia	Shmi	Grandchild-grandparent
Leia	Luke	Full sibling-full sibling (DZ twin-DZ twin)*
Leia	Breha	Adopted child-adopted parent
Leia	Bail	Adopted child-adopted parent
Beru	Luke	Adopted parent-adopted child
Owen	Luke	Adopted parent-adopted child
Breha	Leia	Adopted parent-adopted child
Bail	Leia	Adopted parent-adopted child

Using the data in table 4.2 with additional linked data on disease outcomes, we can design several types of familial aggregation studies:

- *Cohort of dyads*

The most efficient way to use the data is to create a cohort of dyads where each observation is a study person (who experiences the outcome) and they are linked to only one relative at a time (who provides the exposure). In this design we can choose whether the exposure should be latent, meaning that we

disregard at what date or age the relative potentially becomes affected by the exposing disease and just treat it as an ever/never exposure, or time-varying, meaning that if the relative becomes affected during the follow-up period, the study person's observation is split into two observations: one for the unexposed time and one for the exposed time. However, there are also drawbacks to using a cohort of dyads. Firstly, as table 4.2 illustrates, the data from even relatively small excerpts of the MGR grows exponentially when dyads are extracted. When a whole population cohort from the MGR is used, there can easily be 10 million or more dyads included and that may be unmanageable to analyze without very powerful computers, especially if it's survival type data with delayed entry. The problem of dependency in data is very important to handle in cohorts of dyads because not only is there family-level clustering, but also individual-level clustering. Lastly, for communication reasons, a cohort of dyads is not very intuitive as each person has potentially multiple observations and the number of observations is proportional to the number of relatives that person has. For instance, if an outcome is associated with family size an incidence rate or prevalence estimate from a cohort of dyads does not have the same interpretation as that estimate from a cohort of individuals.

- *Cohort of individuals*

If we collapse all dyads for a unique individual and instead consider the relatives as a group we can construct a cohort study such as the one described in paragraph 2.3.1 (study IV, IVtw). The advantage is that the complexity and data volume is reduced and, since it's a cohort of actual persons and not dyads, descriptive statistics will now be directly interpretable in relation to the underlying population. There are also no repeated observations for the same individual, but there is still family-level clustering which means that standard errors still need to be adjusted in analyses.

- *Matched case-control/matched cohort study*

An alternative way to reduce the data volume and complexity of a cohort of dyads but still retain as much information as possible is to use the dyads as a sampling frame where case dyads are matched to control dyads (or compared to control dyads otherwise sampled). Controls dyads are thus matched on characteristics of both the index person and the relative, not just the index person. If a density matching method is used, potentially both time-lines can be "matched away", although this is challenging and thus usually only one of the time-lines is used. Since matching is done on both the case (who experiences the outcome) and the exposing relative, the resulting dataset has characteristics of both a case-control dataset and matched cohort dataset and can potentially be analyzed both ways.

4.2 THE SWEDISH TWIN REGISTRY

STR includes twins born in Sweden since 1886, at present more than 95,000 twin pairs.⁶⁸⁻⁷¹ The majority of the twins have at some point contributed questionnaire data and, increasingly, biological data, to the register. In addition to large volumes of data collected by the STR, the overall register and sub-cohorts of the register are routinely linked to Swedish population-based administrative registers.

4.2.1 Questionnaire data

STR was initiated in the 1950's with the original aim to study adverse health effects of smoking.⁶⁹ At that time birth records in Sweden were administered by parishes and so every parish was contacted and asked to provide contact information on same-sexed twins born between 1886 and 1925. Beginning in 1961 all identified same-sexed twin pairs were sent questionnaires and a register was established that included information on all twin pairs where both twins agreed to be included by responding to the questionnaire. This cohort is now known as the "old cohort" of the STR. In 1973 a second cohort consisting of same-sexed twins born 1926-1958, the "middle cohort", was established in a similar way with mailed questionnaires. Collected data included medical history, occupational information, lifestyle habits and demographic data.

4.2.2 Zygosity determination

Zygosity of same-sexed twins was determined using the question: "During childhood, were you and your twin partner as like as 'two peas in a pod' (*lika som bär*) or not more like than siblings in general?" in the questionnaires. If both twins in a pair replied that they were alike, the pair was classified as MZ. For same-sexed twin pairs from the old and middle cohorts whose zygosity was undetermined after the initial questionnaires and who participated in SALT (see below), this additional question was asked: "How often did strangers have difficulty in distinguishing between you and your twin partner when you were children?". If both twins replied that they were often mixed up by strangers, they were classified as MZ. This method of determining zygosity has been shown to have a validity of 99% when compared to determination based on genotyping.⁶⁸

4.2.3 Screening Across the Lifespan of Twins Study

Screening Across the Lifespan of Twins Study (SALT)^{68,69} was an endeavor between 1998 and 2002 to screen all twins born before 1958 (*i.e.* the old and middle cohorts) for various health outcomes. Opposite sexed twins who were not originally included in the middle cohort were also screened. Twins were first contacted with a letter and then through a computer assisted telephone interview. The battery of questions included common complex diseases, health behavior, medications, lifestyle factors, occupation and much more.

4.2.4 SALT/PARKIN

All twins who were 50 years or older at their screening interview in SALT were (or their proxy was) screened specifically for PD.⁷² Besides being asked directly about whether they had a PD diagnosis there were also questions pertaining to the presence of bradykinesia and tremor, such as presence of small handwriting, poor balance, feet “getting stuck to the floor”, shaking arms or legs, difficulty buttoning buttons, shuffling feet or taking tiny steps when walking, soft voice and slow movements.⁷² Furthermore, questions were also asked about the use of PD medication. This screening procedure has been validated and shown to be highly sensitive.⁷³ The screening interview was the first of three phases of the study which will here be referred to as SALT/PARKIN. The second and third phases were a follow-up telephone interview of suspected cases and subsequent clinical workup taking place in 2002–2004. The purpose of the second screening was to exclude from further follow-up those with suspected PD who reported symptoms due to diseases other than PD. The second screening interview was conducted by telephone by research nurses and included additional questions about parkinsonian symptoms and diagnosis, other diseases and medications. If the twins remained suspects after the second screening, they were invited to a clinical workup or, if they had self-reported PD, were worked up by medical record review.

A parallel study to PARKIN used a similar protocol that included a screening phase and a clinical phase to diagnose dementia and Alzheimer’s disease (HARMONY).⁷⁴

5 METHODS

5.1 THE VALIDATION STUDY (STUDY I)

5.1.1 In brief

Parkinson's disease and overall parkinsonian disorder case ascertainment using the inpatient data of the National Patient Register and the Cause of Death Register was validated against cross-sectional screening and clinical data from the Swedish Twin Register (The SALT/PARKIN study). All twins who had participated in the screening and/or clinical work-ups were eligible to be included in the study population as gold standard positive (n=194) or negative cases (n=35,594). Validity of the register-based case ascertainment was measured by the proportion of register-based positive cases that were confirmed by the gold standard (the positive predictive value) and by the proportion of gold standard positive cases detected by the registers (the sensitivity).

5.1.2 Study population

The study population included all twins with known PD or non-PD parkinsonian disorder status after participation SALT/PARKIN (fig. 5.1). N=132 PD cases and n=62 cases of other parkinsonism were confirmed. These n=194 persons constituted the gold standard positive cases, whereas all who screened negative for PD symptoms or were determined to be parkinsonian disorder-free in the clinical work-up were the gold standard negative cases (n=35,594). The final study population included 35,788 twins with available data.

5.1.3 Gold standard

Twins who self-reported a PD diagnosis in the screening were worked up by review of medical records and telephone interview (n=195). Twins who did not report a PD diagnosis but who were suspected cases based on their responses to the screening questions about symptoms or medications were worked up with a clinical examination in their home, an in-person interview and a review of their medical records (n=225). The Unified Parkinson's Disease Rating Scale (UPDRS)⁷⁵ was used to evaluate parkinsonian signs and symptoms. Diagnoses of PD were assigned according to the National Institute of Neurological Disorders and Stroke (NINDS)⁴ criteria by a study physician together with movement disorder specialists, who independently assigned diagnoses, reviewed cases where there was disagreement and agreed upon final consensus diagnoses. Diagnoses of MSA⁷⁶ (n=3), PSP¹⁰ (n=6), LBD⁴² (n=11) and CvP⁷⁷ (n=32) were based on previously published criteria and clinical experience and, where available, computer tomography imaging, in line with clinical practice at that time. In addition, non-PD parkinsonian cases were diagnosed with parkinsonism in dementia (n=5) and parkinsonism of unknown cause (n=5).

Of all parkinsonian disorder cases, 107 (55.2%) were men. The mean (Standard deviation, SD) age at screening was 63.1 (9.7) years for the study population overall and 74.7 (8.8) years for the parkinsonian disorder cases. Mean (SD) age at onset was 68.7 (11.1) years, and mean (SD) disease duration at screening was 6.3 (6.6) years.

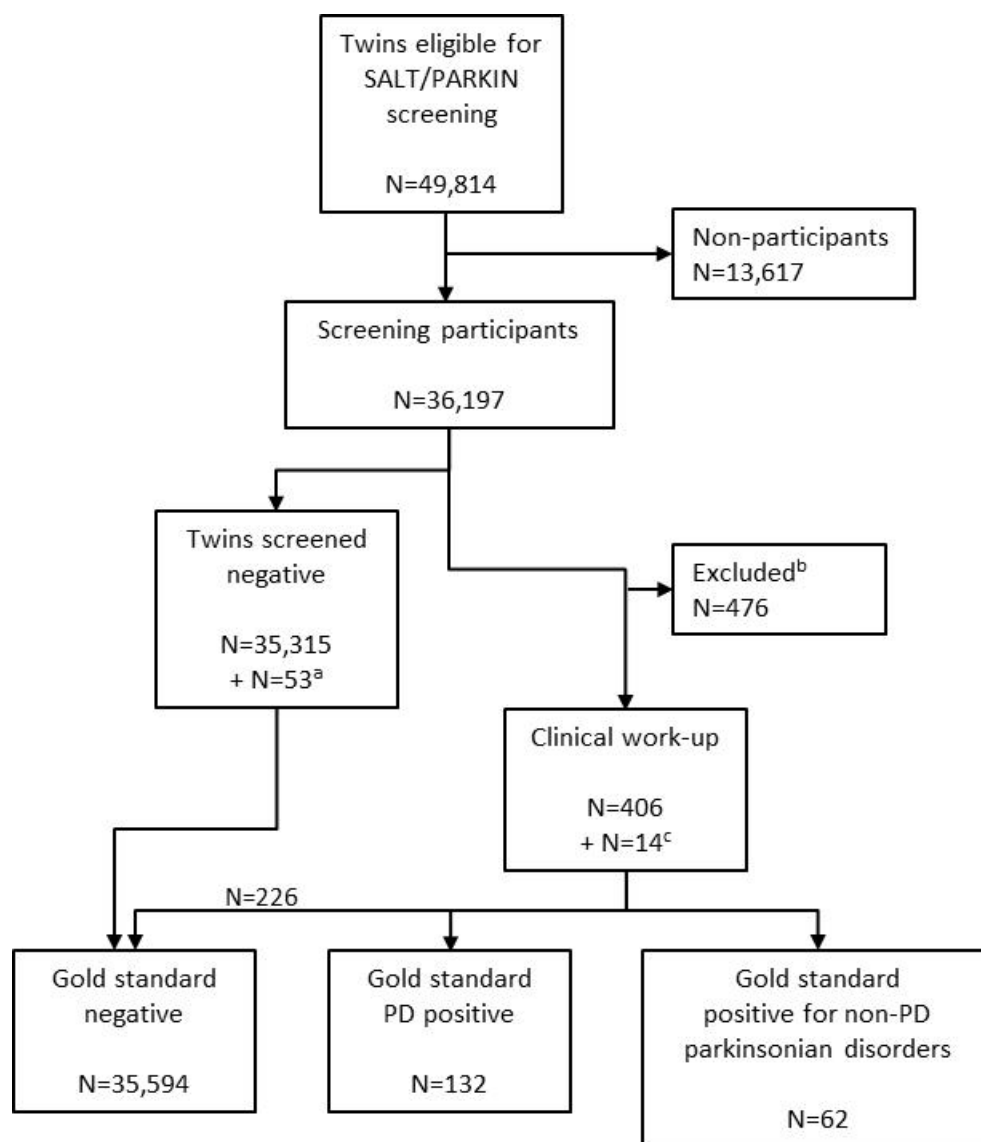


Fig. 5.1. Flow chart depicting the gold standard cohort included in the validation study (Study I). See Wirdefeldt et al.⁷² for a more detailed flow-chart of the PARKIN study. a) N=53 twins were screened for PD in HARMONY and therefore had not been included in the original eligible sample, b) Twins were excluded if lost to follow-up for instance due to refusal to participate in the 2nd screening, c) N=14 twins were non-participants in the screening but still included in the clinical work-up as cotwins. The final cohort with gold standard diagnostic status included n=35,788 participants. PD: Parkinson's disease.

5.1.4 Validation procedure

ICD codes used to identify parkinsonian disorders in study I as well as in studies II, IV and IVtw are listed in table 5.1. To identify cases of parkinsonian disorders in the NPR (inpatient data only) and CDR, we used comprehensive definitions that included all primary and contributory diagnoses in the NPR and all underlying and contributory causes of death in the CDR. To determine the most optimal register-based PD case definition we explored different ways of restricting the definition of PD in the registers, as follows:

1. By excluding register cases who, at any point, had had a non-PD parkinsonian disorder diagnosis (*e.g.* CvP) in the NPR in addition to a PD diagnosis.
2. By using primary diagnoses only in the NPR or underlying causes of death only in the CDR.
3. By using diagnoses (primary or contributory) from neurological, neurosurgical and geriatric departments only in the NPR.

Table 5.1. International Classification of Diseases (ICD) codes used in Studies I, II, IV, IVtw. PD codes in bold. CvP: Cerebrovascular parkinsonism, PD: Parkinson's disease, PSP: Progressive supranuclear palsy.

ICD-code	Diagnosis	Used in Study		
		I	II	IV/IVtw
→1967 (ICD-7)				
350	Paralysis agitans (PD)	X	X	X
1968-1986 (ICD-8)				
342.00	PD	X	X	X
342.08	Other defined parkinsonism	X	X	X
342.09	Unspecified parkinsonism	X	X	X
1987-1996 (ICD-9)				
332.0	PD	X	X	X
332.1	Secondary Parkinsonism		X	
333.0	Other degenerative diseases of the basal ganglia	X	X	X
1997→ (ICD-10)				
F02.3	Dementia in Parkinson's disease			X
G20	PD	X	X	X
G21.0	Malignant neuroleptic syndrome		X	
G21.1	Other drug-induced secondary parkinsonism		X	
G21.2	Secondary parkinsonism due to other external agent		X	
G21.3	Postencephalitic parkinsonism		X	
G21.4	Vascular parkinsonism (<i>i.e.</i> CvP)	X	X	X
G21.8	Other defined secondary parkinsonism	X	X	X
G21.9	Unspecified secondary parkinsonism	X	X	X
G23.1	Progressive supranuclear ophthalmoplegia (<i>i.e.</i> PSP)	X	X	X
G23.2	Striatonigral degeneration	X	X	X
G23.9	Unspecified degenerative disease of basal ganglia	X	X	X
G25.9	Unspecified extrapyramidal and movement disorder	X	X	X
G31.8A	Lewy body disease			X

Validation measures are summarized in fig. 5.2. We were primarily interested in validating register-based case ascertainment in terms of Positive predictive value (PPV) and sensitivity. All validity measures were calculated for two outcomes:

1. *Any parkinsonian disorder*

True positive cases of any parkinsonian disorder did not have to have had the same parkinsonian disorder in the registers and the gold standard, *e.g.* a gold standard CvP case given a PD diagnosis in the registers would still be true positive.

2. *PD*

True positive cases had to have a PD diagnosis in the registers and the gold standard.

Clopper-Pearson exact confidence limits for proportions⁷⁸ were used to construct 95% CI for the validity estimates. Even if cell counts are low and the proportion is close to 0 or 1, these confidence intervals do not cross 0 or 1 which is otherwise a risk with the standard way of calculating CI for proportions.

5.1.4.1 *Positive predictive value*

The PPV was calculated as the proportion (expressed as a percentage) of true positive cases among all register-based positive cases (true positive plus false positive). As the index date for each participant (the date at which their gold standard positive or negative diagnosis status was known), we used the date of initial telephone contact in the SALT screening. For the NPR, all persons in the study population who had been diagnosed with a parkinsonian disorder or PD between the start of follow-up in NPR (1964) and their index date, the PPV represented the proportion whose diagnosis was confirmed by the gold standard. For the CDR, all persons who died within three years of their index date and had a parkinsonian disorder or PD diagnosis listed as a cause of death, the PPV represented the proportion who had also been given a positive diagnosis by the gold standard.

		Gold Standard			
		Positive +	Negative -		
Outcome measure	Positive +	True Positive (TP)	False Positive (FP)	Positive Predictive Value (PPV)	=TP/ (TP+FP)
	Negative -	False Negative (FN)	True Negative (TN)	Negative Predictive Value (NPV)	=TN/ (TN+FN)
		Sensitivity =TP/ (TP+FN)	Specificity =TN/ (TN+FP)		

Fig. 5.2. Validity measures. FN: False negative, FP: False positive, NPV: Negative predictive value, PPV: Positive predictive value, TN: True negative, TP: true positive.

5.1.4.2 Sensitivity

Sensitivity was calculated as the proportion (expressed as a percentage) of true positive cases among all gold standard positive cases (true positive plus false negative). For the NPR, sensitivity represented the proportion of gold standard parkinsonian disorder or PD cases who at any point between 1964 and end of follow-up in 2009 received a register-based diagnosis. For the CDR, sensitivity represented the proportion of all gold standard positive cases who died between their index date and end of follow-up in 2008 and who had a diagnosis listed as a cause of death.

5.2 THE OCCUPATIONAL EXPOSURE STUDY (STUDY II)

5.2.1 In brief

The association between Parkinson's disease or any parkinsonian disorder and 14 occupational risk factors was investigated in a population-based prospective cohort design. Occupational data for Swedish male twins (n=14,169) from questionnaires in 1967 or 1973 was linked to a job exposure matrix to enable grouping of occupations according to level of probability of exposure to 14 occupational exposures. The study population was followed in the National Patient and Cause of Death Registers for incident Parkinson's disease and other parkinsonian disorders from baseline occupational data collection until end of 2009 (up to 43 years). We identified 234 parkinsonian disorder cases including 204 Parkinson's disease cases. Hazard ratios with 95% confidence intervals adjusted for age, smoking and educational status at baseline were used to estimate the relative risk of disease associated with occupational exposure.

5.2.2 Participants

The study population is illustrated in the flow chart in fig. 5.3. All twins in the old (born 1886-1925) and middle (born 1926-1958) cohorts of the STR were eligible to be included. However, due to high levels of missingness on occupational information among women it was decided to include only men. Excluded were also those who were younger than 25 at baseline in the middle cohort (effectively truncating the birth cohort in 1947), those who died before January 1st 1967 in the old cohort and those lost to follow-up from the STR. The final study population included 20,225 men of which 14,167 (70.0%) had complete data on exposure and covariates and could be included in the main analyses. Mean age at baseline in the study population was 44.0 (SD 13.0) and mean follow-up time was 29.1 years (SD 11.1).

5.2.3 Outcome

Incident cases of PD were ascertained by record linkage of the STR with the NPR (only inpatient data) and the CDR. In the present study NPR data was available for 1967-2009 and CDR data was available for 1967-2008 (dates of death for 2009 which was not covered by the CDR were taken from the TPR). Cases were defined as any parkinsonian disorder if there was a relevant diagnosis in the NPR or CDR (table 5.1). Cases were defined as PD if there was a PD diagnosis and no non-PD parkinsonian disorder diagnosis in the NPR or CDR. Please note that this register-based definition of PD is different from that used in studies IV and IVtw. Date of ascertainment was defined as date of first hospital record of a parkinsonian disorder, or, for cases only identified through the CDR, date of death.

We identified 336 incident cases of Parkinsonian disorders, whereof 293 (87.2%) were PD cases. Of 294 cases identified in the NPR, 119 cases (40.5%) were also identified in the CDR. Forty-two cases (12.5%) were identified in death records only. Of 43

parkinsonian disorders cases, 29 (67.4%) had a PD diagnosis, but were classified as non-PD parkinsonian disorder due to an additional non-PD parkinsonian disorder diagnosis. Mean age of ascertainment of outcome for all cases was 75.5 (SD 8.6) years (range 43.4-94.6 years). The overall crude incidence rates per 100,000 PYR were 49.8 for PD and 57.1 all parkinsonian disorders.

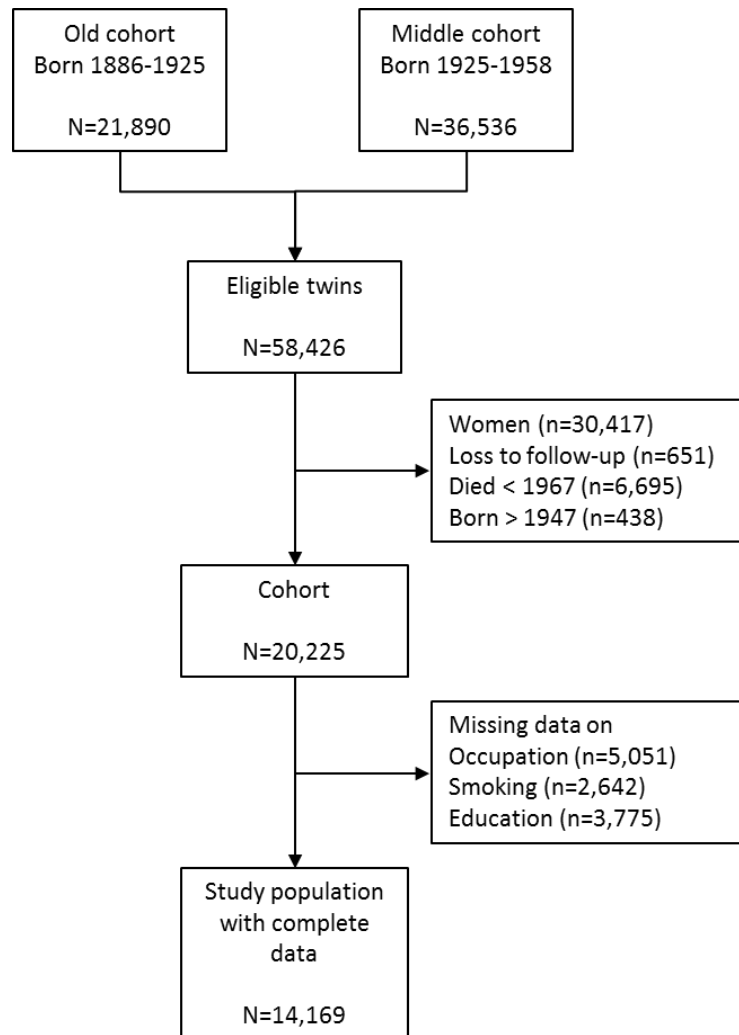


Fig. 5.3. Flow chart depicting inclusion of study persons to the occupational exposure study (Study II).

5.2.4 Exposure

For the old cohort, questions about main occupation in adulthood were included in the 1967 questionnaire when the mean age of the twins was 55.1 (SD 9.6, range 41.0-81.0) years. For the middle cohort, questions about present occupation were included in the 1973 questionnaire when the mean age of the twins was 34.7 (SD 6.5, range 25.0-47.0) years. Among men with occupational data, the crude incidence rate of PD per 100,000 PYR was 49.1, whereas among men without occupational data the rate was 52.2 per 100,000 PYR.

Occupations were coded according to the 3-digit level of occupational codes used in the 1970 census, based on the Nordic Standard Occupational Classification of 1965 (*Nordisk yrkesklassificering*), adapted from the International Standard Classification of Occupations. A job exposure matrix (JEM) was used to assess exposures for each person based on occupation at baseline. The JEM was developed by industrial hygienists at Karolinska Institutet for population-based studies of occupational exposures in the years 1960-1980,^{79,80} corresponding to the average working age of the men in the study population. The JEM assesses probability of occupational exposure to chemical and biological compounds in four classes. Classes 0-3 represent increasing probability of specific exposures within families of occupations, from very low probability (class 0: less than 1/10 of persons within the occupational family exposed), to high probability of exposure (class 3: more than 2/3 exposed). The complete JEM contains information for 29 exposures in 248 occupations. Based on a cut-off level of at least 5% exposed in the study population, 14 of the 29 occupational exposures were selected for analysis.

Education and smoking were included to control for possible confounding. Information was taken from questionnaires in 1961 and 1963 for the old cohort and in 1973 for the middle cohort. Educational level was categorized as having completed compulsory school only or any higher education. In this study population, compulsory school was 6, 7 or 8 years of education depending on birth year. Smoking status was categorized as never, current, or past smoker at the time of the questionnaire.

Of 336 cases, 254 (75.6%) had occupational data and 234 (69.6%) had complete data on occupation, education and smoking status.

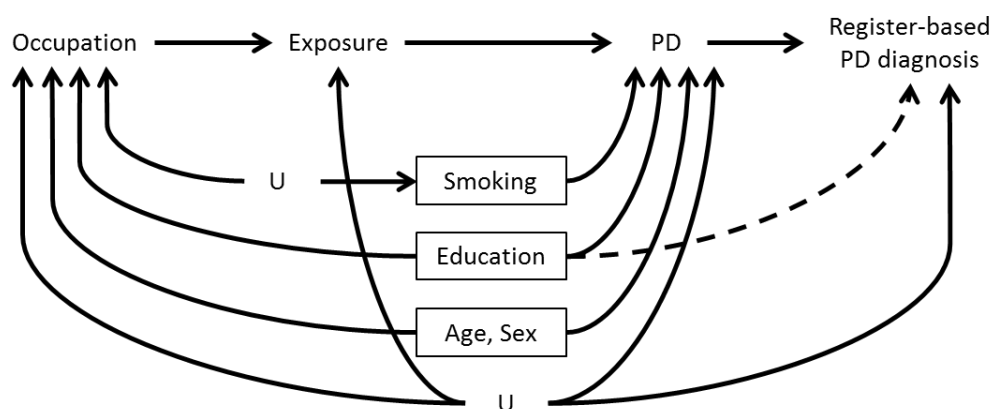


Fig. 5.4. Directed acyclical graph depicting a possible model of the causal pathways between occupational exposure and Parkinson's disease (Study II). PD: Parkinson's disease, U: Unmeasured confounding.

5.2.4.1 Causal pathways

Fig. 5.4 shows a possible model for the causal pathways between PD, occupational exposure, education, smoking, age and sex illustrated as a DAG. The hypothesis is that occupation causes exposure, which may cause PD and subsequent detection of PD in the registers. Smoking is possibly causally (inversely) associated with PD and likely associated with occupation, although not causally. It's more likely that there is a common cause of smoking status and occupation, for instance, socio-economic status (which may be partially captured by education). Education is associated with PD but the mechanism of this association may go through occupation and occupational exposure, in which case education would not be a confounder. However, education may be associated with the detection of PD, although this is not highly likely due to the nature of the public health insurance system in Sweden. We also cannot rule out unmeasured residual confounding. Note that sex is controlled for by stratification.

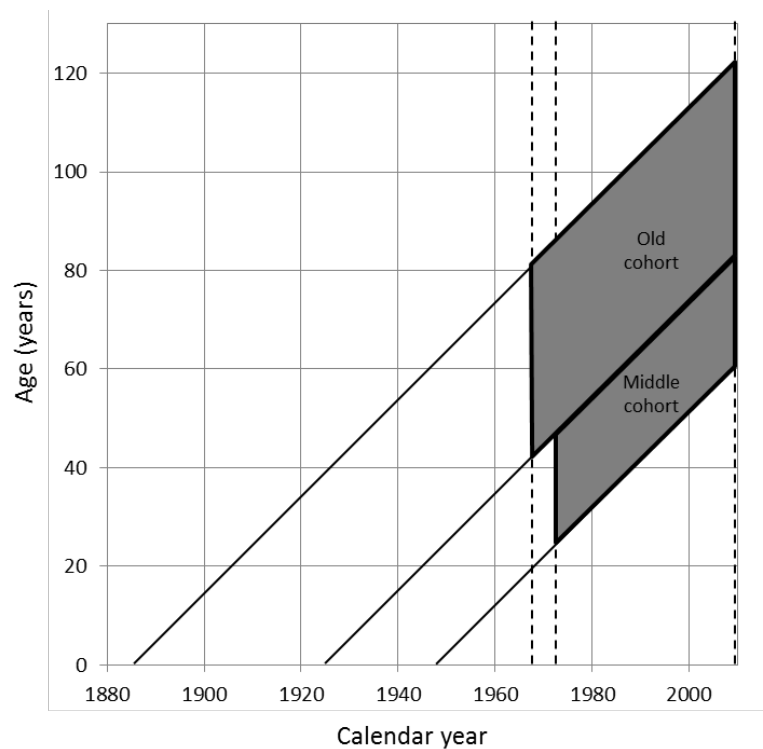


Fig. 5.5. Lexis diagram depicting the follow-up in the occupational exposure study (Study II).

5.2.5 Statistical Methods

The study population was followed for PD and other parkinsonian disorders from baseline collection of occupational data until date of ascertainment of case status, December 31st 2009 (end of follow-up), or date of death, whichever came first. January 1st 1967 was chosen as baseline date for the old cohort and January 1st 1973 was chosen for the middle cohort. Theoretical follow-up of the study population is

illustrated as a Lexis diagram in fig. 5.5. Incidence rates (IR) of PD were calculated as events divided by time-at-risk, reported per 100,000 PYR. The association between occupational exposure and PD was modeled using Cox proportional hazard regression yielding HRs with 95% CI. Age was used as the underlying time scale. Educational level and smoking status were included as covariates in the statistical model.

5.3 THE SYSTEMATIC REVIEW AND META-ANALYSIS (STUDY III)

5.3.1 In brief

Familial coaggregation of Parkinson’s disease and dementia was investigated by conducting a systematic review and meta-analysis. PubMed was searched for studies published through end of October 2012 on dementia risk associated with family history of Parkinson’s disease (or parkinsonism), or Parkinson’s disease risk associated with family history of dementia. Three independent investigators screened publications, extracted data and assigned a quality score based on four indicators to each included study. Hazard ratios or odds ratios from included studies were summarized into meta-estimates using random effects models and illustrated in forest and funnel plots. Heterogeneity and publication bias were tested using the Higgins’ and Egger’s tests, respectively. Of 405 studies found in the initial search, 12 met the inclusion criteria with and an additional 4 studies identified through references and referrals, a total of 16 studies were reviewed, with 14 included in any meta-analysis.

5.3.2 Inclusion of studies

We followed the Meta-Analysis Of Observational Studies in Epidemiology (MOOSE)⁸¹ reporting guidelines which are listed in table 5.2.

Table 5.2. Meta-Analysis Of Observational Studies in Epidemiology (MOOSE)⁸¹ reporting guidelines.

Reporting of background should include
Problem definition
Hypothesis statement
Description of study outcome(s)
Type of exposure or intervention used
Type of study designs used
Study population
Reporting of search strategy should include
Qualifications of searchers
Search strategy, including time period included in the synthesis and keywords
Effort to include all available studies, including contact with authors
Databases and registries searched
Search software used, name and version, including special features used
Use of hand searching

Table 5.2 cont.

List of citations located and those excluded, including justification
Method of addressing articles published in languages other than English
Method of handling abstracts and unpublished studies
Description of any contact with authors
Reporting of methods should include
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
Rationale for the selection and coding of data
Documentation of how data were classified and coded
Assessment of confounding
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results
Assessment of heterogeneity
Description of statistical methods in sufficient detail to be replicated
Provision of appropriate tables and graphics
Reporting of results should include
Graphic summarizing individual study estimates and overall estimate
Table giving descriptive information for each study included
Results of sensitivity testing
Indication of statistical uncertainty of findings
Reporting of discussion should include
Quantitative assessment of bias
Justification for exclusion
Assessment of quality of included studies
Reporting of conclusions should include
Consideration of alternative explanations for observed results
Generalization of the conclusions
Guidelines for future research
Disclosure of funding source

To identify the participating studies we searched (PubMed) using the following search string:

((parkinson*) AND (dement* OR alzheimer*)) AND (risk OR epidemiolog*) AND ((famil* AND (risk OR histor* OR aggreg*)) OR parent* OR sibling* OR twin* OR relative*)

The search was performed on October 30th, 2012 and thus all studies published until that date were eligible to be included. Fig. 5.6 illustrates the inclusion of studies to the qualitative and quantitative review. The initial search returned 405 results, which were exported as

1. A list of titles into a spreadsheet
2. A list of abstracts into a word document

These two documents were screened independently by two investigators in three phases:

1. All titles (n=405) were read to exclude studies which did not appear to fulfill the eligibility criteria (see below).
2. The abstracts of the eligible studies (n=277) were read to exclude studies which did not appear to fulfill the inclusion criteria (see below).
3. The full text of the remaining studies (n=39) was read to ascertain if they fulfilled criteria for inclusion (n=12 with n=4 additional studies identified through references and referrals, totaling n=16 included studies).

Of the 27 studies from the search which were found not to fulfill the inclusion criteria in the last phase, 14 studies were on familial coaggregation between AD and/or PD with another outcome (such as amyotrophic lateral sclerosis (ALS), PDD, LBD, PSP or depression), 3 studies solely concerned the ALS-PD complex of Guam and 10 studies did not include a control group or were missing other crucial data.

Two studies fulfilled the criteria to be included in the systematic review, but were excluded from the meta-analysis; one due to not including enough relevant quantitative data⁸² and one due to unclear results.⁸³ The latter study included Standardized incidence ratios (SIR) for AD risk associated with having a sibling with PD and SIRs for PD risk associated with having a sibling with AD, but it was unclear from the published article which estimate was for which comparison. We were in contact with the authors of the study but did not receive sufficient information to include it. Ultimately, the meta-analyses included estimates from 14 studies.

5.3.2.1 Eligibility criteria

We considered published observational epidemiological studies that met the following criteria:

1. Study design was case-control, cohort, reconstructed cohort or cross-sectional
2. Outcome was disease status (incident or prevalent) of PD, parkinsonism, AD or dementia
3. Exposure was family history of the opposite disease defined as ever/never disease status among one or any first-degree relative (e.g. if exposure was family history of PD, outcome was AD, or vice versa)

Note: We considered studies which included only one type of first-degree relatives to be eligible but in practice all studies included in the meta-analyses assessed family history in both siblings and parents/offspring. Only one study⁸³ had used only siblings but this study had to be excluded from the meta-analysis for other reasons and was only reviewed qualitatively (as discussed above).

4. To be included in the quantitative analysis (meta-analyses) the studies also had to include a crude or adjusted effect estimate of association, such as an OR,

HR/incidence rate ratio (IRR) or SIR, with a standard error (SE) or 95% CI, or enough data allowing a crude effect estimate to be calculated, such as the number of exposed and unexposed among cases and controls in a case-control study.

Following initial independent assessment, the investigators disagreed on the inclusion of two studies.^{84,85} After careful reexamination of the texts of both studies by both investigators it was decided by consensus decision to include one⁸⁵ of the studies and exclude the other.⁸⁴

Relevant data on study design and results was extracted from all included studies (n=16) independently by two investigators. Data on four variables (see below) were combined into a quality score. After extraction results were compared and discussed and a consensus decision was made for each extracted data point and quality score.

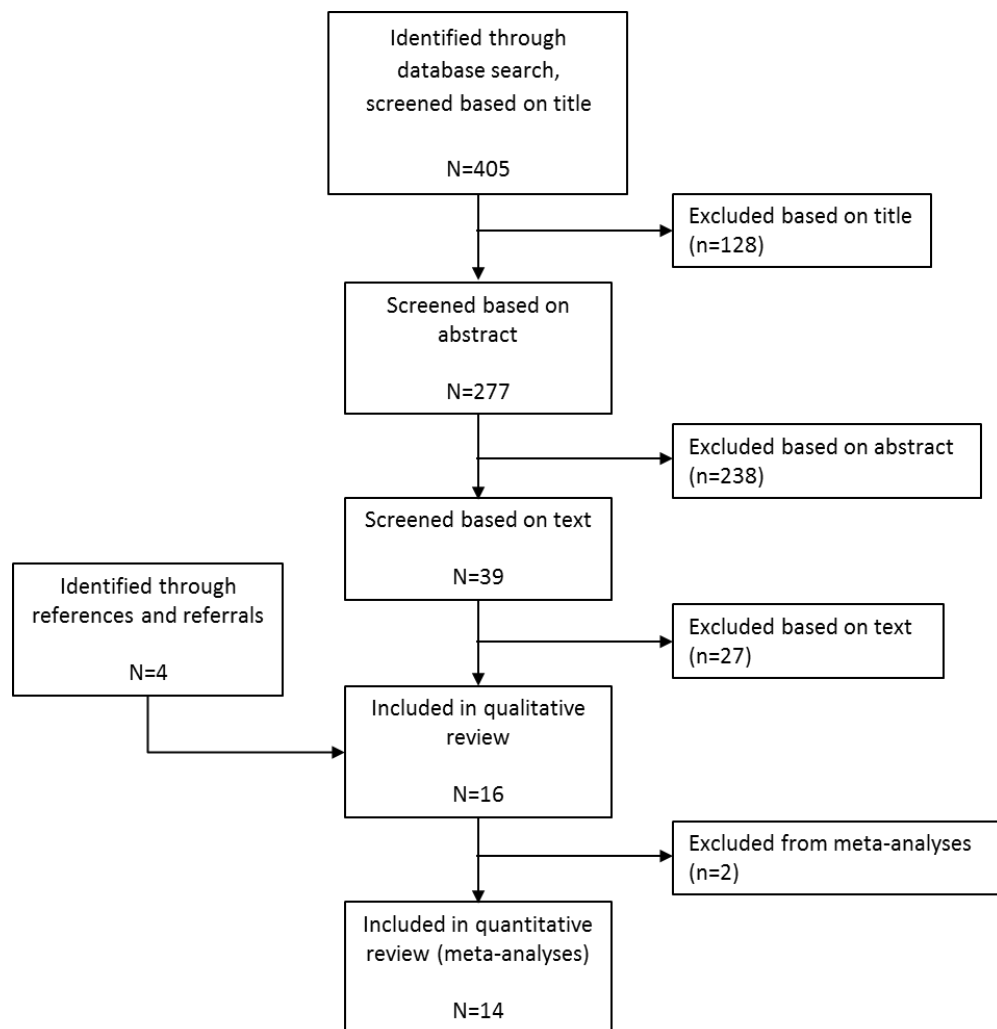


Fig. 5.6. Flow chart depicting the inclusion of studies to the systematic review and meta-analysis (Study III).

5.3.2.2 *Quality score*

We assigned a methodological quality score to all included studies based on four study characteristics. For each characteristic the study was given a score of 1 for more rigorous method, or 0 for less rigorous method. The scores were summed to give a total score ranging from 0 to 4. The quality characteristics were:

1. Setting: If the study participants could be considered to be sampled in a reasonably population-based manner (1) or not (0). For example, using neighborhood controls was considered to be a reasonably population-based method, whereas spouse controls, hospital-based controls or only hospital-based cases with an unspecified catchment area was not.
2. Diagnostic ascertainment in first-degree relatives: If disease status of first-degree relatives was ascertained by family history interviews only (0) or if confirmation by clinical exam or medical record review was sought in some or all instances (1).
3. Study design: If the study was a case-control study that did not enumerate first-degree relatives (0) or a cohort or reconstructed cohort study that did enumerate first-degree relatives (1).
4. Inclusion criteria: If the study excluded (0) or included (1) index cases with co-occurring dementia and parkinsonism, such as in PDD. (See paragraph 5.4.4.1. for a causal inference explanation of why this criterion was included in the quality score.)

5.3.3 **Statistical analysis**

Study specific OR or HR estimates were pooled meta-estimates using the random effects model of DerSimonian and Laird and reported with 95% CI.⁸⁶ In a fixed effects model the assumption is that all differences between included estimates in the meta-analysis are due to sampling variation and not due to variation in the underlying population that each sample is drawn from. A random effects model allows variation not just in sampling but also in the underlying populations. Thus, we chose to use a random effects model because it necessitates fewer assumptions. We did not pool ORs and HRs in the same meta-analysis as some meta-analyses have done before.¹⁷ For studies that reported subgroup specific estimates but no overall estimate, we collapsed the subgroup estimates before entering them into the meta-analysis. This primarily concerned two studies:

1. *Marder et al.*⁵²

This was a reconstructed cohort study which reported relevant HR estimates for the association between risk of AD and family history of PD for two groups:

- I. First-degree relatives of PD cases without dementia vs. first-degree relatives of matched controls
- II. First-degree relatives of PD cases with dementia vs. first-degree relatives of matched controls

These two estimates were collapsed before being entered into the meta-analysis with all other reconstructed cohort studies of dementia risk associated with family history of PD (overall for all first-degree relatives and for the meta-analyses among parent-offspring or siblings only).

2. *Rosen et al.*⁴⁹

This was a case-control study which reported relevant OR estimates for three groups:

- I. AD cases without parkinsonism vs. controls where the exposure was family history of PD
- II. PD cases without dementia vs. controls where the exposure was family history of AD
- III. Cases referred to as AD/PD cases (*i.e.* those with both dementia and parkinsonism including LBD cases, PD cases and cases of AD with extrapyramidal symptoms) vs. controls where the exposure was either family history of PD or family history of AD

The estimates for the association between family history of PD for AD cases vs. controls and for AD/PD cases vs. controls were collapsed before being entered into the meta-analysis with all other case-control studies of dementia risk associated with family history of PD.

The estimates for the association between family history of AD for PD cases vs. controls and for AD/PD cases vs. controls were collapsed before being entered into the meta-analysis with all other case-control studies of PD risk associated with family history of AD or dementia.

Where possible, we calculated pooled estimates for some specific subgroups, such as sibling relationships and parent-offspring relationships. Heterogeneity among studies was evaluated with the Q-statistic from the inverse-variance fixed-effect model and with the Higgins' test yielding the I^2 measure.⁸⁷ Heterogeneity can be said to be a test of whether it is likely that the included studies in the meta-analysis all estimate the same thing. Statistically significant heterogeneity and a high I^2 measure (>25% as a rule of thumb) implies that the studies may not be comparable and that the meta-estimate may not be valid. We constructed forest plots and funnel plots for each meta-analysis and assessed the presence of publication bias using Egger's test for funnel plot asymmetry.⁸⁸ In a funnel plot the estimate of each study is plotted against its inverse log SE. Ideally, studies with high power and thus low SE aggregate at the "tip" of the funnel whereas studies with lower power and thus high SE spread out at the "bottom" of the funnel. If the distribution of estimates appears asymmetric, this may be an indication of publication bias.

5.3.4 Additional studies

After completing this systematic review and meta-analysis, results became known from three additional studies which, if they had been published before October 2012 had been eligible to be included. These studies are Study IV⁸⁹ and IVtw of this thesis, which are described in detail further on in this chapter, and a study by Boot *et al.* published in 2013.⁹⁰ This was a case-control study of LBD and AD cases versus matched controls, where first-degree family history of PD was one of many investigated risk factors for LBD. Estimates (ORs) were reported for the association between family history of PD and LBD in LBD cases vs. age and sex matched controls (conditional logistic model) and in LBD cases vs. unmatched AD cases (unconditional logistic model adjusted for age and sex). There were no estimates reported for LBD plus AD cases vs. controls. As crude numbers were given for both case groups and controls, a crude OR was calculated for the association between AD or LBD and family history of PD before being entered into the meta-analysis of all case-control studies of risk of AD or dementia associated with family history of PD.

5.4 THE FAMILY-BASED STUDY (STUDY IV)

5.4.1 In brief

Familial aggregation and coaggregation of Parkinson's disease and dementia was investigated using the multigeneration register. We constructed two cohorts; a first-degree relative cohort of persons born 1932–1960 (n=2,775,332) and of persons who shared a biological child (spouses) born 1890–1960 (n=4,736,006). The parents of the first-degree relative cohort and study persons in both cohorts were followed up for dementia and parkinsonian disorders between 1969 and 2009 in the National Patient and Cause of Death Registers. We modeled the association between incidence of disease and having at least one relative (any relative, sibling, parent or spouse) affected by the same or opposite disease using Cox proportional hazard regression that estimated hazard ratios with 95% confidence intervals adjusted for age, sex and number of relatives.

5.4.2 Participants

The inclusion of study persons to the cohorts is summarized in the flow chart of fig. 5.7. We used a large database of several linked population-based registers with about 15,000,000 unique individuals including all Swedish residents 1960-1990 and their parents and children if identified in the MGR. In this database we selected study persons for two cohorts:

1. *The first-degree relative cohort*

All individuals born 1932-1960 who were index persons in MGR and could be linked to at least one sibling or a parent (n=2,775,332). In the cohort, 93.5% were linked to both parents, 5.4% were linked to only their mother and 1.1%

were linked to only their father. Via parents we could link full siblings and in this way we identified 745,143 sibling clusters with on average 3.1 (standard deviation (SD) 1.4) siblings per family. The parents of the members of the first-degree relative cohort consisted of 2,749,772 persons (they were not themselves included in the cohort but through them siblings were linked and the exposure status in parent-offspring analyses was ascertained).

2. The spouse cohort

All individuals born 1890-1960 with at least one identified spouse, that is another individual they shared a biological child born between 1932 and 2010 with (n=4,736,006). Most individuals were linked to one spouse but 6.8% were linked to two spouses and 0.8% to more than two spouses.

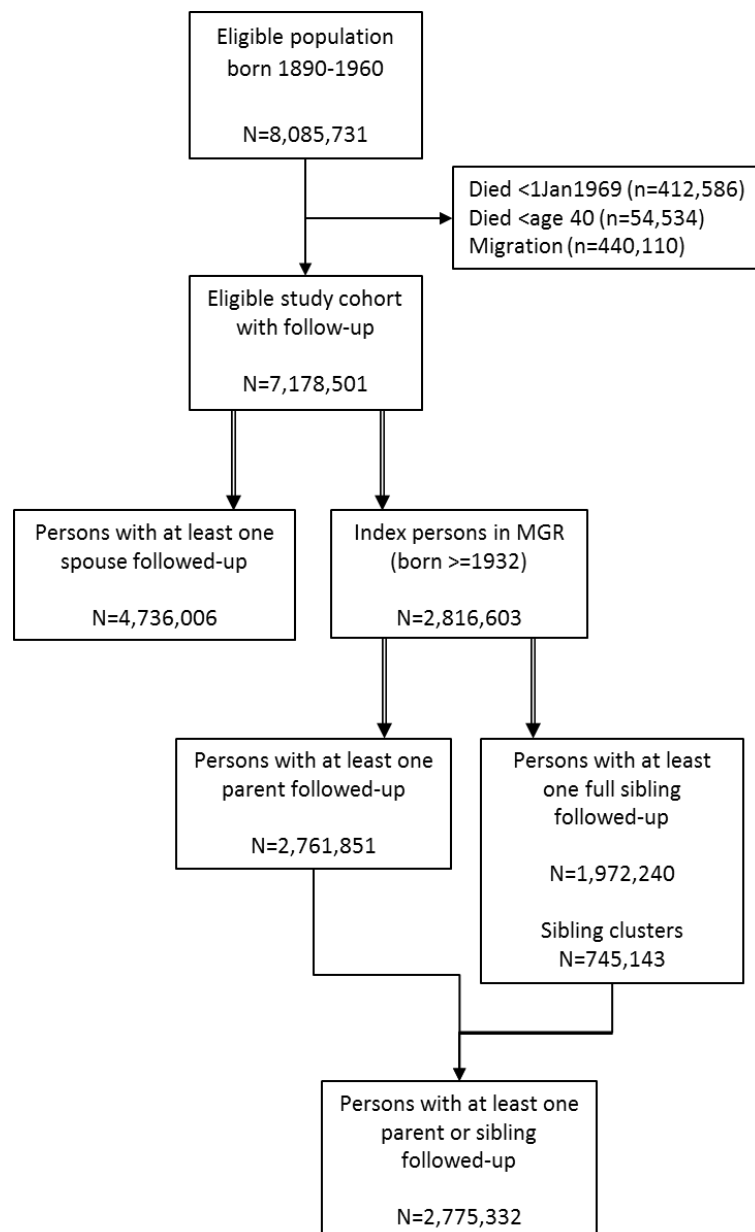


Fig. 5.7. Flow chart depicting the inclusion of study persons to the cohorts in the family-based study (Study IV).

Study persons were eligible to be included in the cohorts if they were alive and residing in Sweden during the follow-up period (1969-2009), and additionally if they had not died or emigrated before their 40th birthday. In order to be able to assess exposure status, the study persons had to have at least one relative who was also alive and residing in Sweden during the follow-up period.

5.4.3 Outcomes

Incident cases of parkinsonian disorders including PD and dementia including AD were ascertained using the outpatient and inpatient NPR and the CDR. See table 5.1 for the ICD codes used to identify cases. We considered four different incident outcomes which were defined like this (when studying a specific outcome, a person was still at risk if any of the other outcomes occurred):

1. Any parkinsonian disorder: Presence of any parkinsonian disorder or PD code in the NPR or CDR
2. PD: Presence of a PD code as a primary diagnosis in NPR or an underlying cause of death in CDR
3. Any dementia: Presence of any dementia or AD code in the NPR or CDR
4. AD: Presence of an AD code as a primary diagnosis in NPR or an underlying cause of death in CDR

First date of hospital admission or outpatient visit with a relevant diagnosis was used as date of ascertainment for cases identified in the NPR whereas date of death was used as date of ascertainment for cases identified in the CDR. An individual could be classified as both a parkinsonian disorder and dementia case.

Age-specific incidence rates of all four outcomes in Study IV (first-degree relative cohort only) and Study IVtw is illustrated in fig. 5.8. In the first-degree relative and spouse cohorts combined we identified 265,620 dementia cases, whereof 92,280 (34.7%) were AD cases, as well as 64,206 parkinsonian disorder cases, whereof 39,504 (61.5%) were PD cases. Among the parents of the first-degree relative cohort, 266,583 (10.4%) had dementia, including 92,659 who had AD and 55,160 (2.2%) had a parkinsonian disorder including 31,967 who had PD.

5.4.4 Exposures

Exposure was dementia, AD, any parkinsonian disorder or PD status (ever or never) in first-degrees relatives (siblings and/or parents) or spouses, depending on the model. In analyses of familial aggregation within diseases the outcome and exposure were the same condition. In analyses of familial coaggregation between diseases, when the outcome was, *e.g.*, dementia or AD, then the exposure was any parkinsonian disorder or PD (or vice versa). We were primarily interested in familial coaggregation of PD with any dementia.

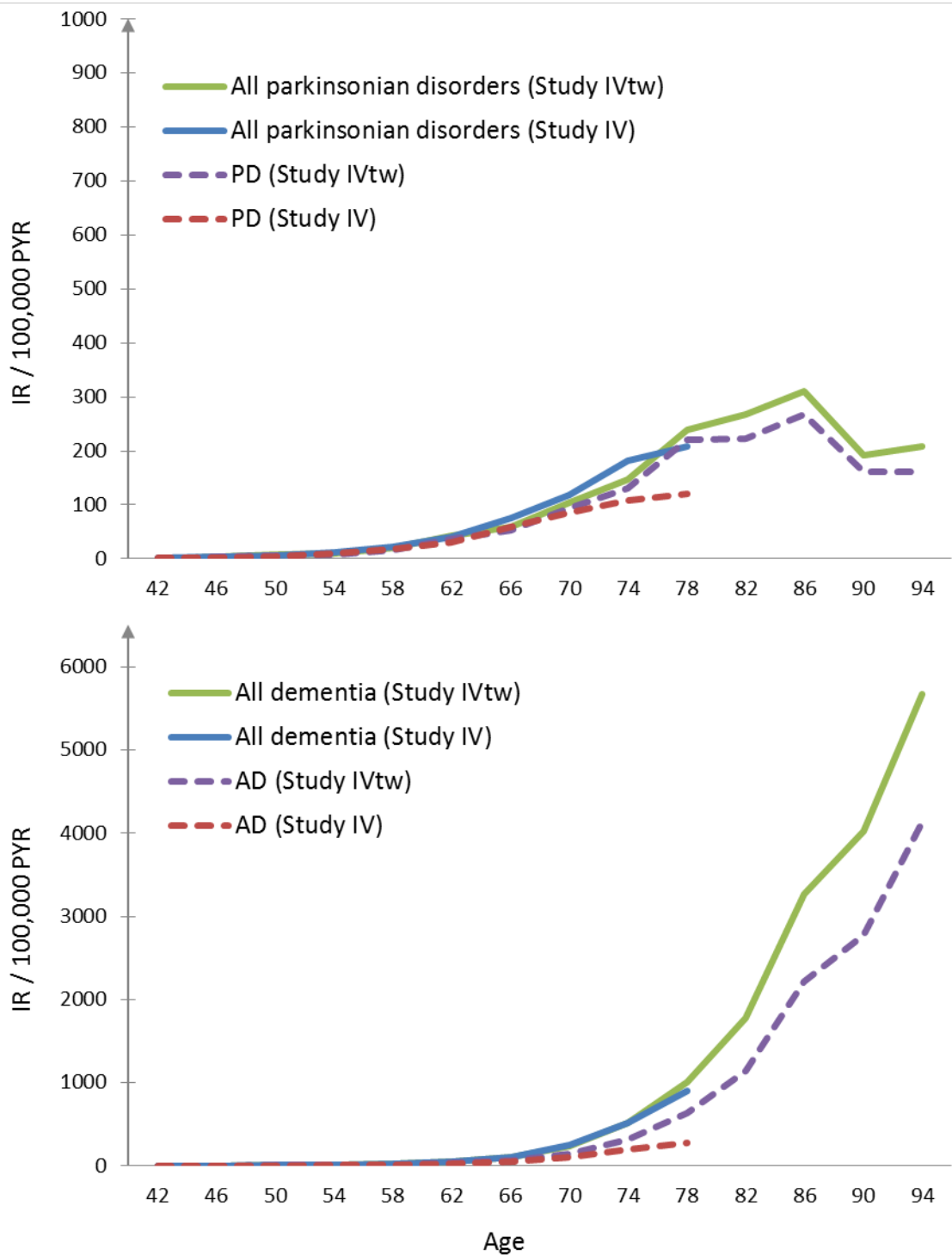


Fig. 5.8. Age-specific incidence rates of parkinsonian disorders and dementia in the family-based study (Study IV) and the twin study (Study IVtw). Note the different scales for IR. AD: Alzheimer's disease, IR: Incidence rate, PD: Parkinson's disease, PYR: Person-years.

5.4.5 Statistical Methods

The follow-up in both cohorts is illustrated in the Lexis diagram in fig. 5.9. The study persons were followed from January 1st 1969, date of immigration or their 40th birthday until December 31st 2009, date of emigration, death or first ascertainment of incident outcome disease, whichever came first. We chose to left-censor the cohort at age 40 to reduce data volume and because there are very few cases of either parkinsonian disorders or dementia diagnosed before age 40. IR per 100,000 PYR were calculated as number of events divided by total person-time at risk. To quantify the association between disease status in relatives and outcome, we estimated HR with 95% CI using Cox proportional hazard regression with age as the underlying time-scale. The models were also adjusted for sex and number of informative relatives as a continuous measure (i.e., number of first-degree relatives, full siblings, parents or spouses). To account for dependencies in the data, we adjusted the standard errors using a robust sandwich estimator clustered on family.

We studied co-occurrence of dementia and parkinsonism by modeling incidence of any dementia or AD in persons who were ever vs. never diagnosed with PD themselves. It was not possible to study co-occurring dementia in PD in time-varying model (PD cases unexposed before they are ascertained with PD) because the time-lag between onset and detection of both PD and dementia in the registers makes it impossible to ascertain which comes first and a large proportion of the comorbid cases were first diagnoses with both conditions on the same date.

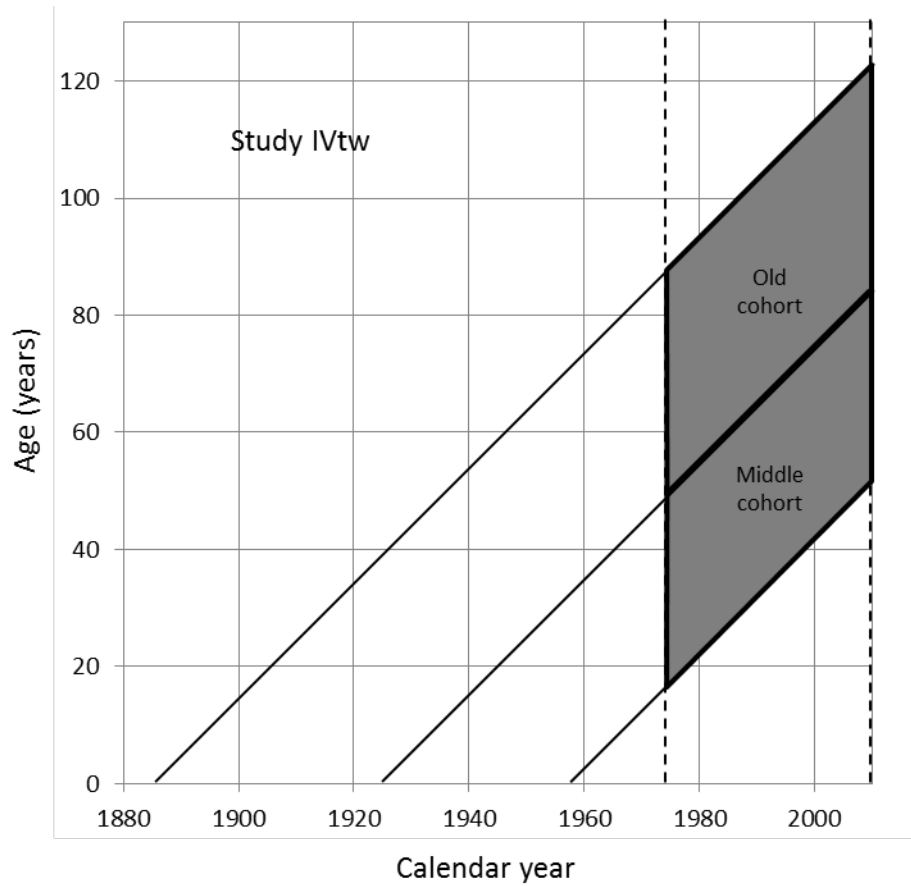
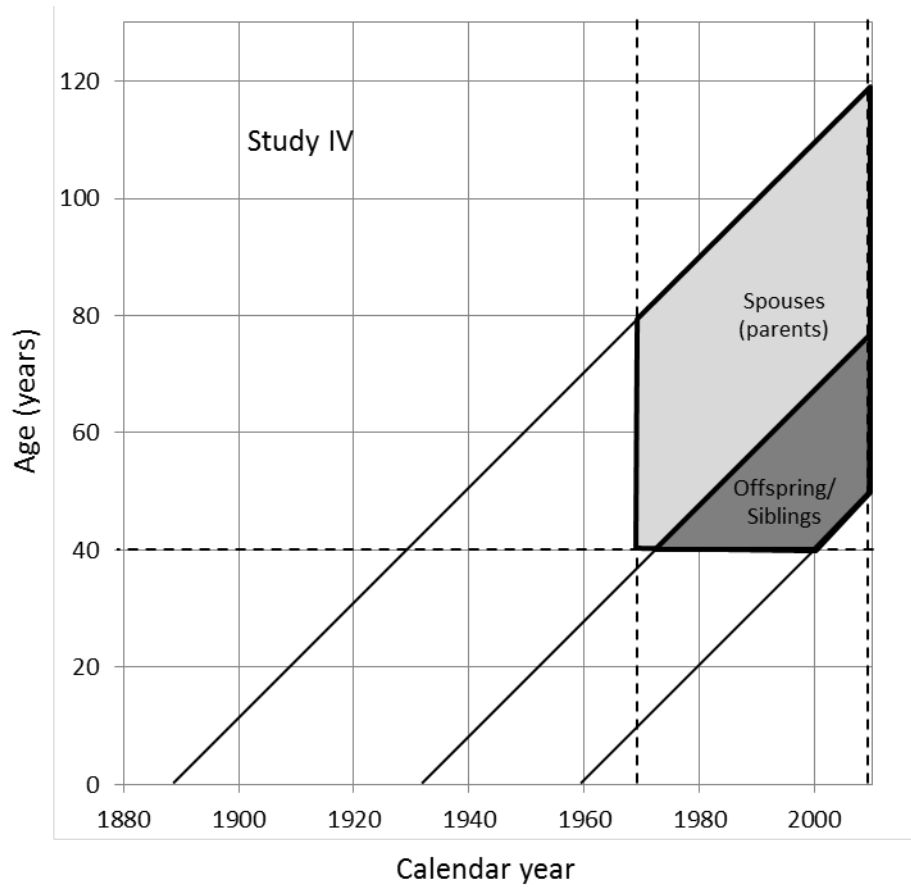


Fig. 5.9. Lexis diagrams illustrating follow-up in the family-based study (Study IV) and the twin study (Study IVtw). Note that cohort overlap in study IV.

5.4.5.1 Causal pathways

A DAG illustrating a possible model of the causal pathways in familial aggregation and coaggregation of PD and dementia is shown in fig. 5.10. This DAG makes several assumptions concerning the association between PD and dementia in individuals:

1. A direct causal pathway from PD to dementia means that in cases with both PD and dementia, such as LBD and PDD, we assume PD is the underlying cause of dementia.
2. Individual (not shared in families) risk factors (C) that are associated causally with both PD and dementia. Please note that C encompasses several potential risk factors.
3. Familial risk factors ($F_{PD/Dementia}$) that are causally associated with both PD and dementia. Please note that $F_{PD/Dementia}$ encompasses several potential risk factors.

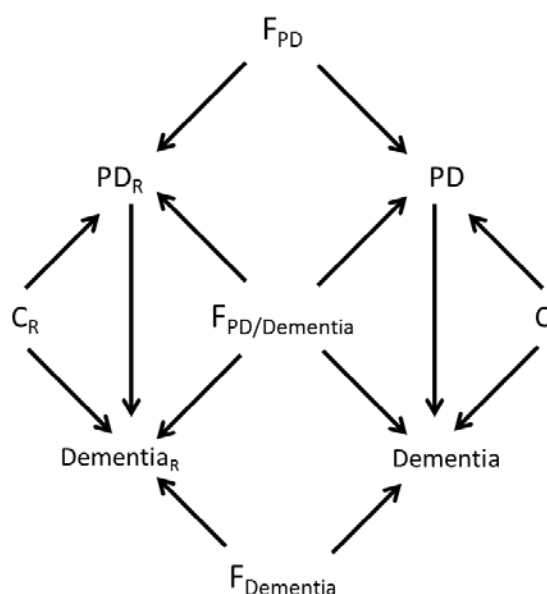


Fig. 5.10. Directed acyclical graph of causal pathways in familial aggregation and coaggregation of Parkinson's disease and dementia (Study III, IV, IVtw),

Please note that the causal pathway between PD_R and Dementia through $F_{PD/Dementia}$ is not the only open pathway and thus if we observe familial coaggregation of PD and dementia, the presence of $F_{PD/Dementia}$ is indicated but not unequivocally proven. We could try to close the pathway through $PD_R \leftarrow F_{PD} \rightarrow PD \rightarrow Dementia$ by, for instance, stratifying (excluding cases with both outcomes), but that action could potentially open the pathway $PD \leftarrow C \rightarrow Dementia$, which is otherwise closed due to the collider at PD. One might also be tempted to exclude relatives with both PD and Dementia,

effectively conditioning on $Dementia_R$, but this action would only open otherwise closed causal pathways through $F_{Dementia}$ and should be avoided.

5.4.6 The Twin Study (Study IVtw)

In parallel with the family-based study we also attempted to investigate familial aggregation and coaggregation of PD and dementia in MZ and DZ twins from the STR. The methods were basically the same as in Study IV with a few differences:

- We included all twins from the old (born 1886-1925) and middle cohorts (born 1926-1958). DZ opposite sexed twins born 1926-1958 were also included. The complete cohort included 76,763 twins in 36,823 complete pairs, of which 9,108 were MZ pairs.
- Follow-up was 1974-2010 for all twins (fig. 5.9).
- As we had available clinical data on parkinsonian disorders and dementia in the STR from the SALT/PARKIN⁷² and SALT/HARMONY⁷⁴ studies, we supplemented the register-based diagnoses with additional clinical diagnoses. For cases only diagnosed clinically, first date of screening contact in SALT was used as the date of ascertainment. We identified 1,000 parkinsonian disorder cases including 882 PD cases and 4474 dementia cases including 2,893 AD cases. 17.3% and 10.9% of respectively the parkinsonian disorder and dementia cases were diagnosed clinically (the rest were register-based cases). Age-specific incidence rates are shown in fig 5.8.

6 ETHICAL CONSIDERATIONS

It is very important to ensure the personal integrity of research participants, whether they consented directly or indirectly to participate in a particular study. When research participants are never contacted directly by researchers or are deceased or unable to consent, an ethical review board may consent to the use of personal data in their place. This was the situation for studies I, II, IV and IVtw in this thesis and the primary ethical issue was data safety to ensure the participants' integrity.

As researchers, we do not have access to personal numbers in our register-based data and reverse identification is strictly prohibited by law. It has never happened that personal data has leaked from a research institution working on registers, but that is not a reason to take the risk lightly. If sensitive data about a particular person become known, even just to the researcher handling the data, it would be a serious ethical breach. An even greater danger is that personal data leaks to, for instance, the press or other unauthorized bodies. This is a real danger that should not be underestimated! The sensitive points in the research process include transferring data and reporting low frequencies in subgroups that may be traced to a specific person.

In theory, the ethical review process is a good format for register-based research, but the problem is that ethical reviews of studies on human subjects treat all studies more or less equally regardless of whether it's an experimental randomized control trial, a purely register-based study (such as study IV) or a study that reuses previously collected data (study I, II, IVtw). For instance, one of the questions on the ethical review application concerns potential harms and benefits to participants. For register-based studies, answering this question becomes a rhetorical exercise. It is my personal opinion that register-based research of the kind that only uses registers and no researcher-generated primary data should not be treated the same way as clinical studies in the ethical review process as the considerations and risk/benefit balance becomes artificial. The key is that participants are not strictly participants. Here the primary consideration should be about data safety, how leaking and safety breaches are prevented, which depends on, for instance, the competence level of those who have access to the data. You need a certificate to conduct animal testing but you don't need any formal training to conduct research on sensitive personal data. Also, what data is being merged should be considered as some data is more vulnerable than other and also may only become vulnerable in combination with other data. It may seem paradoxical but the bigger cohort you have, the safer an individual's data actually is.

7 MAIN RESULTS AND DISCUSSION

7.1 VALIDITY OF REGISTER-BASED PARKINSON'S DISEASE AND PARKINSONIAN DISORDER CASES (STUDY I)

Results for validity of register-based parkinsonian disorder and PD case ascertainment are shown in fig. 7.1.⁹¹ Of all register-based parkinsonian disorder cases who were detected in the NPR before the gold standard screening, 88.0% (95% CI 78.4–94.4) were confirmed (*i.e.* the PPV). For the register-based PD cases among these, a lower proportion were confirmed; 70.8% (95% CI 58.9–81.0). There were 21 register-based PD cases that were misclassified, *i.e.*, not confirmed by the gold standard (the false positives) and of these, 13 were diagnosed with a non-PD parkinsonian disorder by the gold standard, 4 had other types of movement disorders and 4 were free from parkinsonian or movement disorders (fig 7.2). Of these 4, only 1 had not been included in the gold standard second screening nor clinical work-up. Thus, almost 30% of register-based PD cases in NPR were erroneously classified as PD, but the major cause of the errors was misclassification between differential parkinsonian and other movement disorders. The proportion of false positive register-based PD cases decreased when we restricted the case definition to only include cases with a primary PD diagnosis in NPR (PPV 83.0%, 95% CI 70.2–91.9) or only cases who had been discharged with a PD diagnosis from specialized hospital departments (PPV 83.3%, 95% CI 62.6–95.3). As expected, restricting the case definition reduced the sensitivity of detection of PD cases from 72.7% to 50.0% when using primary diagnoses and to 23.5% when using specialized hospital departments. Restricting the case definition by excluding all register-based PD cases who also had non-PD parkinsonian disorder diagnosis did not improve PPV. Thus, taking into account both PPV and sensitivity, the most optimal PD case definition in NPR was to include only cases with primary PD diagnoses. However, if the outcome of interest was all parkinsonian disorders, all primary and contributory diagnoses in NPR may be used.

Of 194 parkinsonian disorder cases, 42 (31.6%) died within 3 years from screening and among these 17 were correctly given a parkinsonian disorder diagnosis in the CDR, with one additional false-positive register-based case, giving a PPV of 94.4% (95% CI 72.7–99.9) for all parkinsonian disorder cases in CDR. For register-based PD cases in CDR the PPV was 66.7% (95% CI 41.0–86.7). Until the end of follow-up in 2008, 127 parkinsonian disorder cases had died and of these 55 were detected by CDR, giving an overall sensitivity of 43.3% (95% CI 34.6–52.4) for all parkinsonian disorders and 57.1% (95% CI 45.4–68.4) for PD. Restricting the PD case definition to having a diagnosis as an underlying cause of death only improved the PPV to 80.0% (95% CI 28.4–99.5) but reduced the sensitivity to 19.5% (95% CI 11.3–30.1). However, only 7.2% of all register-based cases in NPR and CDR were detected by CDR only and thus the reduced sensitivity of PD detection in CDR when restricting the case definition would not affect the overall sensitivity in any great way. The low sensitivity in CDR is in line with previous reports of underreporting of PD on death certificates.^{92–95}

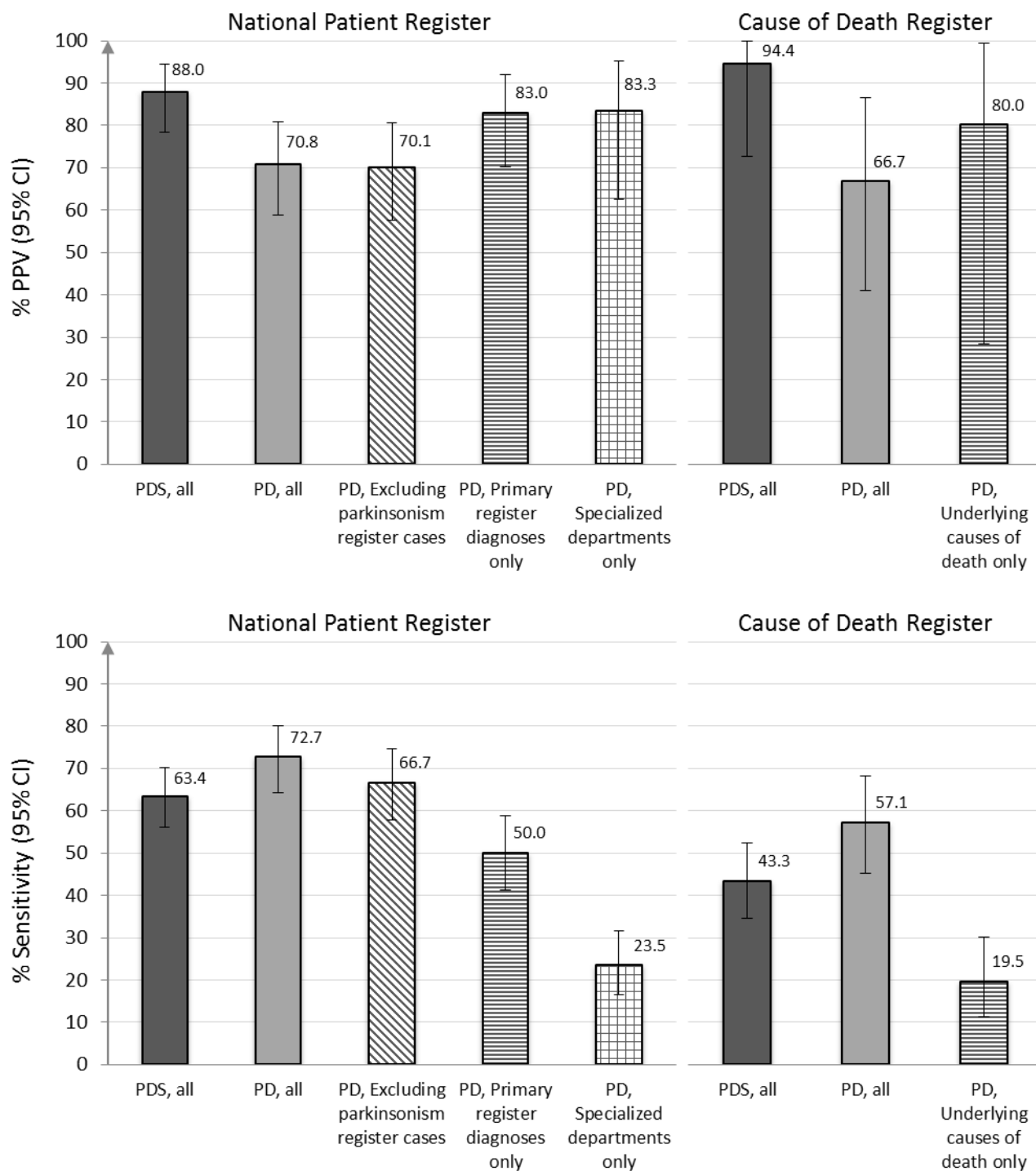


Fig 7.1. Validity of register-based parkinsonian disorder and Parkinson's disease cases ascertained in the National Patient Register (on the left) and the Cause of Death Register (on the right), overall and with restrictions for definition of Parkinson's disease cases (Study I). Positive predictive value and sensitivity reported as percentages with 95% Clopper-Pearson exact confidence intervals for proportions. CI: Confidence interval, PD: Parkinson's disease, PDS: parkinsonian disorder, PPV: Positive predictive value.

In conclusion, the NPR and CDR are valid data sources in epidemiological studies of parkinsonian disorders. However, there are certain caveats that researchers should be wary of:

1. Age at first detection of PD or a parkinsonian disorder in NPR is a poor estimate of age at onset of the disease. On average, register-based cases were first detected 6.8 years (SD 5.5) after the first parkinsonian symptoms appeared. This is not surprising as NPR detection is dependent on hospital admission and PD is typically treated on an out-patient basis, at least in the beginning of the disease.
2. The reduced sensitivity of register-based detection of PD in particular makes the NPR and CDR poor sources of data in studies of occurrence of parkinsonian disorders or PD (*i.e.* overall incidence and prevalence).
3. If studies depend on differential parkinsonian disorder diagnoses, there will be misclassification. By restricting the register-based PD definition misclassification can be reduced, but not without reducing the sensitivity of detection of an already low-prevalent outcome.
4. The validity of register-based differential non-PD parkinsonian disorder diagnoses such as PSP and CvP is outside the scope of this study, but considering the results for PD, the validity may be poor.

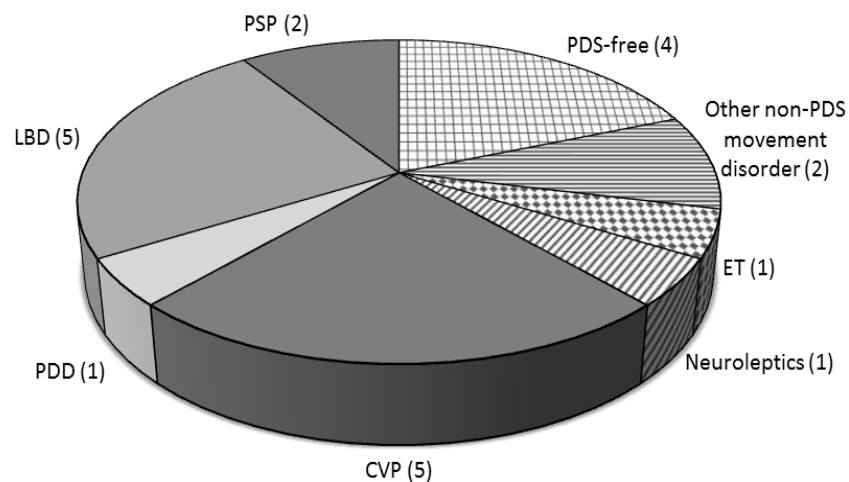


Fig 7.2. Gold standard diagnoses among false positive register-based Parkinson's disease cases in the National Patient Register (Study I). Total $n=21$, number of each diagnosis in parenthesis. Solid pieces represent various parkinsonian disorder diagnoses whereas patterned pieces represent non-parkinsonian disorder diagnoses. CVP: Cerebrovascular parkinsonism, ET: Essential tremor, LBD: Lewy body dementia, PDD: Parkinson's disease dementia, PDS: Parkinsonian disorders, PSP: Progressive supranuclear palsy.

7.1.1 Methodological Considerations

We can be fairly certain that all the positive gold standard parkinsonian disorder cases are correct, as the clinical work-up was quite rigorous. However, since we did not have access to post-mortem neuropathology, there may be uncertainty regarding the differential parkinsonian diagnoses. It's also worth to question how certain we can be of the negative gold standard cases. The majority of gold standard parkinsonian disorder-free participants were classified as such because they (or their proxy) denied having any parkinsonian symptoms in a screening telephone interview. It is possible that mild parkinsonian disorder cases could have been missed, and if they were given a diagnosis in the registers, they would erroneously be classified as false positives which would mean that the PPV was underestimated. However, only 1 of 21 false positive register-based PD cases had been classified as a non-case after the first screening, all others had been included in the clinical work-up, or at least the second telephone screening. We excluded all participants with unclear diagnoses because of non-participation in any of the screening phases or clinical work-up. Non-participants were on average older than participants. Additional analyses showed that both PPV and sensitivity may be lower for older cases, so these measures may have been overestimated.

7.2 ASSOCIATION BETWEEN OCCUPATIONAL EXPOSURE AND PARKINSON'S DISEASE (STUDY II)

Results from the study of association between PD risk and occupational exposure are shown in fig 7.3.⁹⁶ There was no association between PD risk and occupational exposure to metal dust, wood dust, animal handling, pesticides, stone and concrete dust, chrome and nickel dust, quartz dust, organic dust, welding smoke, oil, asbestos, organic solvents or irritating gas (HRs ranging from 0.73 to 1.16). For men with any probability of inorganic dust exposure there was a significantly increased risk of PD: HR 1.63 (95% CI: 1.09-2.44).

Inorganic dust exposure was found in several occupations including farm workers, dental technicians, painters, construction workers, glass workers, dock laborers and cleaners. Farmers^{26,97,98} and cleaners^{98,99} have previously been associated with PD risk. Inorganic dust is a constituent of the particulate matter (PM) in air pollution. Inhaled PM may reach not only the systemic circulation but also the central nervous system, where it can cause neuroinflammation through the activation of microglia, which are found in the substantia nigra of PD patients.^{100,101} One neuropathological study¹⁰¹ showed that young adult residents in cities with high air pollution (*i.e.* Mexico City) had neuroinflammation, blood-brain barrier disruption and accumulation of α -synuclein. It has been hypothesized that α -synuclein-related pathology found in PD is possibly initiated in the periphery, for instance via airborne toxin exposure to the olfactory epithelium. This is supported by the association between olfactory dysfunction and prevalent PD.¹⁰² Thus, there is a plausible mechanism through which

inorganic dust exposure may increase PD risk. However, it should be considered that this may have been a chance finding as when we adjusted for multiple testing using the Bonferroni correction for 14 comparisons, the association between inorganic dust exposure and PD was no longer significant.

We found no association between occupational pesticide exposure and PD. Although this association is not universally seen, it is fairly well substantiated^{29,30} and there are plausible molecular mechanisms through which it could work, particularly for insecticides which by design are often neurotoxic.¹⁰³ In Sweden, herbicides have been more commonly used than insecticides and pesticide use overall has been strictly regulated, which could explain the lack of an association in our data. However, other Swedish studies have found associations to PD, not with pesticides specifically, but with farming as an occupation.²⁶ However, in light of our results, that association with farming could be explained at least partially by inorganic dust exposure.

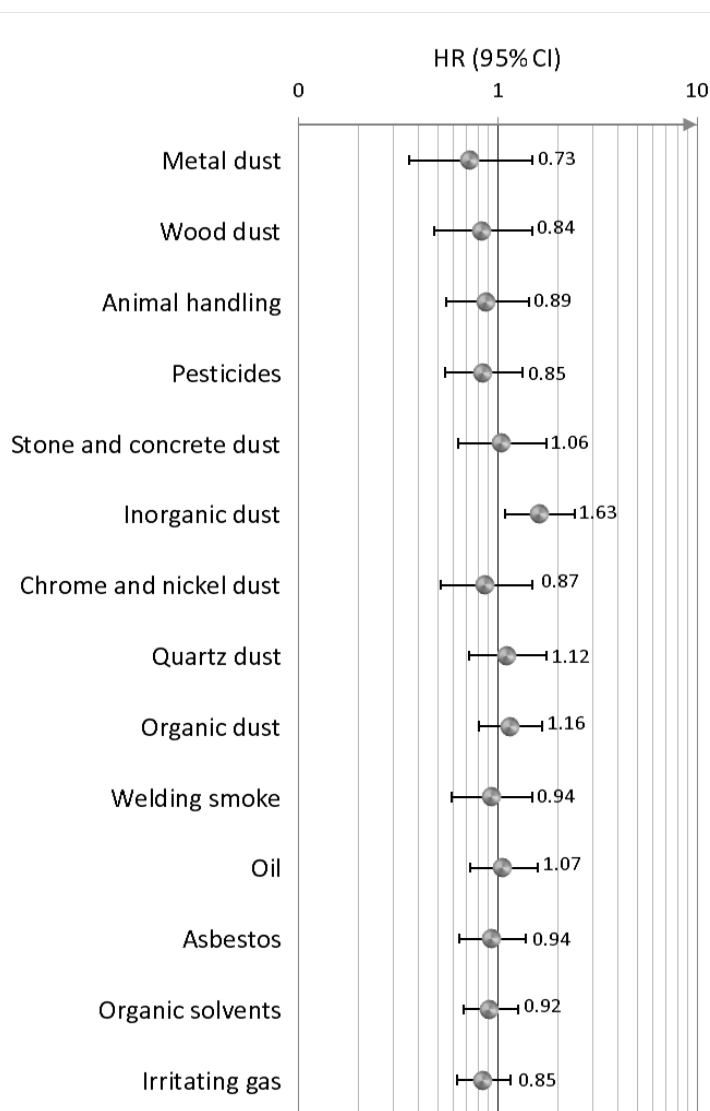


Fig 7.3. Association between Parkinson's disease risk and 14 occupational exposures (Study II). Total n=14,169. All exposed (any exposure probability level) vs. non-exposed as the reference. HRs are from Cox proportional hazard regression models with age as the underlying time-scale and adjusted for smoking and education as covariates. HRs shown on a log scale. CI: Confidence interval, HR: Hazard ratio.

7.2.1 Methodological Considerations

As the results from Study I (which was performed chronologically after this study) showed, restricting the register-based PD case definition by excluding cases who also had a non-PD parkinsonian disorder, reduced sensitivity but did not increase PPV of PD. Thus, it was not surprising that results were very similar for PD and all parkinsonian disorders (see table 3 in published article). We used a JEM to assess occupational exposure, which is a useful tool to explore which specific occupational aspects may confer risk for a disease. It is also a convenient and sensitive way of assessing exposure for a large historical cohort such as in the present study. However, as exposure is not assessed on an individual level, the specificity of the exposure measure may be low.^{79,80} Another limitation is that this particular JEM assessed probability of any exposure as opposed to actual level of exposure, which may further reduce specificity. We used self-reported occupational information from mailed questionnaires in the 1960's and 70's. A large proportion of men (25.0%) were missing information on occupation, which reduced our power. If we had repeated the study today we may have used the 1970 census as the source of occupational data instead which may have reduced missingness. We adjusted the analyses for age, sex, smoking status at baseline and educational level. However, as is shown in the DAG in fig. 5.4, smoking and education are not clear-cut confounders of the association between occupational exposure and PD. Smoking has an effect on PD risk, but the causal pathway between smoking and occupational exposure is less clear. What is certain is that smoking does not "cause" occupation, however, occupation may "cause" smoking, in which case smoking is a mediator between occupation and PD. Most likely there is unmeasured confounding ("U") which causes both smoking and occupation, and it's plausible that at least part of this confounding is captured by adjusting for education. Education on its own is certainly causally associated with occupation, but perhaps the only association between education and PD is through occupation, in which case education is not a confounder and should not have been adjusted for.

7.3 FAMILIAL AGGREGATION OF PARKINSON'S DISEASE (STUDY IV, IVtw)

Results for familial aggregation of all parkinsonian disorders and PD in all first-degree relatives, siblings, parent-offspring, twins and spouses are shown in fig 7.4.⁸⁹ As expected, there was a clear indication of a trend of greater familial aggregation associated with higher genetic correlation (MZ > DZ > Siblings > Parent-offspring > Spouses), although the estimate for DZ twins appeared to be somewhat lower than expected. One explanation could be the younger age range in the siblings compared to the twins as it's likely that familial aggregation is more commonly observed in families of young onset cases. Nevertheless, the results add to evidence from several previous studies, including those on the same data as here,^{16,104} supporting strong familial aggregation of PD and parkinsonian disorders overall.¹⁷

Similarly strong familial risks were found for AD and dementia, although there was not as clear a trend of genetic correlation as for parkinsonian disorders and PD. The highest estimate for familial aggregation in AD was among siblings (HR 5.79, 95% CI 4.84-6.92). AD is likely affected more than PD by the right truncation at age 78 of the cohort and this could explain why the siblings show such high familial risks. Interestingly, there was a slight significant familial risk of AD among spouses (HR 1.07, 95% CI 1.03-1.10). This association could be explained by correlation in educational level among spouses.

7.4 CO-OCCURRENCE OF PARKINSONISM AND DEMENTIA (STUDY IV, IVtw)

Among all participants in Study IV (all persons in the first-degree relative and spouse cohorts, born 1890-1960), there were 15,048 individuals who had co-occurring dementia and parkinsonian disorders, corresponding to 23.4% of parkinsonian disorder cases or 5.7% of dementia cases. There was a strong association between PD status and risk of co-occurring dementia (HR 2.83, 95% CI 2.76-2.89). Among the twins (Study IVtw) there were 195 dementia cases among 882 PD cases (22.1%), with a similar association between PD status and risk of co-occurring dementia (HR 2.23, 95% CI 1.92-2.60). These estimates are similar to what others have found.⁴⁶ Undoubtedly the association between PD and dementia in individuals is strong.

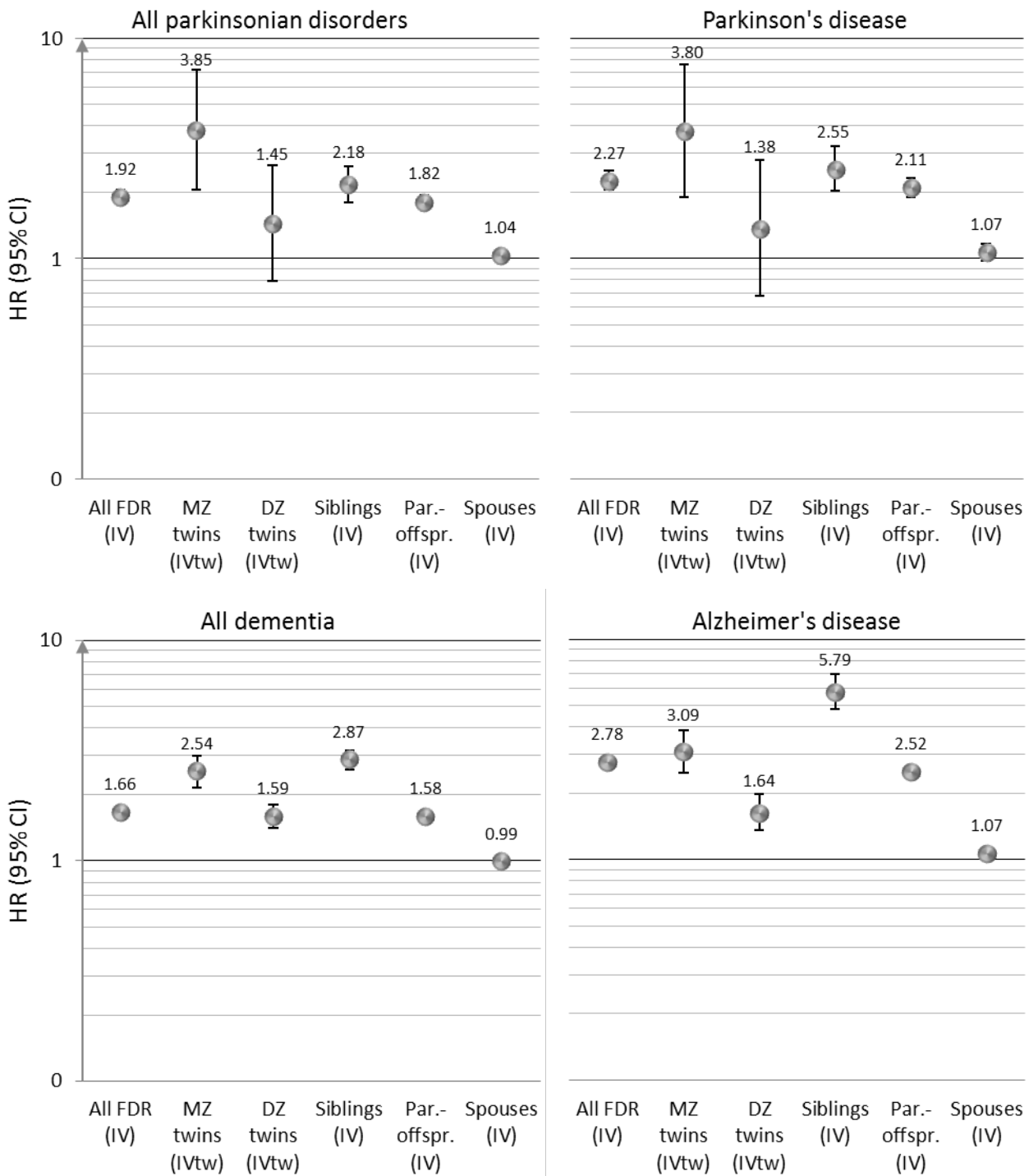


Fig 7.4. Results for familial aggregation of Parkinson's disease, all parkinsonian disorders, Alzheimer's disease and all dementia (Study IV, IVtw). Association between disease risk and having an affected relative, overall and by relative type. All FDR refers to siblings and offspring in parent-offspring comparisons. All FDR born 1932-1960 and followed 1972-2009, Spouses born 1890-1960 and followed 1969-2009, twins born 1886-1958 and followed 1974-2010. HRs shown on a log scale. CI: Confidence interval, DZ: Dizygotic, FDR: First-degree relatives, HR: Hazard ratio, MZ: Monozygotic, Par-offspr.: Parent-offspring.

7.5 FAMILIAL COAGGREGATION OF PARKINSON'S DISEASE AND DEMENTIA (STUDY III, IV, IVtw)

Results for dementia risk associated with family history of PD and parkinsonism risk associated with family history of AD are shown in fig. 7.5 (study IV, IVtw). When dementia was the outcome there was no association with family history of PD for parent-offspring, twins or spouses, but there was a modest significant association for siblings, HR 1.20 (95% CI 1.02-1.41). However, as the power was higher in the parent-offspring analysis, the sibling association did not make a large impact on the result for all first-degree relatives. When parkinsonism was the outcome the association in siblings was somewhat stronger, HR 1.35 (95% CI 1.11-1.65). There was also an association of similar magnitude between parkinsonism risk and having a parent with AD, HR 1.21 (95% CI 1.13-1.29). As expected, there was no association among spouses in either direction. According to the assumptions we have made about the causal pathways between risk of dementia or PD and family history of the opposite disease, we would expect the association to be similar in both directions and thus the different results depending on direction for parent-offspring analyses is puzzling. In parent-offspring analyses the family history exposure was ascertained among the parents

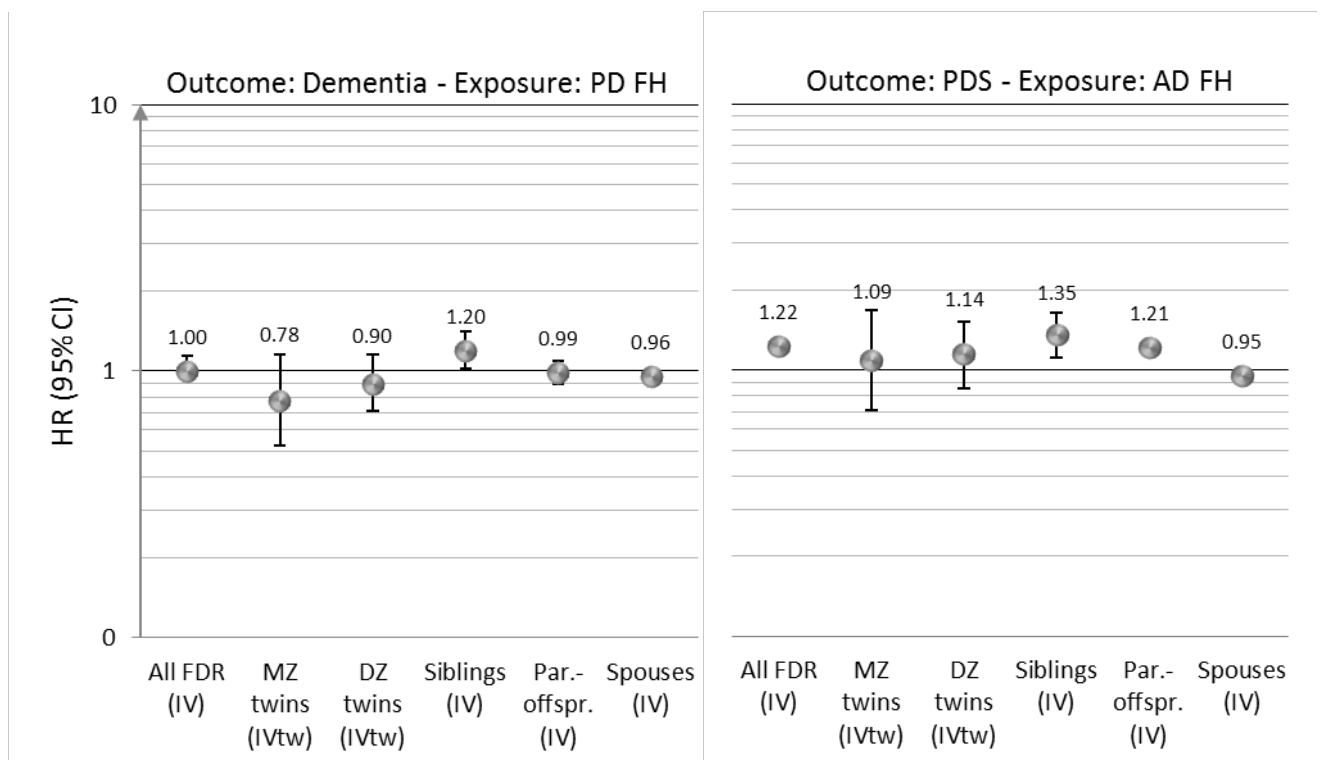


Fig 7.5. Results for familial coaggregation of Parkinson's disease and dementia (Study IV, IVtw). Association between dementia risk and having a relative with PD (left) and association between parkinsonism risk and having a relative with AD (right), overall and by relative type. All FDR refers to siblings and offspring in parent-offspring comparisons. All FDR born 1932-1960 and followed 1972-2009, Spouses born 1890-1960 and followed 1969-2009, twins born 1886-1958 and followed 1974-2010. HRs shown on a log scale. AD: Alzheimer's disease, CI: Confidence interval, DZ: Dizygotic, FDR: First-degree relatives, FH: Family history, HR: Hazard ratio, MZ: Monozygotic, Par-offspr.: Parent-offspring, PDS: Parkinsonian disorders.

who were not subjected to the same right censoring as their offspring in whom the outcome was ascertained. Therefore the parent history exposure better represents the onset age spectrum in both dementia and parkinsonism. However, parkinsonism has an earlier onset than dementia and as was shown in fig. 5.8, the parkinsonism cases in the first-degree relative cohort seem to better represent the overall case population than the dementia cases who obviously represent a young onset case population. Perhaps the risk of dementia associated with family history of PD is not apparent among young onset dementia cases and thus we would have to follow the offspring for an additional 10-20 years before the results are comparable in both directions. On the other hand, if we were to follow the cohort in older ages it would become more and more similar to the twin cohort, in which no associations were shown indicating that perhaps the sibling associations could only be found in a younger population.

Results for meta-analysis of cohort and case-control studies on dementia risk associated with family history of PD are shown in fig 7.6¹⁰⁵ (Study III, IV). Before adding the estimate from Study IV to the meta-analysis, the meta-estimate was pointing towards a conclusion in support of a modest familial coaggregation of PD and dementia, just as Study IV did overall. However, as described above, there was no association for familial aggregation among parent-offspring in Study IV and this association drove the estimate for all first-degree relatives to also show no association, despite the small but significant association among siblings. So, when the estimate from study IV for first-degree relatives was added to the meta-analysis, as indeed it would have been if we had published study IV before commencing the data collection for study III, the conclusion of Study IV is changed to that there is no evidence, not even of a modest association, in support of familial coaggregation of PD and dementia.

But what about among siblings, where there was an association in study IV? When adding the sibling estimate from study III to the other published estimates for dementia risk associated with family history of PD among siblings, the meta HR was 1.17 (95% CI 1.02-1.36, I^2 : 0.0%, $p=0.942$, weight of study IV = 79.5%). However, when the estimates from Study IVtw, which did not show any association (twins are siblings after all), were also added, these counteracted the estimate from study IV and the resulting meta-estimate was of no association (fig 7.7). As the population in study IVtw overlaps partly with the siblings of Study III, these estimates should not strictly speaking be entered together in the same meta-analysis, or their weights should be reduced to account for the dependence of the data. Nevertheless, the conclusion must be that the meta-analysis of familial coaggregation of PD and dementia among siblings does no longer indicate any significant association when all data are combined.

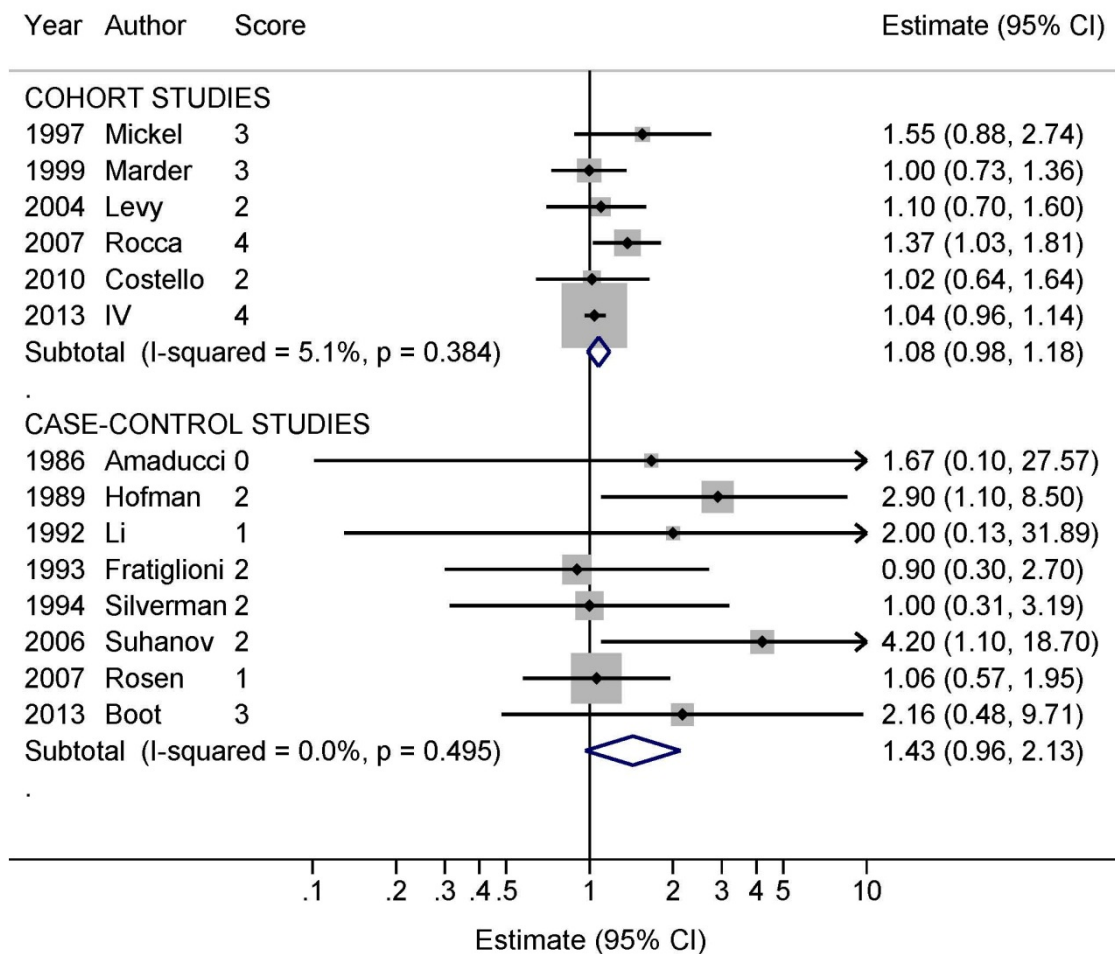


Fig 7.6. Forest plot depicting meta-analysis of studies of dementia risk associated with first-degree family history of Parkinson's disease (Study III with added data from study IV and Boot (2013)). Note that cohort studies estimate HRs and case-control studies estimate ORs. Estimates shown on log scale. The grey boxes around each estimate illustrate the relative sample size of each study, which is proportional to the weight. Without the added estimate from Study IV to the cohort studies the meta HR was 1.20 (95% CI 1.03-1.40, I^2 : 0.0%, $p=0.513$). Without the added estimate from Boot (2013) to the case-control studies the meta OR was 1.40 (95% CI 0.92-2.12, I^2 : 1.3%, $p=0.414$). CI: Confidence interval, HR: Hazard ratio, OR: Odds ratio.

A funnel plot of all cohort and case-control studies of dementia risk associated with first-degree family history of PD is shown in fig 7.8. When all data were included (all studies originally included in study III plus the estimate from study IV and Boot (2013)), there appears to be funnel plot asymmetry indicating a small study effect, or, as it is otherwise called, publication bias. However, considering the cohort studies and case-control studies separately there was no evidence of publication bias; Egger's test of funnel plot asymmetry $p=0.275$ and 0.232 , respectively. The choice to show all cohort and case-control studies together in the same funnel plot was made to illustrate how the cohort studies were consistently more powerful but the case-control studies on average produced higher estimates of association.

Thus, considering studies III, IV and IVtw together, the overall conclusion is that there may be no familial coaggregation of PD and dementia. Or if there is, it is of a very modest magnitude and perhaps only present among young onset cases. Indeed, if it is there it is so small that the only clinical significance of this result is that families of PD patients may know that their family history probably does not put them at an increased risk of dementia and clinicians may know that family history of either PD or dementia is not a useful biomarker for the diagnosis of the opposite disease.

But why is the magnitude of the associations for familial coaggregation so small (or not there at all)? Under the assumption that the DAG in fig 5.10 represents the true underlying causal pathways between PD and dementia in families and that all pathways have positive effects (fig 7.9a), we would expect to see an association between PD_R and Dementia, even if we are wrong about there being shared familial risk ($F_{PD/Dementia}$) (fig 7.9b). This because there is an open path going

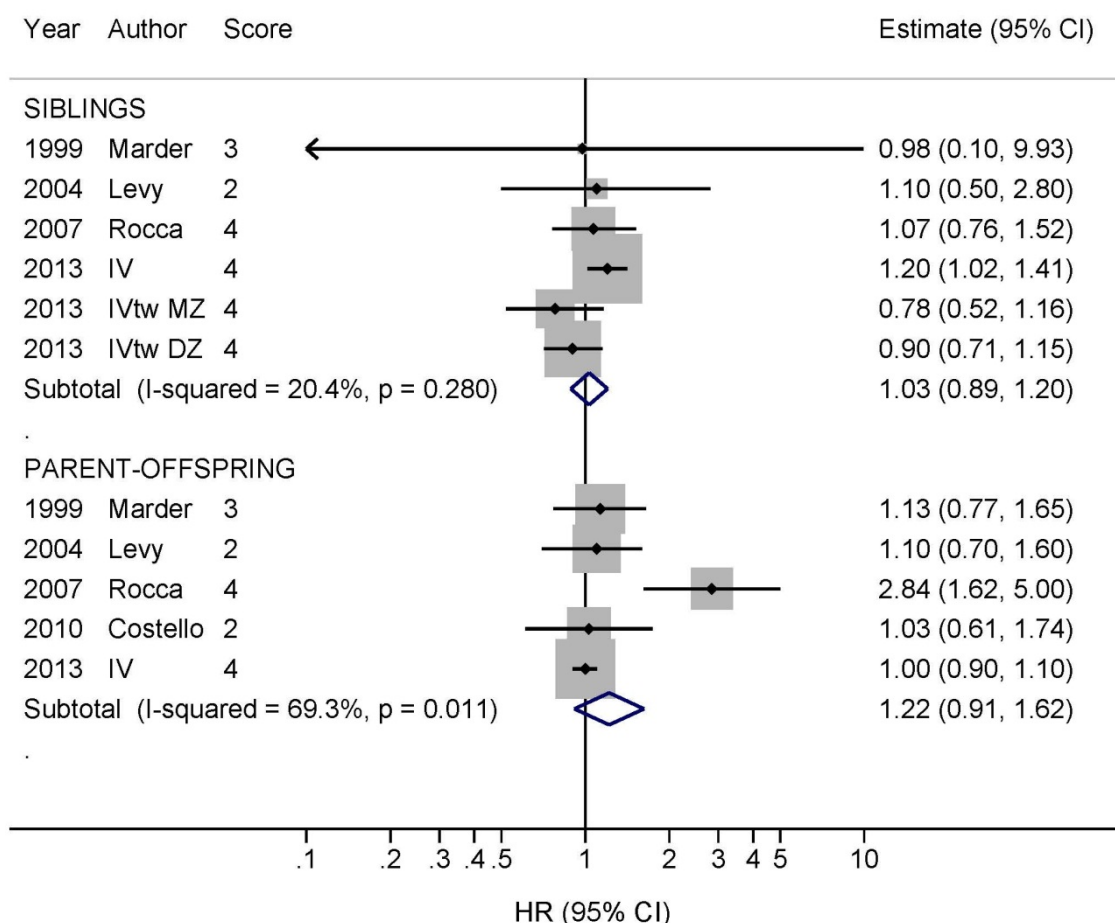


Fig 7.7. Forest plot depicting meta-analysis of cohort studies of dementia risk associated with first-degree family history of Parkinson's disease (Study III with added data from studies IV and IVtw). Estimates shown on log scale. The grey boxes around each estimate illustrate the relative sample size of each study, which is proportional to the weight. Without the added estimates from Studies IV and IVtw to the sibling risk analysis the meta HR was 1.07 (95% CI 0.78-1.47, I^2 : 0.0%, $p=0.995$). Without the added estimate from Study IV to the parent-offspring risk analysis the meta HR was 1.34 (95% CI 0.89-2.02, I^2 : 67.7%, $p=0.026$). CI: Confidence interval, HR: Hazard ratio.

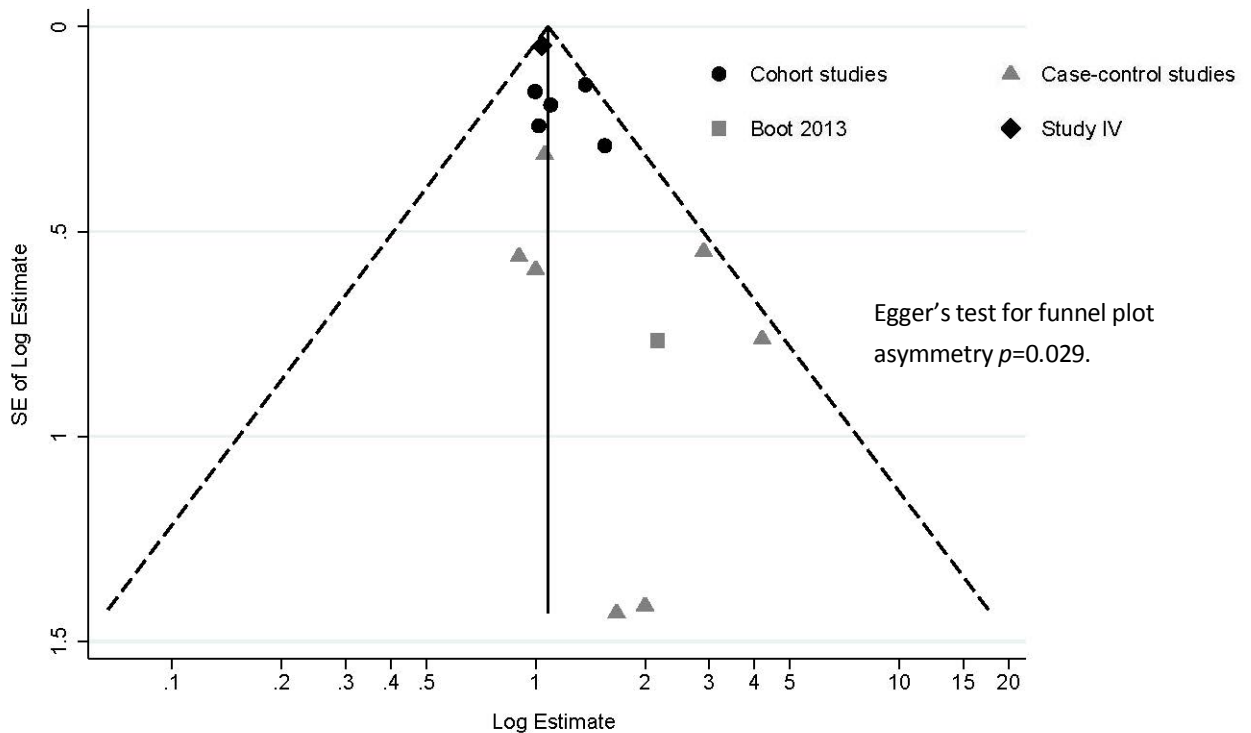


Fig 7.8. Funnel plot with 95% pseudo confidence intervals of all studies of dementia risk associated with first-degree family history of Parkinson's disease (Study III with added data from study IV). Note that cohort studies estimate HR and case-control studies estimate OR. The overall meta-estimate represented by the vertical line was 1.11 (95% CI 1.00-1.22). Without the added estimated from Boot (2013) and Study IV the overall meta-estimate was 1.20 (95% CI 1.03-1.41) and $p=0.244$ for funnel plot asymmetry. HR: Hazard ratio, OR: Odds ratio, SE: Standard error.

$PD_R \leftarrow F_{PD} \rightarrow PD \rightarrow Dementia$ and there is a strong association between PD_R and PD, and between PD and comorbid Dementia. The absence of an association between PD_R and Dementia (or the fact that the association is so modest) leads us to think that the DAG may be wrong. But wrong in which way? There are several options:

- i. Perhaps there is no direct causal pathway $PD \rightarrow Dementia$ (and $PD_R \rightarrow Dementia_R$) (fig 7.9c). Then, since the path between PD and Dementia though C is blocked in the collider at PD, $F_{PD/Dementia}$ is the only remaining open path and then the observed weak association or no association is a true effect. However, is it plausible that there is no direct effect of PD on dementia and that the whole association between PD and comorbid dementia is due to a non-shared familial confounding (C)? In light of the neuropathological and molecular evidence, this does not seem likely. For instance, α -synuclein is a precursor to the non-amyloid component of amyloid- β plaques, and may, when abundant, thus promote dementia-related pathology.^{106,107}

Perhaps one of the pathways is actually an inverse association which counteracts the effect of $F_{PD/Dementia}$ making it appear as a weak association between PD_R and Dementia? Let's explore which pathway could be inverse:

- ii. If $PD \rightarrow Dementia$ (and $PD_R \rightarrow Dementia_R$) was inverse, meaning that the presence of PD actually decreased the risk of dementia, it could counteract the effect of $F_{PD/Dementia}$ and the association between PD_R and Dementia would disappear or appear weak (fig 7.9d). However, this is not a likely explanation due to the evidence for a positive causal effect listed above and the fact that C would then have to be super strong to counteract the inverse causal effect of $PD \rightarrow Dementia$ since the association between PD and Dementia is so strong.
- iii. One of the pathways $F_{PD/Dementia} \rightarrow PD / F_{PD/Dementia} \rightarrow PD_R$ or $F_{PD/Dementia} \rightarrow Dementia / F_{PD/Dementia} \rightarrow Dementia_R$ could be inverse, meaning that the effect of $F_{PD/Dementia}$ is an increased risk of one of the diseases and a decreased risk of the other disease which would appear as a net effect of no association or very weak association between PD_R and Dementia (fig 7.9e). This is actually the most likely model, and the results from the spouse risk analyses support it. Firstly, within PD the HR for spouses was 1.07 (95% CI 0.98-1.17, fig 7.4), which is not significant but may point towards the existence of a weak environmental F_{PD} among spouses. However, there was a slight negative association between PD_R and Dementia among spouses, HR 0.96 (95% CI 0.93-0.99). Again, the association is very weak but it's notable that it's in the opposite direction, as if some pathway is counteracting the effect of F_{PD} . And that could be the presence of an environmental $F_{PD/Dementia}$ which weakly increases risk of PD while decreasing risk of Dementia. In fact, there are two possible environmental risk factors that may have this effect and that are correlated in families; nonsmoking^{108,109} and higher education.^{25,110}

7.5.1 Methodological Considerations

The greatest methodological problem in Study IV is short follow-up age-wise. As shown in the lexis-diagrams (fig 5.9), follow-up in the first-degree relative cohort was only until 78 years for the oldest of the cohort. To grasp the impact of this right censoring we can compare age-specific incidence rates of disease in the first-degree relative cohort to the twin cohort which was followed to a theoretical age of 124 (fig. 5.9). For PD and parkinsonian disorders, many cases are missed after age 78, but the younger cases do not seem to be unusually young as incidence does not increase that dramatically after 78. For dementia however, a very small proportion of cases are detected before age 78 and it is apparent that these cases are not representative of typical dementia cases in the population. The impact of this censoring is thus that when dementia was the outcome, it was really young onset dementia specifically. It is possible that the association would have been lost if we had been able to follow the cohort for as long as the twin cohort. Unfortunately we could not test to see if a sub cohort of the twin cohort with the same follow-up as the first-degree relative cohort would also show an association due to too low power. Lastly, the imperfect sensitivity of the register-based ascertainment may have caused non-differential misclassification and bias towards the null.

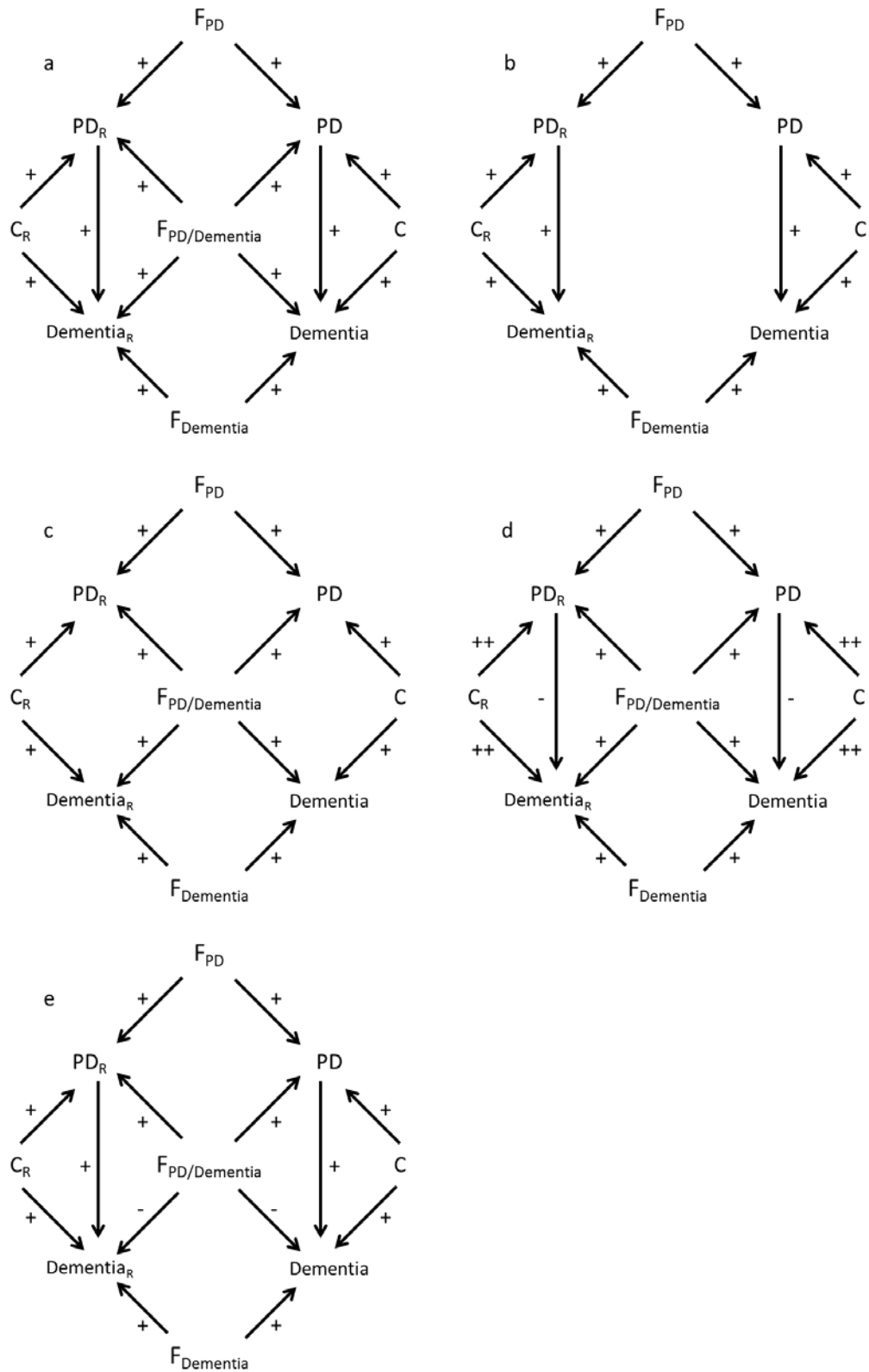


Fig 7.9a-e. Directed acyclical graphs representing different causal models for the association between first-degree family history of PD and dementia (Study III, IV, IVtw). DAG adapted from Hudson et al.⁵⁹ F_{PD} : Familial risk factors for PD, $F_{Dementia}$: Familial risk factors for dementia, $F_{PD/Dementia}$: Familial risk factors common to PD and Dementia, C: Non-shared risk factors for both PD and Dementia, PD: Parkinson's disease.

8 CONCLUSIONS

- I. Parkinson's disease and parkinsonian disorder register-based case ascertainment in the national patient and Cause of Death Registers is valid for use in epidemiological studies, although not perfect. The most optimal case definition is achieved when using primary hospital discharge diagnoses.
- II. Occupational exposure to inorganic dust may be a risk factor for Parkinson's disease in Swedish men who were of working age in the 1960's and 1970's. In this population, there may not be any association between Parkinson's disease risk and occupational exposure to metal dust, wood dust, animal handling, pesticides, stone and concrete dust, chrome and nickel dust, quartz dust, organic dust, welding smoke, oil, asbestos, organic solvents and irritating gas.
- III. There may be an association of modest magnitude between dementia risk and positive first-degree family history of Parkinson's disease, indicating the presence of familial coaggregation of these disorders.
- IV. There is an association of modest magnitude between dementia risk and positive family history of Parkinson's disease among siblings, but not among parents-offspring, and an association between parkinsonian disorder risk and positive family history of Alzheimer's disease in siblings and parents-offspring. In monozygotic or dizygotic twins there may not be any familial coaggregation of Parkinson's disease and dementia.

There is strong familial aggregation among siblings and parents-offspring for parkinsonian disorders overall, Parkinson's disease, dementia overall and Alzheimer's disease.

Separately, the results from study III and study IV indicate that there is familial coaggregation of Parkinson's disease and dementia, whereas the results from study IVtw are inconclusive. However, when the results from all three studies are pooled there is no longer any association between Parkinson's disease and dementia, and the conclusion is that there may not be familial coaggregation of these disorders.

9 STRESZCZENIE PO POLSKU

Ruchy ciała ludzkiego wykonywane są przez mięśnie sterowane impulsami wysyłanymi z ośrodkowego układu nerwowego za pośrednictwem układu nerwowego obwodowego. Olbrzymia ilość komórek nerwowych tworzących układ nerwowy, zarówno obwodowy jak i ośrodkowy, łączy się ze sobą w rozległą sieć. Każda pojedyncza komórka nerwowa składa się z pękatego ciała komórkowego zawierającego jądro komórkowe oraz z krótkich i długich wypustkach zwanych dendrytami i aksonami. Komórki nerwowe komunikują się ze sobą wysyłając związki chemiczne – neuroprzekaźniki. Cząsteczki neuroprzekaźników, wydzielane przez wypustki nerwowe jednej komórki, reagują z receptorami znajdującymi się na powierzchni sąsiadującej komórki nerwowej.

Na najwyższej części pnia mózgowego znajduje się obustronnie rozmieszczone, wydłużone blaszkowate ciało zwane istotą czarną (substantia nigra). Wyspecjalizowane komórki nerwowe istoty czarnej wytwarzają neuroprzekaźnik dopaminę, której funkcją jest między innymi inicjowanie i koordynacja ruchów w ośrodkowym układzie nerwowym. Zwrodnienie tych właśnie komórek nerwowych, powodujące z kolei lokalne zakłócenia w wydzielaniu dopaminy, jest głównym czynnikiem ograniczenia sprawności ruchowej, a tym samym patogenezą choroby Parkinsona.

Objawy choroby Parkinsona są drżenie spoczynkowe (tremor), sztywność mięśniowa oraz spowolnienie i zubożenie ruchów. Objawy pojawiają się zwykle po jednej stronie ciała i narastają powoli i stopniowo. Po pewnym czasie choroba prowadzi do poważnego inwalidztwa. Ryzyko zachorowania zwiększa się z wiekiem i jest największe po 65 roku życia. Choroba Parkinsona dotyka około 1% populacji.

Dotychczas odkryto kilka genów, zmiany w których wiąże się obecnie ze zwiększonym ryzykiem występowania choroby Parkinsona. Przypadki, kiedy chorobę można przypisać wyłącznie czynnikom genetycznym, zdarzają się jednak bardzo rzadko. W większości wypadków wystąpienie choroby Parkinsona następuje w wyniku wzajemnego oddziaływania zarówno czynników genetycznych, jak i środowiskowych. Jeżeli chodzi o czynniki środowiskowe, tylko niektóre udało się jak dotąd zidentyfikować. Do tych, o których wiemy, że mają istotny związek z chorobą Parkinsona, należy palenie tytoniu, które wykazuje działanie ochronne, tzn. – palacze mają mniejsze ryzyko zachorowania niż osoby niepalące. Wśród innych dotychczas zbadanych czynników środowiskowych, które przypuszczalnie zwiększają ryzyko zachorowania, są między innymi biocydy używane w rolnictwie, produkty mleczne, oraz metale.

Poza ograniczoną zdolnością ruchową, chorzy na chorobę Parkinsona wykazują również inne objawy neurologiczne, takie jak otępienie (demencja) czyli zaburzenie procesów poznawczych obejmujących funkcje pamięci, koncentracji oraz myślenia i rozumienia. Dlaczego chorzy na Parkinsona cierpią również na otępienie? Dwie

przyczyny uważa się za najbardziej prawdopodobne: 1) Uszkodzenie komórek nerwowych, znajdujących się w śródmózgowiu i wytwarzających dopaminę, rozprzestrzenia się również do innych części mózgu, które kierują procesami poznawczymi; 2) Te same – dziedziczne lub środowiskowe – czynniki, które powodują neurodegenerację w śródmózgowiu, prowadzą również do neurodegeneracji w innych częściach mózgu związanych z czynnościami poznawczymi.

Ogólnym celem niniejszej rozprawy doktorskiej jest lepsze poznanie przyczyn choroby Parkinsona oraz lepsze zrozumienie powodów, dla których u pacjentów cierpiących na chorobę Parkinsona pojawia się również otępienie.

Wiele badań epidemiologicznych, które przeprowadza się zarówno w Szwecji, jak i w innych krajach, opiera się na danych gromadzonych do innych celów niż cele naukowe – na przykład dane gromadzone w rejestrach administracyjnych służby zdrowia. Dlatego powinno się zbadać, jak należy używać danych z tego typu rejestrów w badaniach naukowych. W części I niniejszej pracy przedstawione zostały wyniki badania przydatności diagnoz choroby Parkinsona, znajdujących się w szwedzkim Rejestrze Pacjentów i Rejestrze Przyczyn Zgonów, do badań epidemiologicznych. Diagnozy zawarte w Rejestrze Bliźniąt zostały potraktowane jako złoty standard, czyli jako najlepsze dostępne diagnozy. W Rejestrze Bliźniąt zgromadzone są dane z obszernych badań przeprowadzonych w Szwecji w latach 1998–2004 na ponad 35 tysiącach bliźniąt w wieku 50 lat i więcej.

Rezultaty wykazują, że z zarejestrowanych w Rejestrze Pacjentów przypadków, kiedy u pacjenta zdiagnozowano chorobę Parkinsona w wyniku przynajmniej jednego pobytu w szpitalu, aż 83% można było potwierdzić przez złoty standard (dodatnia wartość predykcyjna). Jeżeli chodzi o Rejestr Przyczyn Zgonów, 80% przypadków zostało potwierdzonych. Większość osób, których choroba została błędnie zarejestrowana, cierpiało na choroby, które, podobnie jak choroba Parkinsona, atakują aparat ruchowy (parkinsonizm). Czułość rejestru wynosi 50%, co oznacza, że ze wszystkich przypadków choroby Parkinsona (według złotego standardu), połowa została zarejestrowana również w Rejestrze Pacjentów przynajmniej jeden raz w czasie życia chorego.

W części II rozprawy zbadany został związek pomiędzy ryzykiem zachorowania na chorobę Parkinsona a czternastoma różnymi czynnikami ekspozycji zawodowej w grupie ponad 14 tysięcy męskich bliźniąt. W latach 1960-tych i 1970-tych uczestniczyli oni w badaniach ankietowych i dane o nich, uzupełnione danymi zawartymi w szwedzkich rejestrach służby zdrowia i zbieranymi przez okres następnych 43 lat, pozwoliły na przeprowadzenie niniejszych badań. Stopień narażenia osób badanych został zmierzony za pomocą matrycy szacującej ryzyko zawodowe (tzn. w zależności od zawodu wykonywanego przez osobę badaną) ekspozycji na różne substancje chemiczne. Rezultaty wykazały, że wśród tych, którzy byli narażeni na pyły nieorganiczne, nastąpiło zwiększenie ryzyka choroby Parkinsona o 63% (współczynnik hazardu 1,63 (przedział ufności 1,09–2,44)). Do grupy substancji określanych jako „pyły nieorganiczne” zalicza się między innymi pył drogowy, podnoszący się w

miejskim ruchu ulicznym. Powyższy rezultat został otrzymany po raz pierwszy i powinien w związku z tym stać się przedmiotem dalszych badań. Żaden związek nie został natomiast zaobserwowany w przypadku pozostałych trzynastu czynników ekspozycji zawodowej na substancje chemiczne, takich jak na przykład biocydy używane w rolnictwie.

Celem analiz przedstawionych w części III i IV niniejszej pracy było zbadanie, czy istnieją rodzinne czynniki dziedziczne i/lub środowiskowe (na przykład z powodu podobnego stylu życia w rodzinie, itp.), mogące spowodować zarówno chorobę Parkinsona jak i otępienie. Jednym ze sposobów podejścia do tego problemu było zbadanie, czy choroby tego typu gromadzą się w rodzinach – innymi słowy, czy krewni pierwszego stopnia (rodzeństwo biologiczne lub rodzice i dzieci) chorują na obie te choroby równocześnie częściej, niż można byłoby tego oczekiwać. W analizie zostały porównane ryzyka zachorowania na chorobę X u ludzi, w których rodzinach są osoby chorujące na chorobę Y lub w których rodzinach ta choroba (Y) nie występuje.

Część III jest publikacją przeglądową, gdzie została przedstawiona metaanaliza rezultatów z 16 publikacji badających albo a) związek pomiędzy ryzykiem otępienia a występowaniem choroby Parkinsona u członków najbliższej rodziny, albo b) związek pomiędzy ryzykiem choroby Parkinsona a występowaniem otępienia w rodzinie.

W części IV zostały przedstawione wyniki badania związków (a) i (b) wśród ponad 2 milionów mieszkańców Szwecji. W badaniu wykorzystano dane o związkach pokrewieństwa z Rejestru Pokoleń oraz dane o stanie zdrowia z rejestrów służby zdrowia. Rezultaty badania z części III wykazały, że osoby mające krewnych pierwszego stopnia cierpiących na chorobę Parkinsona mają, ogólnie rzecz biorąc, zwiększone o 18% ryzyko zachorowania na otępienie w porównaniu z tymi, u których w rodzinie otępienie nie występuje (współczynnik hazardu 1,18 (przedział ufności 1,00-1,39)). Rezultaty badania w części IV pokazują, że u osób, których biologiczne rodzeństwo cierpi na chorobę Parkinsona, ryzyko zachorowania na otępienie zwiększa się o 20% (współczynnik hazardu 1,20 (przedział ufności 1,02-1,41)).

Z przedstawionych tu rezultatów należy wysnuć wniosek, że istnieje niewielka akumulacja tych dwóch chorób w rodzinach. To również oznacza, że nawet jeżeli jest prawdopodobne, że choroba Parkinsona i otępienie mają wspólne rodzinne czynniki ryzyka, nie mogą owe czynniki w pełni wytłumaczyć dlaczego pacjenci cierpiący na chorobę Parkinsona również w tak wysokim stopniu sami cierpią na otępienie.

10 AFTERWORD

The historically relatively recent and dramatic improvement in life expectancy in world-wide is a truly striking image (fig. 10.1). This is the result of reforms in economic, social and medical conditions and it's an immensely positive development. However, there are also serious challenges. For instance, concerning social and economic conditions, an increasing number of persons need to be supported in old age. In the medical field, an increasing number of persons will live with chronic aging-

related diseases as the relative and absolute population at risk of these diseases grows. We need more basic research to understand what is happening with these diseases in the population and if we are ever going to have a chance to actively prevent them. Luckily, it appears that at least for dementia the incidence may be decreasing as this was reported by two recent papers from England¹¹¹ and Sweden.¹¹²

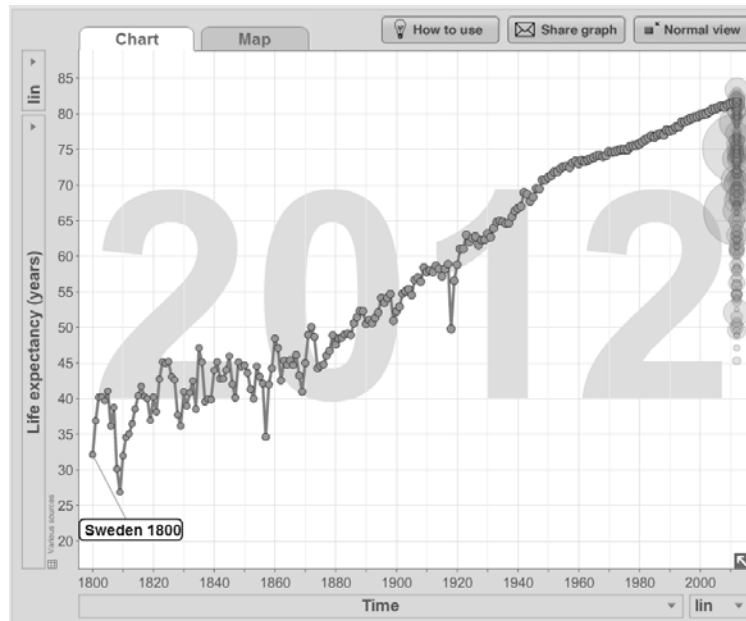


Fig. 10.1. Life expectancy at birth in Sweden from 1800 to 2012 as well as status in 2012 of life expectancy in the rest of the world's countries (seen on the right). Source: Free material from www.gapminder.org.

10.1 ERRATA

No thesis is perfect and that certainly includes this one. Before even sending it to the printers I can already report two formatting errors:

- **Paper II**

There is an error in the column headings of tables 2 and 3, the correct headings should read: N; IR/100,000 PYR; HR; 95% CI.

- **Paper IV**

The columns are misaligned in the bottom part of table 1.

If any more errors are discovered after the publication of the thesis I will post them online at this address: <http://goo.gl/Q4kxed> (also accessible through the QR-code).



11 FUNDING SOURCES

This data collection for the STR that made the work in studies I and IVtw possible was supported by The National Institutes of Health. Other funding came from Karolinska Institutet, the Swedish Research Council, the Swedish Society for Medical Research, the Swedish Medical Society and the Parkinson Foundation in Sweden.

12 ACKNOWLEDGEMENTS

Karolinska Institutet (KI) has been my academic home for just over a decade and I couldn't have asked for a better environment to study and work in, thank you.

My principal supervisor, **Nancy Pedersen**: Thank you for taking a chance on me and proposing that I do a PhD in your group. You believed in me and supported me from the beginning and I want to thank you for not giving up, even though I deserted MEB (twice) before making up my mind about wanting to stay. Thank you for giving me the freedom to pursue all my side projects in research and outside of research. With your support I have been able to travel to and present at conferences in Seoul, Barcelona and Florence, among other places, and I will be forever grateful for that.

My principal co-supervisor, **Karin Wirdefeldt**: Thank you for originally inviting me to do my master thesis project with you at MEB; this was my first practical experience of neuroepidemiology after which I was hooked for life. I have learned a lot from your insight and experience from clinical neurology and I am very grateful that you let me meet some of your patients.

My co-supervisors, **Patrik Magnusson** and **Anna Bennet**: Thank you for letting me be a part of the Magnificent 7, for all your positive energy, warm support and wise advise; my only regret is that I didn't seek you out more often.

My constant co-author, **Margaret Gatz**: Thank you for all your great feedback on my projects, I always looked forward to hearing from you and to your coming to Stockholm.

My external mentor, **Denny Vågerö**: Thank you for letting me work at CHES when I had just got my MSc; it was certainly a very interesting year. I think my genuine interest in register-based data can be traced back to working on the Stockholm Birth Cohort with you. I am very happy that we stayed in touch even after I left to return to MEB.

Biostatistician extraordinaire, **Anna Johansson**: Having you around from the very beginning has been a real comfort and luxury. Thank you for all your patience with me and thank you for the music! MEB's resident DAG expert, **Arvid Sjölander**: Thank you for being a wonderful teacher of statistics and causal inference, for trusting me to be your TA and for letting me bother you with my questions over and over again. I would also like to thank **Paul Lambert** for your contribution to the systematic review and meta-analysis (study III); I couldn't have done it without you.

To all my co-workers at **MEB**, thank you for daily making our department such a good place to work. I would especially like to thank **Gunilla Sonnebring** for being my friend and most reliable source of Charlaïne Harris-related literature, **Marie Krushammar** for expertly keeping track of my supervisor's schedule and helping me navigate a jungle of ethical permits in the highest of spirits, **Camilla Ahlquist** for being our calm port in the

ocean of doctoral education administration; we would be lost without you, and **Christina Hultman** for your encouragement and frequent kind words.

I have been very lucky to be a part of a wonderful and active community of PhD students at MEB over these last few years, all the members of which have contributed to the awesomeness of this journey.

Thank you to my predecessors, in particular **Christina Persson**, **Fang Fang** and **Sara Öberg**: As an epidemiologist, you are my professional role-models and I would be lucky and proud if I manage to follow in either one of your footsteps.

Among my contemporaries (more or less) I am so happy to be able to call so many not just colleagues but primarily friends. **Maria Sandberg**: I predict that in a century, when the neighbors of room 5529 at MEB are suddenly disturbed by the haunting whisper of a loud epi discussion interspersed with laughter, it'll be our ghosts reliving the awesome three years when we were roomies – while we still doodle 2*2 tables on a huge white board somewhere and drink wine with sushi, as if nothing ever changed. **Karin Sundström**: If you were a handbag, you would be a Birkin bag – smart, warm, friendly, elegant, sophisticated, timeless; truly one of a kind. **Stephanie Bonn**: Your productivity level is inspiring, as is your humility, generosity, sewing talent and cultural taste. **Therese Ljung**: I'm sitting opposite you at Espresso House when I'm writing this and thinking about how lucky we are to not be alone during this stressful phase of applying, writing and planning for the future. I want to thank you in particular for gracefully enduring my many recent anxious rants. **Miriam Elfström**: You swept in on the west wind two years ago and made everything sunnier. **Lisa Möller**: Your bright smile always brings me joy ☺ **Sara Christensen**: Thank you for your constant level-headedness and great kindness. **Lovisa Högberg**: The dancing queen with the big heart. **Martin Fransson**: The deep thinker and rock of the group. **Thomas Frisell**: The politician geek genius sommelier, I feel cooler by proxy for knowing you. **Alexander Grankvist**: Who knew when we met you that you would become such a sophisticated gentleman? You never cease to surprise me. **Kaavya Narasimhalu**: You made the rest of us look bad by being so much more efficient and creative, but I never minded because your friendship is as spicy as your cooking and you know I've missed you terribly since you left Stockholm. **Iffat Rahman**: Just like we once wandered around Reykjavik, I look forward to travelling with you again and I'm so happy it's coming up so soon. **Ci Song**: Thank you for being so smart, funny and grounded.

The next generation of PhD students, in particular **Johanna Sieurin**, **Ida Karlsson** and **Vilhelmina Ullemar**: Thank you for giving me hope for the future of MEB.

I am also grateful to have gotten to know some of the brilliant Postdocs at MEB, including **Anna Kähler**, **Lisen Arnheim Dahlström**, **Anna Dahl**, **Lotte Gerritsen** and **Sara Hägg**.

The **PubMed** group, including **DJ Robert Karlsson, Sandra Eloranta, Anna Johansson, Favelle Lamb, Anne Örtqvist, Ralf Kuja-Halkola, Henrik Olsson** and **Sara Christensen**: Thank you for the beer, the laughs and the epic parties!

The **Book circle**, including **Maria Sandberg, Karin Sundström, Therese Ljung, Thomas Frisell, Miriam Elfström, Carolyn Cesta** and **Michael Jan**: Thank you for all the stimulating conversation and for broadening my literary horizons.

Former and current members of **Framtidsrådet** (sometimes known as the KI Development Council), including **Maria Sjögren, Sophia Savage, Christina Bäckman, Jacob Kjell, Andreas Nyström, Stefan Plantman, Indira Chivukula, Ulrika Eriksson** and **Vilhelmina Ullemar**: Working with you during these last two years and finally launching our grand project KI Evolution has been an amazing and challenging opportunity and I am very proud to have been able to contribute to it. I look forward to seeing how Framtidsrådet will develop in the future and I promise that you can count on me as an active alumnus.

A big thank you to my colleagues and friends from other corners of KI, including **Anna Aminoff, Sofi Eriksson, Adina Welander, Kaziwe Mollazadegan, Annika Österdahl** and **Susanne Szydowski**.

My two friends without whom I can't imagine what the last decade would have been like, **Rebecca David** and **Cecilia Björklund**: You make my life richer and I love you.

To all my friends outside of Stockholm, including **Maria Warnefors, Rebecca Litzell** and **Caroline Norén**: Your friendship means the world to me, thank you for sticking around even though we don't get to see each other so often.

An extra thank you to the artistic talents **Rebecca Litzell, Vilhelmina Ullemar** and **Noemi Naszarkowski** for helping me make the thesis look so beautiful on the outside.

Thank you to my small but close extended family, including all the **Flatos** and my aunt **Irena**. To my cousin **Noemi**: you are my sistah, my 6-foot hummingbird, my consigliere. I am so proud of you.

Last but not least, thank you to my mama, **Danuta**.

13 REFERENCES

1. Jacobson S, Marcus EM. Neuroanatomy for the Neuroscientist. Boston, MA: Springer US, 2011.
2. Parkinson J. An essay on the shaking palsy. 1817. *J Neuropsychiatry Clin Neurosci*. Spring 2002;14(2):223-236; discussion 222.
3. Poirier J. James Parkinson (1755-1824) revisited. *Geriatr Psychol Neuropsychiatr Vieil*. Mar 1 2013;11(1):65-72.
4. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. *Arch Neurol*. Jan 1999;56(1):33-39.
5. Holdorff B, Rodrigues e Silva AM, Dodel R. Centenary of Lewy bodies (1912-2012). *Journal of neural transmission*. Apr 2013;120(4):509-516.
6. 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. *Neurobiology of aging*. Mar-Apr 2003;24(2):197-211.
7. Williams DR, Litvan I. Parkinsonian syndromes. *Continuum (Minneapolis)*. Oct 2013;19(5 Movement Disorders):1189-1212.
8. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*. Aug 26 2008;71(9):670-676.
9. Papp MI, Kahn JE, Lantos PL. Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome). *Journal of the neurological sciences*. Dec 1989;94(1-3):79-100.
10. Litvan I, Agid Y, Calne D, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology*. Jul 1996;47(1):1-9.
11. Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. Jan 29 2013;80(5):496-503.
12. Zijlmans JC, Daniel SE, Hughes AJ, Revesz T, Lees AJ. Clinicopathological investigation of vascular parkinsonism, including clinical criteria for diagnosis. *Movement disorders : official journal of the Movement Disorder Society*. Jun 2004;19(6):630-640.
13. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet Neurol*. Jun 2006;5(6):525-535.
14. 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.
15. Tanner CM, Ottman R, Goldman SM, et al. Parkinson disease in twins: an etiologic study. *JAMA*. Jan 27 1999;281(4):341-346.
16. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging*. Apr 9 2011.
17. Thacker EL, Ascherio A. Familial aggregation of Parkinson's disease: a meta-analysis. *Movement disorders*. Jun 15 2008;23(8):1174-1183.
18. Lill CM, Roehr JT, McQueen MB, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics*. Mar 2012;8(3):e1002548.
19. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science*. Oct 31 2003;302(5646):841.
20. 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. *Annals of neurology*. Sep 2002;52(3):276-284.
21. Quik M. Smoking, nicotine and Parkinson's disease. *Trends in neurosciences*. Sep 2004;27(9):561-568.

22. Fang F, Chen H, Feldman AL, Kamel F, Ye W, Wirdefeldt K. Head injury and Parkinson's disease: a population-based study. *Movement disorders : official journal of the Movement Disorder Society*. Nov 2012;27(13):1632-1635.
23. Rughjerg K, Ritz B, Korbo L, Martinussen N, Olsen JH. Risk of Parkinson's disease after hospital contact for head injury: population based case-control study. *BMJ*. 2008;337:a2494.
24. Wirdefeldt K, Adami HO, Cole P, Trichopoulos D, Mandel J. Epidemiology and etiology of Parkinson's disease: a review of the evidence. *European journal of epidemiology*. Jun 2011;26 Suppl 1:S1-58.
25. Frigerio R, Elbaz A, Sanft KR, et al. Education and occupations preceding Parkinson disease: a population-based case-control study. *Neurology*. Nov 22 2005;65(10):1575-1583.
26. Li X, Sundquist J, Sundquist K. Socioeconomic and occupational groups and Parkinson's disease: a nationwide study based on hospitalizations in Sweden. *Int Arch Occup Environ Health*. Jan 2009;82(2):235-241.
27. 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.
28. Goldman SM, Tanner CM, Olanow CW, Watts RL, Field RD, Langston JW. Occupation and parkinsonism in three movement disorders clinics. *Neurology*. Nov 8 2005;65(9):1430-1435.
29. Priyadarshi A, Khuder SA, Schaub EA, Shrivastava S. A meta-analysis of Parkinson's disease and exposure to pesticides. *Neurotoxicology*. Aug 2000;21(4):435-440.
30. Van Maele-Fabry G, Hoet P, Vilain F, Lison D. Occupational exposure to pesticides and Parkinson's disease: a systematic review and meta-analysis of cohort studies. *Environ Int*. Oct 1 2012;46:30-43.
31. Tanner CM, Kamel F, Ross GW, et al. Rotenone, paraquat, and Parkinson's disease. *Environmental health perspectives*. Jun 2011;119(6):866-872.
32. Le W, Sayana P, Jankovic J. Animal Models of Parkinson's Disease: A Gateway to Therapeutics? *Neurotherapeutics*. Oct 25 2013.
33. Mortimer JA, Borenstein AR, Nelson LM. Associations of welding and manganese exposure with Parkinson disease: review and meta-analysis. *Neurology*. Sep 11 2012;79(11):1174-1180.
34. Vishal S, Sourabh A, Harkirat S. Alois Alzheimer (1864-1915) and the Alzheimer syndrome. *J Med Biogr*. Feb 2011;19(1):32-33.
35. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*. May 2011;7(3):263-269.
36. Braak H, Alafuzoff I, Arzberger T, Kretschmar H, Del Tredici K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol*. Oct 2006;112(4):389-404.
37. Goldman JS, Hahn SE, Catania JW, et al. Genetic counseling and testing for Alzheimer disease: joint practice guidelines of the American College of Medical Genetics and the National Society of Genetic Counselors. *Genet Med*. Jun 2011;13(6):597-605.
38. Lobo A, Lopez-Anton R, Santabarbara J, et al. Incidence and lifetime risk of dementia and Alzheimer's disease in a Southern European population. *Acta Psychiatr Scand*. Nov 2011;124(5):372-383.
39. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet neurology*. Dec 2007;6(12):1106-1114.
40. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Archives of general psychiatry*. Feb 2006;63(2):168-174.

41. van Duijn CM, Clayton D, Chandra V, et al. Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. *Int J Epidemiol.* 1991;20 Suppl 2:S13-20.
42. McKeith IG, Perry EK, Perry RH. Report of the second dementia with Lewy body international workshop: diagnosis and treatment. Consortium on Dementia with Lewy Bodies. *Neurology.* Sep 22 1999;53(5):902-905.
43. Aarsland D, Londo E, Ballard C. Parkinson's disease dementia and dementia with Lewy bodies: different aspects of one entity. *Int Psychogeriatr.* Apr 2009;21(2):216-219.
44. Lee JE, Park B, Song SK, Sohn YH, Park HJ, Lee PH. A comparison of gray and white matter density in patients with Parkinson's disease dementia and dementia with Lewy bodies using voxel-based morphometry. *Movement disorders : official journal of the Movement Disorder Society.* Jan 15 2010;25(1):28-34.
45. Savica R, Grossardt BR, Bower JH, Boeve BF, Ahlskog JE, Rocca WA. Incidence of dementia with lewy bodies and Parkinson disease dementia. *JAMA Neurology.* Nov 1 2013;70(11):1396-1402.
46. Aarsland D, Kurz MW. The epidemiology of dementia associated with Parkinson disease. *Journal of the neurological sciences.* Feb 15 2010;289(1-2):18-22.
47. Aarsland D, Zaccai J, Brayne C. A systematic review of prevalence studies of dementia in Parkinson's disease. *Movement disorders.* Oct 2005;20(10):1255-1263.
48. Hofman A, Schulte W, Tanja TA, et al. History of dementia and Parkinson's disease in 1st-degree relatives of patients with Alzheimer's disease. *Neurology.* Dec 1989;39(12):1589-1592.
49. Rosen AR, Steenland NK, Hanfelt J, Factor SA, Lah JJ, Levey AI. Evidence of shared risk for Alzheimer's disease and Parkinson's disease using family history. *Neurogenetics.* Nov 2007;8(4):263-270.
50. Costello S, Bordelon Y, Bronstein J, Ritz B. Familial associations of Alzheimer disease and essential tremor with Parkinson disease. *European journal of neurology.* Jun 1 2010;17(6):871-878.
51. Levy G, Louis ED, Mejia-Santana H, et al. Lack of familial aggregation of Parkinson disease and Alzheimer disease. *Arch Neurol.* Jul 2004;61(7):1033-1039.
52. Marder K, Tang MX, Alfaró B, et al. Risk of Alzheimer's disease in relatives of Parkinson's disease patients with and without dementia. *Neurology.* Mar 10 1999;52(4):719-724.
53. Mickel SF, Broste SK, Hiner BC. Lack of overlap in genetic risks for Alzheimer's disease and Parkinson's disease. *Neurology.* Apr 1997;48(4):942-949.
54. Rocca WA, Bower JH, Ahlskog JE, et al. Risk of cognitive impairment or dementia in relatives of patients with Parkinson disease. *Arch Neurol.* Oct 2007;64(10):1458-1464.
55. Amaducci LA, Fratiglioni L, Rocca WA, et al. Risk factors for clinically diagnosed Alzheimer's disease: a case-control study of an Italian population. *Neurology.* Jul 1986;36(7):922-931.
56. Penrose LS, Mackenzie HJ, Karn MN. A genetical study of human mammary cancer. *British journal of cancer.* Jun 1948;2(2):168-176.
57. Lichtenstein P, Yip BH, Bjork C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* Jan 17 2009;373(9659):234-239.
58. Schnack TH, Poulsen G, Myrup C, Wohlfahrt J, Melbye M. Familial coaggregation of cryptorchidism, hypospadias, and testicular germ cell cancer: a nationwide cohort study. *Journal of the National Cancer Institute.* Feb 3 2010;102(3):187-192.
59. Hudson JI, Javaras KN, Laird NM, VanderWeele TJ, Pope HG, Jr., Hernan MA. A structural approach to the familial coaggregation of disorders. *Epidemiology.* May 2008;19(3):431-439.

60. Elbaz A, McDonnell SK, Maraganore DM, et al. Validity of family history data on PD: evidence for a family information bias. *Neurology*. Jul 8 2003;61(1):11-17.
61. Statistics Sweden, Background Facts, Population and Welfare Statistics 2011:2, Multi-generation register 2010. A description of contents and quality.
62. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekblom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009;24(11):659-667.
63. National Board of Health and Welfare. The National Patient Register. <http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish> [Accessed December 2013].
64. Socialstyrelsen. Kvalitet och innehåll i patientregistret: Utskrivningar från slutenvården 1964–2007 och besök i specialiserad öppenvård (exklusive primärvårdsbesök) 1997–2007. 2009.
65. World Health Organization. Manual of the International Classification of Diseases, injuries and causes of death. 7th rev. Geneva: World Health Organization, 1955.
66. Socialstyrelsen. Dödsorsaksstatistik: Historik, produktionsmetoder och tillförlitlighet. Komplement till rapporten Dödsorsaker 2008. 2010.
67. National Board of Health and Welfare. The Cause of Death Register. <http://www.socialstyrelsen.se/register/dodsorsaksregistret#> (in Swedish) [Accessed December 2013].
68. 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.
69. Lichtenstein P, Sullivan PF, Cnattingius S, et al. The Swedish Twin Registry in the third millennium: an update. *Twin Res Hum Genet*. Dec 2006;9(6):875-882.
70. Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin research and human genetics : the official journal of the International Society for Twin Studies*. Feb 2013;16(1):317-329.
71. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the third millennium. *Twin Res*. Oct 2002;5(5):427-432.
72. Wirdefeldt K, Gatz M, Bakaysa SL, et al. Complete ascertainment of Parkinson disease in the Swedish Twin Registry. *Neurobiol Aging*. Dec 2008;29(12):1765-1773.
73. Tanner CM, Ellenberg J, Mayeux R, Ottman R, Langston JW. A sensitive and specific screening method for Parkinson's disease (PD). *Neurology*. 1994;44(Suppl. 2):A136.
74. Gatz M, Fratiglioni L, Johansson B, et al. Complete ascertainment of dementia in the Swedish Twin Registry: the HARMONY study. *Neurobiol Aging*. Apr 2005;26(4):439-447.
75. Fahn S, Elton RL. The UPDRS Development Committee. Unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Calne D, Goldstein M, editors. Recent Developments in Parkinson's Disease, vol. 2. Florham Park, NJ: MacMillan Healthcare Information; 1987. pp. 153-163.
76. Gilman S, Low P, Quinn N, et al. Consensus statement on the diagnosis of multiple system atrophy. American Autonomic Society and American Academy of Neurology. *Clin Auton Res*. Dec 1998;8(6):359-362.
77. Winikates J, Jankovic J. Clinical correlates of vascular parkinsonism. *Archives of neurology*. Jan 1999;56(1):98-102.
78. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26:404-413.
79. Lope V, Perez-Gomez B, Aragonés N, et al. Occupational exposure to chemicals and risk of thyroid cancer in Sweden. *Int Arch Occup Environ Health*. Jan 2009;82(2):267-274.
80. Plato N, Steineck G. Methodology and utility of a job-exposure matrix. *Am J Ind Med*. Mar 1993;23(3):491-502.

81. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA : the journal of the American Medical Association*. Apr 19 2000;283(15):2008-2012.
82. Huff FJ, Auerbach J, Chakravarti A, Boller F. Risk of dementia in relatives of patients with Alzheimer's disease. *Neurology*. May 1988;38(5):786-790.
83. Hemminki K, Sundquist K, Li X. Familial risks for main neurological diseases in siblings based on hospitalizations in Sweden. *Twin Res Hum Genet*. Aug 2006;9(4):580-586.
84. Kurz MW, Larsen JP, Kvaloy JT, Aarsland D. Associations between family history of Parkinson's disease and dementia and risk of dementia in Parkinson's disease: A community-based, longitudinal study. *Mov Disord*. Dec 2006;21(12):2170-2174.
85. Silverman JM, Raiford K, Edland S, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part VI. Family history assessment: a multicenter study of first-degree relatives of Alzheimer's disease probands and nondemented spouse controls. *Neurology*. Jul 1994;44(7):1253-1259.
86. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. Sep 1986;7(3):177-188.
87. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. Jun 15 2002;21(11):1539-1558.
88. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. Sep 13 1997;315(7109):629-634.
89. Feldman AL, Wirdefeldt K, Johansson AL, Gatz M, Pedersen NL. Evidence for modest familial co-aggregation between dementia and parkinsonism. *European journal of epidemiology*. Nov 19 2013.
90. Boot BP, Orr CF, Ahlskog JE, et al. Risk factors for dementia with Lewy bodies: a case-control study. *Neurology*. Aug 27 2013;81(9):833-840.
91. 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.
92. Beyer MK, Herlofson K, Arslan D, Larsen JP. Causes of death in a community-based study of Parkinson's disease. *Acta Neurol Scand*. Jan 2001;103(1):7-11.
93. Elbaz A, Bower JH, Peterson BJ, et al. Survival study of Parkinson disease in Olmsted County, Minnesota. *Arch Neurol*. Jan 2003;60(1):91-96.
94. 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.
95. Paulson GW, Gill WM. Are death certificates reliable to estimate the incidence of Parkinson's disease? *Mov Disord*. Sep 1995;10(5):678.
96. Feldman AL, Johansson AL, Nise G, Gatz M, Pedersen NL, Wirdefeldt K. Occupational exposure in Parkinsonian disorders: A 43-year prospective cohort study in men. *Parkinsonism Relat Disord*. Nov 2011;17(9):677-682.
97. Kirkey KL, Johnson CC, Rybicki BA, Peterson EL, Kortsha GX, Gorell JM. Occupational categories at risk for Parkinson's disease. *Am J Ind Med*. Jun 2001;39(6):564-571.
98. Tuchsén F, Jensen AA. Agricultural work and the risk of Parkinson's disease in Denmark, 1981-1993. *Scand J Work Environ Health*. Aug 2000;26(4):359-362.
99. Fall PA, Fredrikson M, Axelson O, Granerus AK. Nutritional and occupational factors influencing the risk of Parkinson's disease: a case-control study in southeastern Sweden. *Mov Disord*. Jan 1999;14(1):28-37.
100. Block ML, Calderon-Garciduenas L. Air pollution: mechanisms of neuroinflammation and CNS disease. *Trends Neurosci*. Sep 2009;32(9):506-516.
101. Calderon-Garciduenas L, Solt AC, Henriquez-Roldan C, et al. Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and

- accumulation of amyloid beta-42 and alpha-synuclein in children and young adults. *Toxicol Pathol.* 2008;36(2):289-310.
102. Doty RL, Deems DA, Stellar S. Olfactory dysfunction in parkinsonism: a general deficit unrelated to neurologic signs, disease stage, or disease duration. *Neurology.* Aug 1988;38(8):1237-1244.
 103. Hatcher JM, Pennell KD, Miller GW. Parkinson's disease and pesticides: a toxicological perspective. *Trends Pharmacol Sci.* Jun 2008;29(6):322-329.
 104. Sundquist K, Li X, Hemminki K. Familial risks of hospitalization for Parkinson's disease in first-degree relatives: a nationwide follow-up study from Sweden. *Neurogenetics.* Nov 2006;7(4):231-237.
 105. Feldman AL, Johansson AL, Lambert PC, et al. Familial Coaggregation of Alzheimer's Disease and Parkinson's Disease: Systematic Review and Meta-Analysis. *Neuroepidemiology.* Nov 30 2013;42(2):69-80.
 106. Crews L, Tsigelny I, Hashimoto M, Masliah E. Role of synucleins in Alzheimer's disease. *Neurotox Res.* Oct 2009;16(3):306-317.
 107. Iwai A. Properties of NACP/alpha-synuclein and its role in Alzheimer's disease. *Biochimica et biophysica acta.* Jul 26 2000;1502(1):95-109.
 108. 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.* Sep 2002;52(3):276-284.
 109. Rusanen M, Kivipelto M, Quesenberry CP, Jr., Zhou J, Whitmer RA. Heavy smoking in midlife and long-term risk of Alzheimer disease and vascular dementia. *Archives of internal medicine.* Feb 28 2011;171(4):333-339.
 110. Sharp ES, Gatz M. Relationship between education and dementia: an updated systematic review. *Alzheimer Dis Assoc Disord.* Oct 2011;25(4):289-304.
 111. Matthews FE, Arthur A, Barnes LE, et al. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet.* Oct 26 2013;382(9902):1405-1412.
 112. Qiu C, von Strauss E, Backman L, Winblad B, Fratiglioni L. Twenty-year changes in dementia occurrence suggest decreasing incidence in central Stockholm, Sweden. *Neurology.* May 14 2013;80(20):1888-1894.

“ Why, anybody can have a brain.
That's a very mediocre commodity.
Every pusillanimous creature that crawls on the
earth or slinks through slimy seas has a brain.
Back where I come from, we have universities,
seats of great learning, where men go to become
great thinkers.
And when they come out, they think deep
thoughts and with no more brains than you have.
But they have one thing you haven't got:
A diploma. ”

-The Wizard, *The Wizard of Oz* (1939)



Portrait of the author by Jorge Cham (2011)