From the Department of Medicine, Solna
Clinical Epidemiology Division
Karolinska Institutet, Stockholm, Sweden

GENETIC AND ENVIRONMENTAL INFLUENCES ON MULTIPLE SCLEROSIS COMPLICATIONS AND AETIOLOGY

Kelsi Alexandra Smith

Karolinska Institutet

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023
© Kelsi Alexandra Smith, 2023
ISBN 978-91-8016-998-1
Cover illustration by Ricky Schaede 2023. *A lifetime of multiple sclerosis.*

Genetic and Environmental Influences on Multiple Sclerosis Complications and Aetiology

Thesis for Doctoral Degree (Ph.D)

By

Kelsi Alexandra Smith

The thesis will be defended in public at Inghesalen, Widerströmska Huset, Karolinska Institutet Tomtebodavägen 18, 171 65, and online via Zoom at https://bit.ly/42kxJYO on the 15th of June 2023 at 13:00.

Principal Supervisor:
Professor Scott Montgomery
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

Co-supervisor(s):
Dr. Pernilla Stridh
Karolinska Institutet
Department of Clinical Neuroscience

Professor Lars Alfredsson
Karolinska Institutet
Institute of Environmental Medicine

Professor Ingrid Kockum
Karolinska Institutet
Department of Clinical Neuroscience

Opponent:
Dr. Ruth Dobson
Queen Mary University of London
Preventative Neurology Unit at the Wolfson Institute of Preventative Medicine

Examination Board:
Professor Jonas Björk
Lund University
Division of Occupational and Environmental Medicine

Dr. Joachim Burman
Uppsala University
Department of Medical Sciences

Dr. Anna Sidorchuk
Karolinska Institutet
Department of Clinical Neuroscience

To Mama & Papa Bear.

Thank you for always supporting and encouraging me.

I finally learned how to roar.

Popular science summary of the thesis

Understanding the causes and consequences of multiple sclerosis

The purpose of the research reported by this thesis was to better understand the causes and consequences of multiple sclerosis (MS). MS is a neurological disease affecting the brain and spinal cord, but it is also an immune mediated disease, meaning the body mistakenly attacks itself causing damage to nerves. It is hard to diagnose initially because it has many symptoms that are not the same among different people. Almost three times as many females are diagnosed with MS than males. When studying MS, it is important to combine information about a person's genes, lifestyle, and their environment as all of these factors together might influence MS to develop and to make it worse. This thesis investigated these three factors in four studies, and tried to understand differences in MS between females and males. Individual data from Swedish residents from nationwide registers were linked to genetic and lifestyle data using epidemiological methods. Epidemiology uses statistics to understand the patterns and frequency of disease among large groups of people.

Study I's aim was to see if being diagnosed with pneumonia as a child or teenager was a risk factor for MS. A risk factor is associated with whether a disease develops, like MS. Some research has shown that risk factors happening in teenage years might have the largest magnitude association with risk of MS developing, so pneumonia in different age-groups was included. Overall, this study showed being admitted to hospital with pneumonia aged 11–15 years was associated with double the risk of developing MS, compared with never being diagnosed with pneumonia at this age. As few people had pneumonia, analysing males and females separately was not possible. To understand the results better, urinary tract infections were also included in case undiagnosed MS was increasing the risk of infections, instead of the other way around. Urinary tract infections had no association with MS, so it is unlikely undiagnosed MS explains the association of pneumonia in adolescence with MS risk. Including infectious mononucleosis, the infection usually caused by Epstein-Barr virus which is thought to be a cause of MS, also did not change the association of pneumonia with MS risk.

The purpose of **Study II** was to better understand the treatment of spasticity, a frequent consequence of MS. Spasticity affects muscles, causing them to be painfully stiff, twitch or spasm on their own. The timing of a treatment called baclofen, which is used to treat spasticity, was examined. Spasticity treatment has been studied before, but rarely among people newly diagnosed with MS. This study investigated three things: when did people receive baclofen, what were the factors influencing who got baclofen, and when did they start and stop using it? Surprisingly, most often people with MS received baclofen within a half year to three years after their MS diagnosis. Previous studies showed that it was after many years of MS that people needed treatment. We also showed that younger

people with MS who are more disabled received treatment more often than older people with MS who had similar disability. No differences between males and females were found, which was important because it was previously thought that males most often received spasticity treatment. And lastly, we found that most people stopped using baclofen within two years after they started it, showing that treatments that are either better at stopping spasticity, or that do not have as many side effects, are needed for people with MS.

Study III and IV focused on possible genetic causes and consequences of MS. This study investigated whether the melanocortin-1 receptor (*MC1R)* gene is a risk factor for MS developing, and if it is also a risk factor for making MS worse. The *MC1R* makes a receptor in the skin that controls the skin's reaction to ultraviolet radiation from sunlight, but it also affects the immune system. This receptor controls if you tan or burn in sunlight, and since MS is thought to be influenced by ultraviolet radiation and vitamin-D (vitamin-D is made by ultraviolet radiation when it hits the skin), it could be that the *MC1R* might influence how MS develops and progresses. **Study III** found that some variations in the *MC1R* were associated with a reduced risk of MS, but only among females. The *MC1R* was not influenced by other MS risk factors, such as low vitamin-D, whether people in your family have MS, individual body mass index and smoking. This is relevant because it shows that the *MC1R* might be a specific risk factor among females, and that other factors might be involved in MS developing among females. **Study IV** looked to see if the *MC1R* affected the severity of MS, using an Age-Related MS Severity Score, which measures MS severity correcting for a person's age. Although the *MC1R* affected MS risk among females, it did not seem to affect MS severity.

In **conclusion**, this thesis adds to the field of MS in several ways. It considered the timing of risk factors because knowing when the most critical time is for MS developing, such as pneumonia in adolescence, can help understand future ways to avoid getting MS. Knowing when treatments for MS symptoms, like spasticity, are most used can help doctors and patients understand when to expect and treat the symptoms to avoid further disability. Finding that MS patients stop taking a treatment such as baclofen, is useful in showing that better treatment alternatives might be needed. Using a combination of data about individuals and genetics can help to better understand how different factors work together in MS. In this case, the *MC1R* gene variations decreased MS risk among females, but did not influence MS severity for females or males. Making sure to consider causes and consequences of MS separately among males and females can help to better understand MS, and why females are diagnosed with MS more often. In conclusion, it is the hope of this author that this thesis helps to better understand risk factors before and after MS diagnosis both for future studies, and to help people with MS.

Populärvetenskaplig sammanfattning

Syftet med denna avhandling var att skapa en bättre förståelse för orsaker och konsekvenser av multipel skleros (MS). MS är en neurologisk sjukdom som påverkar hjärnan och ryggmärgen, men det är också en autoimmun sjukdom, vilket innebär att kroppen attackerar sig själv och åsamkar nervskador. Det är till en början svårt att diagnosticera då MS kan ge upphov till många olika symptom som kan skilja sig mellan olika personer. Nästan tre gånger fler kvinnor än män diagnosticeras med MS. När man studerar MS är det viktigt att kombinera information om personens gener, livsstil och deras miljö då alla dessa är faktorer som tillsammans kan påverka utvecklingen av MS och förvärra sjukdomen. Denna avhandling utreder dessa tre faktorer i fyra studier, och försöker förstå hur MS skiljer sig mellan kvinnor och män. Individuella data från svenska invånare i nationella register länkas till genetisk och livsstilsdata genom epidemiologiska metoder. Inom epidemiologi använder man statistik för att förstå mönstren och frekvensen av sjukdomen hos större grupper av människor.

Studie I; Denna studie syftade till att utreda om diagnos av lunginflammation som barn eller tonåring var en riskfaktor för MS. En riskfaktor är en faktor som ökar sannolikheten att en sjukdom utvecklas, såsom MS. Viss forskning har visat att riskfaktorer som man utsätts för i tonåren har störst påverkan på risken att utveckla MS. Av denna anledning har ett flertal olika åldersgrupper inkluderats i studien. Resultaten visade att sjukhusinläggning för lunginflammation i åldern 11–15 år var associerad med dubbelt så stor risk att utveckla MS jämfört med någon som inte diagnosticerats med lunginflammation i den åldern. Då det var få personer som hade lunginflammation gick det inte att analysera kvinnor och män separat. För att utesluta att inte odiagnostiserad MS ökar risken för infektioner, istället för det motsatta, inkluderades även urinvägsinfektioner i studien. Urinvägsinfektioner hade ingen association med MS, så det är inte troligt att odiagnostiserad MS förklarar den ökade MS-risken hos ungdomar med lunginflammation. Inte heller när körtelfeber (infektionen som ofta orsakas av Epstein-Barrvirus, som också misstänks orsaka MS) inkluderades förändrades associationen mellan lunginflammation och MS-risk.

Syftet med **Studie II** var att bättre förstå behandling av spasticitet, ett frekvent symptom vid MS. Spasticitet påverkar muskler och får dem att smärtsamt stelna, krampa eller rycka av sig själva. Tidpunkt för behandling med baklofen, ett läkemedel mot spasticitet, undersöktes. Spasticitetsbehandling har studerats tidigare, men sällan hos personer som nyligen diagnosticerats med MS. Denna studie undersöker tre saker: När fick personer baklofen, vilka faktorer påverkade vem som fick baklofen, samt när började och slutade de använda baklofen? Överraskande nog så upptäcktes det att de flesta personer med MS som fått baklofen fick det inom ett halvår till 3 år efter deras MS-diagnos. Tidigare studier hade visat att det var först många år efter MS-diagnos som personer behövde behandling. Vi såg också att yngre personer med MS högre svårighetsgrad fick behandling

oftare än äldre personer med MS som hade motsvarande svårighetsgrad. Inga skillnader kunde urskiljas mellan män och kvinnor, vilket var viktigt då man tidigare trott att män oftare får spasticitetsbehandling. Avslutningsvis såg vi att de flesta personerna slutade använda baklofen inom två år efter att de börjat använda det, vilket tyder på att behandlingar som antingen är bättre på att behandla spasticitet eller som har mindre biverkningar behövs för personer med MS.

Studie III och IV undersökte möjliga genetiska orsaker och konsekvenser av MS. Mer specifikt undersöktes om gener för melanocortin-1-receptor (*MC1R*) var en riskfaktor för utveckling av MS, och om det även var en riskfaktor för försämring av MS. *MC1R* producerar en receptor i huden som kontrollerar hudens reaktion till ultraviolett (UV-) strålning från solljus, men den påverkar även immunsystemet. Denna receptor kontrollerar huruvida en person blir solbränd eller blir röd i solljus. Då MS misstänks påverkas av UV-strålning och vitamin-D (vitamin-D produceras när huden utsätts för UV-strålning) kan det innebära att MC1R påverkar hur MS utvecklas och försämrats. I **Studie III** fann vi att vissa variationer i *MC1R* kunde associeras med minskad risk för MS, men enbart hos kvinnor. *MC1R* påverkades inte av andra riskfaktorer för MS, såsom låga nivåer vitamin-D, huruvida andra personer i ens familj har MS, ens body mass index (BMI) eller rökning. Detta är relevant då det visar att *MC1R* kan vara en specifik riskfaktor hos kvinnor och att olika faktorer kan vara involverade i hur MS utvecklas hos kvinnor jämfört med män. **Studie IV** undersökte om *MC1R* påverkade svårighetsgraden av MS, genom att använda den så kallade Age-Related Multiple Sclerosis Severity Score, vilket mäter sjukdomens svårighetsgrad korrigerat för en persons ålder. Även om *MC1R* påverkade risken för MS hos kvinnor verkade det inte påverka svårighetsgraden av sjukdomen.

Sammanfattningsvis bidrar denna uppsats till MS-forskningen på flera sätt. Den har undersökt hur olika tidpunkter i livet påverkar MS-risk, då det är viktigt att veta om det finns en period i livet som är mest kritiskt för utvecklingen av MS, såsom lunginflammation i ungdomen. Det bidrar till att förstå hur man i framtiden kan hitta sätt att förhindra insjuknande i MS. Att veta när behandlingar för MS-symptom, såsom spasticitet, används mest frekvent kan hjälpa läkare och patienter att förstå när de kan förvänta sig och behandla symptom för att förhindra utvecklingen av rörelsehinder. Upptäckten att MS-patienter slutar ta behandlingar såsom baklofen är viktigt för att visa att bättre behandlingsalternativ kan behövas. Genom att använda en kombination av data om individer och genetik kan förståelsen för hur olika faktorer samverkar i MS förbättras. Vissa *MC1R*-genvariationer minskar MS-risken hos kvinnor, men påverkar inte svårighetsgraden av sjukdomen hos varken kvinnor eller män. Genom att studera orsaker och konsekvenser av MS separat mellan män och kvinnor kan förståelsen för MS förbättras, exempelvis varför kvinnor i större grad diagnosticeras med MS. Som slutsats hoppas författaren att denna avhandling ska öka förståelsen för riskfaktorer före och efter MS-diagnos inför framtida studier, men även hjälpa personer med MS.

Abstract

Multiple sclerosis (MS) is an immune mediated, neurological disease that results in a chronic accumulation of disability over time. Many advances in the understanding, diagnosis and treatment of MS have been made, yet questions surrounding causes and consequences still remain. The scope of this thesis was to examine aspects of MS aetiology and complications using individual level data from large, nationwide registers in Sweden. Register data was linked with genetic and lifestyle data and analyses used classic and genetic epidemiological methods. Using a multi-disciplinary approach in MS research is of importance given MS is believed to develop through a combination lifestyle and environmental factors coupled with underlying genetic susceptibility. This thesis includes a total of four studies with the overall goal of further understanding the timing of factors in MS risk and complications, while considering sex-specific differences.

Study I aimed to understand whether pneumonia, a common occurrence after MS diagnosis, is also involved in MS risk and if there are specific ages at which risk is elevated. The timing of risk factors is important to consider as there is evidence to support a critical risk period in adolescence for MS development. MS cases and matched general-population controls were included between 1968-2012. The National Patient Register and MS Register were the main exposure (hospital-diagnosed pneumonia) and outcome (MS) data sources. This study showed that pneumonia between ages 11-15 years was associated with increased risk of MS by 200% compared with no pneumonia between these ages. Examining sex differences was not possible due to small numbers. The association remained even after controlling for additional infectious risk factors. These were infectious mononucleosis, an infection that is usually a result of Epstein-Barr virus which is thought to be a cause of MS, and a control infection of urinary tract infections for prodromal or reverse causation effects.

Study II was a cohort study among people with MS that identified factors associated with first oral baclofen treatment. This study used various national registers including the Prescribed Drug Register and included MS patients between 2005-2014. Baclofen is a specific pharmacological treatment for spasticity, and its use among MS patients, particularly newly diagnosed patients, has not been well studied. This study included both people with incident and prevalent MS, to identify possible time-specific patterns of spasticity treatment. A larger proportion of new MS patients received baclofen than previously thought, as patients with incident and prevalent MS received baclofen most often 0.5 to 3 years after diagnosis, respectively. Males and females were similar in their spasticity treatment patterns after controlling for MS-specific factors. Age and disability specific patterns emerged with younger, more disabled patients receiving baclofen treatments at a rate three times that of older, similarly disabled patients. This is of importance as baclofen use previously has been reported more often among males and

more disabled, older patients. Discontinuation of baclofen occurred rapidly, with over 75% of individuals discontinuing treatment within two years. This demonstrates the need for more effective, tolerable treatments for spasticity among people with MS.

The final two studies investigated associations with a human pigmentation gene, melanocortin-1 receptor gene (*MC1R)* for MS risk (**Study III**) and severity (**Study IV**)**.** The *MC1R* codes for a receptor in the skin controlling skin sensitivity to ultraviolet radiation, which may be important in MS given low exposure to ultraviolet light and low levels of vitamin-D are risk factors. To investigate MS risk, a case-control study examined whether *MC1R* single nucleotide polymorphisms (SNPs) were associated with MS risk. This also included a possible overlapping, additional gene region with *MC1R*, tubulin beta class III (*TUBB3)*. This identified rs885479-A, a SNP in the overlapping *MC1R-TUBB3* region, as being inversely associated with MS only among females, with a 25% reduction in MS risk among female carriers. This SNP is known as a red hair colour variant in the *MC1R*, as it is also associated with red hair colour and high skin sensitivity to ultraviolet radiation. The association of this SNP with MS was further investigated and was shown to be independent of other known MS risk factors including genetic risk for low vitamin-D level, familial history of MS, body mass index and smoking. To investigate whether the *MC1R-TUBB3* region was associated to MS severity, a cross-sectional analysis was used. Red hair colour *MC1R* variants were investigated for their association with the Age-Related Multiple Sclerosis Severity Score. There were possible sex-specific differences, but no association with MS severity withstood correction for multiple testing.

In **conclusion**, this thesis contributes to the understanding of the risk factors for MS aetiology and consequences. It adds to the growing evidence of a critical risk period in adolescence for MS risk and highlights the importance of assessing risk factors such as pneumonia prior to the potential prodromal period for MS. This thesis underlines the importance of using cohort studies to assess treatment timing for aspects such as spasticity to be able to identify age- and disability-specific trends. It further emphasizes the importance of long-term, effective management of MS symptoms given the high rates of spasticity treatment discontinuation. Ensuring that causes and consequences of MS are considered separately among males and females can help to better understand MS, and why females are diagnosed with MS more often. Using a combination of information types, including genetic data, can help to better understand how factors work together in MS. In this case, *MC1R* variants decreased MS risk among females only, but did not influence MS severity. Lastly, examining the same risk factors prior to and after MS diagnosis can help to better determine the causes and consequences of MS to inform future study directions, with the ultimate goal of helping people with MS.

Keywords: Multiple sclerosis, risk factors, severity, pneumonia, spasticity, melanocortin-1 receptor gene, register-based, Sweden

Funding

This PhD project was in part funded by external sources. It should be noted that the funders had no role in hypotheses, study designs, data analysis or interpretation, in drafting of manuscripts or other documents, and did not influence decisions to publish work.

MS Canada (formerly known as the MS Society of Canada) was the primary funder of my PhD and is a Canadian non-profit organisation. MS Canada financially supported my growth and development as a young MS researcher through an endMS Doctoral Studentship Award (MSSC-3835). This was awarded in 2020, 2021, 2022 and 2023 for a combined total of 88,000 CAD; approximately 680,000SEK. Specifically, funding was awarded for projects III and IV, with a combined title of *Pigmentation genes and the timing of sun exposure in MS development and progression*.

NEURO Sweden (Neuroförbundet) is a Swedish non-profit organisation that financially supported this PhD project in 2022 (F2021-0070) and 2023 (F2022-0088) for a combined total of 103,610 SEK. Specifically, they provided funding for projects III and IV, with a combined title of *Pigmentation genes and the timing of sun exposure in MS development and progression*.

Grants for participation in conferences were obtained through the Karolinska Institute Travel Grant and the European Committee for the Treatment and Research in Multiple Sclerosis (ECTRIMS), and the joint efforts of the American Committee for the Treatment and Research in Multiple Sclerosis – ECTRIMS collaboration.

Additional funds held in Professor Scott Montgomery's name (my main supervisor) were also a source of funding for this PhD and included funding in no particular order from: Astra Zeneca, Nyckelfonden, the UK Economic and Social Research Council (ESRC) to the International Centre for Life Course Studies, F.Hoffmann-La Roche Ltd, and Novartis Pharma AG.

List of scientific papers in thesis

I. Smith KA, Hiyoshi A, Burkill S, Bahmanyar S, Öckinger J, Alfredsson L, Olsson T, Montgomery S. **Hospital diagnosed pneumonia before age 20 years and MS risk.** *BMJ Neurology Open* 2020; 2:e000044. doi: 10.1136/bmjno-2020-000044. PMID: 33681783.

II. Smith KA, Piehl F, Olsson T, Alfredsson A, Hillert J, Kockum I, Stridh P, Montgomery S. **Spasticity treatment patterns among people with multiple sclerosis: a Swedish cohort study.** *Journal of Neurology, Neurosurgery & Psychiatry* 2022; Published Online First: 20 December 2022. doi: 10.1136/jnnp-2022-329886. Epub ahead of print. PMID: 36539267.

III. Smith KA, Shchetynsky K, Hedström AK, Huang J, Kockum I, Hillert J, Olsson T, Montgomery S, Stridh P. **Sex differences in the association of the melanocortin-1 receptor gene with multiple sclerosis risk.** [Manuscript].

IV. Smith KA, Shchetynsky K, Hedström AK, Huang J, Kockum I, Hillert J, Olsson T, Montgomery S, Stridh P. **Melanocortin-1 receptor gene variants and multiple sclerosis severity.** [Manuscript].

Scientific papers not included in the thesis

V. Smith KA, Burkill S, Hiyoshi A, Olsson T, Bahmanyar S, Wormser D, Geissbühler Y, Moore A, Kharat V, Montgomery S. **Comorbid disease burden among MS patients 1968-2012: A Swedish register-based cohort study.** *Mult Scler*. 2021 Feb;27(2):268-280. doi: 10.1177/1352458520910497. Epub 2020 Mar 12. PMID: 32162580.

VI. Burkill S, Smith KA, Stridh P, Kockum I, Hillert J, Lindahl H, Alfredsson L, Olsson T, Piehl F, Montgomery S, Bahmanyar S. **The DQB1*03:02 genotype and treatment for pain in people with and without multiple sclerosis.** *Front Neurol*. 2020 Sep 4;11:993. doi: 10.3389/fneur.2020.00993. PMID: 33013655.

VII. Xu Y, Hiyoshi A, Brand JS, Smith KA, Bahmanyar S, Alfredsson L, Olsson T, Montgomery S. **Higher body mass index at ages 16 to 20 years is associated with increased risk of a multiple sclerosis diagnosis in subsequent adulthood among men.** *Mult Scler*. 2021 Jan;27(1):147-150. doi: 10.1177/1352458520928061. Epub 2020 Jun 8. PMID: 32507076.

VIII. McKay KA, Smith KA, Smertinaite L, Fang F, Ingre C, Taube F. **Military service and related risk factors for amyotrophic lateral sclerosis**. *Acta Neurol Scand*. 2021 Jan;143(1):39-50. doi: 10.1111/ane.13345. Epub 2020 Oct 12. PMID: 32905613.

IX. Xu Y, Smith KA, Hiyoshi A, Piehl F, Olsson T, Montgomery S. **Hospital-diagnosed infections before age 20 and risk of a subsequent multiple sclerosis diagnosis.** *Brain*. 2021 Sep 4;144(8):2390-2400. doi: 10.1093/brain/awab100. PMID: 33693538.

X. Xu Y, Hiyoshi A, Smith KA, Piehl F, Olsson T, Fall K, Montgomery S. **Association of infectious mononucleosis in childhood and adolescence with risk for a subsequent multiple sclerosis diagnosis among siblings.** *JAMA Netw Open*. 2021 Oct 1;4(10):e2124932. doi: 10.1001/jamanetworkopen.2021.24932. PMID: 34633426.

XI. Brand JS, Smith KA, Piehl F, Olsson T, Montgomery S. **Risk of serious infections in multiple sclerosis patients by disease course and disability status: Results from a Swedish register-based study.** *Brain Behav Immun Health*. 2022 May 11;22:100470. doi: 10.1016/j.bbih.2022.100470. PMID: 35607517.

Contents

List of abbreviations

alpha-MSH	alpha-melanocortin stimulating hormome
ARMSS	Age-Related Multiple Sclerosis Severity Score
ATC	Anatomic Therapeutic Classification
BBB	Blood brain barrier
BMI	Body mass index
cAMP	cyclic adenosine monophosphate
CI	Confidence interval
CNS	Central nervous system
DMT	Disease-modifying therapy
DNA	Deoxyribonucleic acid
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
EDSS	Expanded Disability Status Scale
EIMS	Epidemiological Investigation of Multiple Sclerosis
GEMS	Genes and Environment in Multiple Sclerosis
GREML	Genomic relatedness based restriction maximum-likelihood model
GRM	Genetic relatedness matrix
GRS	Genetic risk score
GWAS	Genome wide association study
HLA	Human leukocyte antigen gene region
HR	Hazard ratio
ICD	International Classification of Disease
IM	Infectious mononucleosis
IMSE	Immunomodulation and Multiple Sclerosis
LISA	Longitudinal Integrated Database for Health Insurance and Labour Market Studies
MAF	Minor allele frequency
MC1R	Melanocortin-1 receptor gene
MC1R	Melanocortin-1 receptor
MLMA	Mixed linear model association
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MSSS	Multiple Sclerosis Severity Scale
OR	Odds ratio
PC	Principle component
PCA	Principle component analysis
PPMS	Primary progressive MS
RHC	Red hair colour
RRMS	Relapsing-remitting MS
SE	Standard error
SEK	Swedish Kronor
SNP	Single nucleotide polymorphism
SPMS	Secondary progressive MS
STOPMS	Stockholm Prospective Assessment of Multiple Sclerosis
TUBB3	Tubulin beta class III gene
TUBB3	Tubulin beta class III
UTI	Urinary tract infection
UVR	Ultra-violet radiation

1 Introduction and literature review

1.1 MS: A brief lookback through time

The earliest text outlining what was likely multiple sclerosis (MS) was in the early 1300s describing a palsy-like disease. This occurred hundreds of years prior to the lectures of Jean Charcot, the French physician who has been credited with first describing MS as a specific medical affliction.[1] Charcot and his colleague Edmé Vulpian observed lesions in the spinal cord of patients presenting with symptoms of what Charcot would name *scleros en plaques* in his series of lectures and lessons.[2] It is known today as MS.[1]

Great advancements in the understanding and treatment of MS have occurred since Charcot's characterisation of the disease, especially in the past 30 years. Diagnosis has progressed with new and specific disease diagnostic criteria allowing for earlier diagnoses. Treatments have also rapidly become available with the first specific disease modifying therapies (DMTs) arriving in the 1990's, subsequently followed with additional, highly effective treatments a short ten years later. In conjunction with diagnosis and treatment progress, understanding of MS also expanded. Various studies determined the relationship between the central nervous system (CNS), the body's immune system and their involvement in MS development and progression. Prominent risk factors emerged such as female sex, smoking, and specific MS risk genes. Yet, despite all advancements, specific causes and biological mechanisms of MS development and progression remain stubbornly elusive.

MS is a notoriously complex, chronic, neurological, immune mediated disease with disease processes beginning years prior to overt clinical manifestation. The description of the disease itself alludes both to its complexity, but also what remains to be fully understood about its causes and consequences. It is thought to be caused by an interaction of underlying genetic susceptibility, triggering environmental and lifestyle factors[3,4] with exposures increasing risk among individuals as early as childhood leading eventually to pathogenesis of MS. MS development is likely a result of the combined influence of accumulated risk factors over time, rather than one single source. Once these processes begin to develop and manifest, there is currently no way to stop the inevitable diagnosis and progression of MS. Symptoms of MS can be non-specific, subtle and fleeting, which can often lead to their dismissal or misinterpretation, especially in the very earliest pre-clinical stages of the disease, years prior to a diagnosis. Symptoms can affect multiple areas of the body including numbness, tingling, or limb weakness, vision and balance problems, difficulties with grasping objects, bladder or bowel issues among others. Symptoms are individual and varied and are often difficult to attribute a symptom to MS. Once diagnosis is made, the variation in initial symptoms continues throughout the course of the disease, necessitating a multi-disciplinary approach to treatment and management of MS. This multi-disciplinary approach is also reflected in the research of MS with a

growing number of consortia, multi-centre, multi-national efforts and cross-field approaches to understand how MS develops, and how to effectively treat patients.

1.2 Epidemiology and impact of MS

The variability in MS makes it both an interesting, albeit difficult disease to study. The variability makes it difficult to directly pinpoint precise biological mechanisms of how it develops. It also means that a variety of people worldwide with different backgrounds, genetics, and exposures, all with various MS characteristics, are affected by MS. Yet, commonalities still exist and are described through epidemiological efforts. In 2020, the Atlas of MS conducted the largest world-survey of MS to date, collecting data from 87% of countries worldwide. An estimated 2.8 million people live with MS, which has increased from the estimate of 2.3 million in 2013.[5] The increase in prevalence could be due to a number of reasons, including increased life expectancy for people with MS, increased awareness and surveillance, clearer and earlier diagnostic criteria, or even an increased risk of MS resulting in a higher incidence of the disease. The prevalence overall could also be higher than currently recorded, as data are mostly collected from high income countries than low and lower middle-income countries.[5] Trends in multiple countries in several studies have also shown increases in incidence and prevalence over time.[6] Globally, the prevalence is 36 per 100,000 individuals, or 1 in 3 000 adults, but varies greatly among countries.[5,7] Historically, it has also been noted that Scandinavians seem to have a high prevalence of MS,[8] a pattern that is observed today. The top ten countries with the highest prevalence are in Europe, with Sweden among them in seventh place with a prevalence of 218 per 100,000 individuals, and incidence of 8.7 per 100,000 individuals.[5] A total of 766 people are diagnosed with MS in Sweden per year (Figure 1).

MS is most often diagnosed among females as compared with men: nearly 70% of people with MS are female.[3,7] This pattern is present in a multitude of countries worldwide including Sweden (Figure 2).[3,7,9] Historically, MS was previously thought to primarily affect males, with even a review of disseminated sclerosis from the 1930s stating that populations where females were affected to a higher proportion was "*owing to small numbers, this [sex difference] is probably of no significance*."[8] Today, the sex difference is large, and has steadily increased over time. The change in the sex demographic of MS has been hypothesized to be due to lifestyle changes among females,[6] but could also be due to increased recognition of women's health in general, and overall awareness of MS.

MS is often diagnosed between 20-40 years of age, with the average age at diagnosis in Sweden of 38 years.[5] MS poses significant burden and cost to individuals, society and health-care systems. It is diagnosed during a person's early adulthood resulting in substantial indirect costs such as loss of productivity and early-retirement.[10,11] Direct costs to the individual patient such as primary care, in-or out-patient hospital care, disease modifying therapies, diagnostic tests, nursing care, travel costs, and wages lost

to time off from work are also high. Annual costs increase over time with increasing disease severity from an estimated 20 000 Swedish krona (SEK) or approximately 2 000 Euro annually on low ends of disease severity, to 74 000 SEK (7 300 Euro) for more severely affected patients in Sweden.[10,12]

Figure 1: Sweden's MS epidemiological factsheet. Reproduced with permission from the MS International Federation. Sweden's Factsheet, The Multiple Sclerosis International Federation – Atlas of MS – 3rd Edition, downloaded 2023-04-26. These values were identified by using the Swedish National Patient Register and MS Register for all individuals with MS alive with MS or newly diagnosed up to Dec 31, 2018 in Sweden. Contributors to Sweden's data: Kelsi A Smith, Scott Montgomery, Peter Alping, Leszek Stawiarz, and Jan Hillert.

Figure 2: Sex proportion trends in MS in Sweden and globally. Additional country classifications based on the World Bank and WHO region country classifications. Reproduced with permission from the MS International Federation– Atlas of MS – 3rd Edition, downloaded 2023-04-26.

These direct patient costs are reduced in Sweden in comparison to countries without a comprehensive social welfare and health-care system, but with an average total cost of illness per patient of over 200 000 SEK (20 000 Euro) per year, societal costs are high.[13] Years lost to disability are high among people with MS[14] given their wide variety of symptoms requiring treatment and management in their daily lives. This is partly because life expectancy for individuals with MS has greatly improved[15] coupled with access to earlier and more effective treatments.[16] Despite improvements, people with MS still experience a lowered life expectancy of approximately 7-14 years[15,17] and increased disability not only due to their MS, but also due to their additional comorbid disease burden over time, occurring up to 10 years earlier than the general population.[18]

1.3 Understanding MS: Diagnosis, symptoms and consequences

MS is a heterogeneous disease, with varying presentation and symptoms among individuals. Yet, patterns of development, and presenting clinical course have been described succinctly and consistently with classifications today strongly resembling the classified clinical courses named 100 years ago.[1] In particular, some people with MS have an episodic re-occurrence of symptoms, but others instead progress directly.[1] New discoveries about MS in recent years may yet come to challenge the current understanding of the disease, with discussions surrounding MS moving towards describing it as a collection of syndromes, or of a spectrum of diseases.[19]

Differing clinical courses among males and females have been hypothesised, and even new ways of analysing cerebrospinal fluid propose new ways to classify MS.[20] These new approaches are interesting, though the most widely accepted disease course and severity classifications are included in this thesis. Additionally, understanding the

pathogenesis of MS is also becoming better understood with respects to timing of exposures, and even the addition of a prodromal period for MS is changing the research landscape and are explored in this thesis. The current perspectives are presented below that are relevant for setting the works of this thesis in context.

1.3.1 Diagnosing MS: McDonald criteria & disease courses

Several changes to the diagnostic criteria for MS have taken place over time with the current criteria being the 2017 McDonald criteria.[21] The diagnosis of MS is mainly an exclusionary diagnosis in which other possible causes are ruled out in conjunction with a combination of clinical, imaging and laboratory investigations. This combination of information has made it possible to diagnosis MS earlier, reduce misdiagnoses and facilitate timely access to treatment options and care. In the research context, changes in diagnostic criteria over time are important to consider when considering time-periods of pre-clinical activity, as well as in the severity of the disease. The timing of the diagnosis in particular with respects to start of DMTs plays an important role in the overall long-term outcomes of MS.[22]

Two main criteria for MS diagnosis are that findings have a dissemination in time and space, meaning that lesions in the CNS should have occurred at least twice (time) and should be in two distinct anatomical CNS areas (space).[21] In some cases where magnetic resonance imaging (MRI) is not possible (such as in countries where access to MRI are not possible, or lack of health-care resources for example) or in the case of baseline criteria for MS not being met, a patient's cerebrospinal fluid is additionally examined. Here, oligoclonal bands may be present in cerebrospinal fluid and have been found to be indicative of MS[21] especially in individuals presenting with clinically isolated syndrome. Diagnosis of MS is now possible from a first demyelinating event and clinical lesion, much earlier than previously used in other MS criteria where it was most often defined as clinically isolated syndrome.[19]

1.3.1.1 Clinical courses of MS

MS is typically classified into two major subtypes: relapsing remitting MS (RRMS), and progressive course.[21] Progressive course has two subtypes, primary progressive MS (PPMS) and secondary progressive MS (SPMS). The majority of MS patients present with RRMS at diagnosis with approximately 10–15% making up diagnoses of PPMS,[3] though as of 2018 in Sweden, 7% were diagnosed with PPMS.[23]

For years prior to diagnosis, individuals can experience temporary periods of neuroinflammatory activity that do not reach a clinical diagnosable threshold (Figure 3). These periods of neuroinflammatory activity manifest with an increase in severity of underlying MS that becomes clinically observable and diagnosable, and are called relapses. Despite signs and symptoms not crossing a clinically relevant threshold early in

MS development, MS disease mechanisms are underway, causing an accumulation of axonal loss within the CNS and decreases to overall brain volume, as well as disease progression.[24] The disease progresses until it results in an initial clinically observable and diagnosable episode of neuroinflammation, which may be diagnosed as clinically isolated syndrome if accompanied by MRI changes, or radiologically isolated syndrome if only MRI changes are observed in absence of a demyelinating event. At the second episode of neuroinflammation or evidence of lesions within the CNS visible on an MRI, a diagnosis of MS can be made, and if fitting a relapsing remitting disease profile, patients are diagnosed as having RRMS.[21]

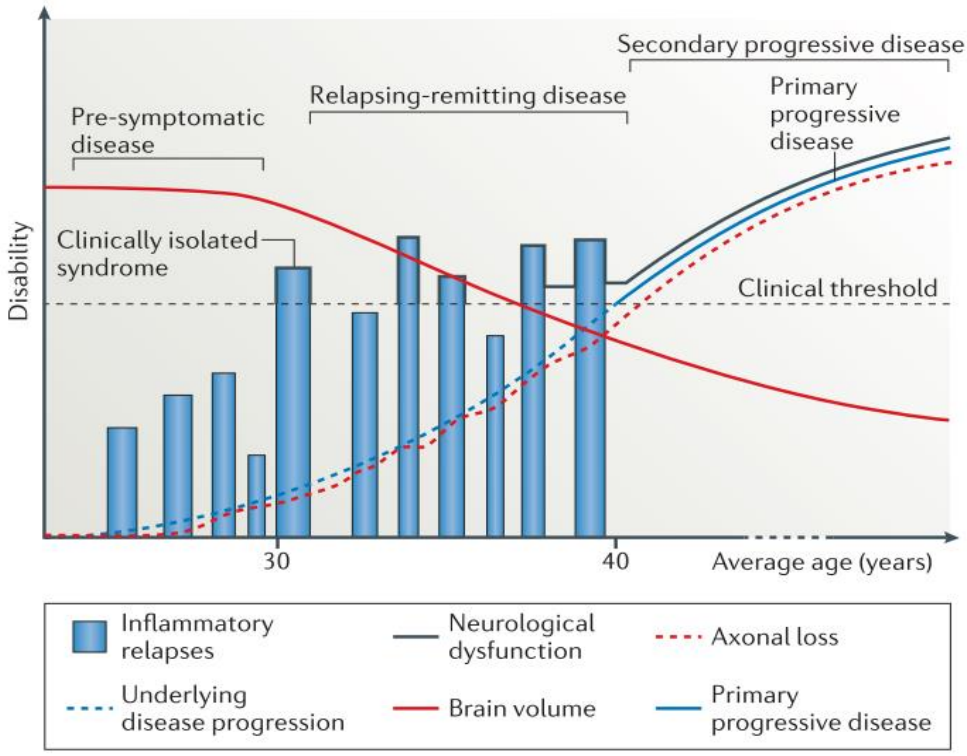


Figure 3: MS clinical disease course trajectories describing changes with increasing age and increases to disability due to inflammatory relapses, neurological dysfunction, axonal loss, and brain volume changes. Figure from Dendrou et al.[24] *Immunopathology of multiple sclerosis.* 2015. Nat Rev Immunol. 15(9):545-558 reproduced with permission from Springer Nature (SNCSC).

Many RRMS patients transition from a relapsing phase, to a progressive phased described as SPMS, that parallels aspects of the disease trajectory of individuals diagnosed with PPMS. The transition from RRMS to SPMS is possibly characterised by a reduction in neuroinflammatory markers within lesions in the CNS as a transition to neurodegeneration occurs.[24] A change in diagnosis is often made retrospectively, as it is difficult to measure when a relapsing type course seemingly switches to a steadier accumulation of disability outside of relapse activity.[25] Making the distinction from RRMS to SPMS is difficult and has

no uniform criteria.[26] Disability among people with SPMS increases as compared with RRMS, with additional brain volume and axonal losses. Few DMT treatment options exist for SPMS and PPMS, and for patients with severe MS severity. Males and females also seem to have different disease trajectories from one another. Females may experience more inflammatory activity as they typically have more relapses than males, and males more often have a progressive course with a faster accumulation of disability over time.[27]

1.3.2 Assessing the severity of MS over time

Measuring MS disease severity is typically measured using the Expanded Disability Status Scale (EDSS), which largely measures evidence of motor disability. Other clinical scales measuring fatigue, cognitive impairments, or gait impairments are also used, though they are not included in this thesis. This thesis used EDSS-based and related measures as they are available in the MS Register, and most consistently and uniformly used as an indicator of severity changes in MS clinics across Sweden.

1.3.2.1 Expanded Disability Status Scale

The EDSS was developed by John Kurtzke in the 1980s to measure and evaluate MS using a combination of information about overall function in a variety of functional systems, and walking abilities.[28] The original Disability Status Scale was developed in the 1950s, to which the functional system assessments were added to be able to identify possible differences in severity due to MS lesions. These functional systems describe different regions of neuronal networks and include aspects of pyramidal, cerebellar, brain stem, sensory, bowel and bladder, visual, and cerebral functions.

The EDSS comprises of a step-wise ordinal scale from 0-10 that increases by 0.5. A score of 0 describes no impairments and 10 is death from MS. The scale increases in steps of 0.5. Scores from 0 to 4.5 are made using a combination of functional system scores, and walking distances. Scores from 5-9.5 are primarily defined by various impairments to walking capabilities. Limitations of the EDSS score include that it is primarily based on motor symptoms that may only reflect part of the total disability or severity of MS over time. It is also somewhat subjective, and criticism has been made with respects to its inter-rater reliability.[29] Despite the limitations, it still is a useful tool in benchmarking changes over time that is relatively easy to administer within a busy clinic environment.

1.3.2.2 Age-Related Multiple Sclerosis Severity Scale

In addition to EDSS, the Multiple Sclerosis Severity Score (MSSS) was developed to correct EDSS with MS disease duration, as severity is likely to increase with increased time from MS diagnosis. The MSSS groups patients of a similar disease duration in which they are ranked and compared to derive a composite score.[30] A drawback of the MSSS is that disease duration is based on date of MS onset, which can often be missing in patient data, or unknown. Additionally, onset date is set retrospectively. Therefore, a similar measure,

the Age-Related Multiple Sclerosis Scale (ARMSS) was developed.[31] This score is generated similarly to MSSS, but instead of using MS disease duration, patient age is used. People with MS of similar ages are compared with respects to their EDSS, and ranked to give an ARMSS score.

1.3.3 Consequences of MS

Consequences of MS, similarly to the symptoms of MS, widely vary among individuals. Initial symptoms can remain and become more troublesome, increasing in severity over time. These consequences include increases in pain due to damage to sensory nerve tracts, to bladder and bowel incontinence, to cognitive impairments and fatigue.[16] People with MS also have an increased comorbid disease burden, developing additional diseases earlier than people without MS and a greater number of them. These comorbid diseases include renal disease, depression, and respiratory disease[18] that have direct implications for their long-term care and treatment plans.[32] One particular and commonly occurring consequence of MS explored directly in this thesis is spasticity.

1.3.3.1 Spasticity among people with MS

Spasticity is a common occurrence among people with MS. It is also a common complication among people who have experienced brain/spinal cord trauma due to external forces[33] or stroke.[34] In particular, spasticity is observed among people with general damage to upper motor neurons[35] resulting in upper motor neuron syndromes as seen in cerebral palsy[36] and as amyotrophic lateral sclerosis[37] among others, can also result in spasticity.[39]

The overlap of spasticity among a number of neurological and other conditions results in a highly variable clinical presentation, even among individuals with the same diagnosis and a similar level of functioning.[38,39] Muscle stiffness, tightness or cramping, clonus, flexor and extensor muscle spasms, and aching pains for example, are some of the symptoms. Varying levels of severity can occur even over a short period of time appearing unilaterally or bilaterally in the body, occurring most commonly in lower extremities, especially among people with MS, but are also often present in the upper body.[38,40,41]

Treatments for spasticity vary by country, but involve a multi-disciplinary approach[42] that include pharmacological and non-pharmacological treatments such as massage and exercise.[43–45] Approved first-line pharmacological treatments in Sweden include oral baclofen, diazepam, clonazepam, gabapentin, and cannanibinoids[46] though the latter is rarely used. Unresponsive spasticity can be additionally treated with second-line pharmaceuticals including tizanidine, dantrolene, intrathecal baclofen, or botulinum toxin A injections.[46]

The underlying pathophysiology of spasticity among people with MS is unknown. One main hypothesis supposes that spasticity is due to lesions within the CNS within upper

motor neuron pathways. This damage may increase in muscular reflex stretches and muscle tone in one or many muscles that is similar to spasticity observed within diseases resulting in upper motor neuron syndromes.[35,41] Other proposed mechanisms suggest that spasticity could be affected by immunological signalling that alters excitability and plasticity of neuronal signalling.[41]

Few studies of treatment patterns and the particular timing of treatment among MS patients have been conducted. Studies have mainly included prevalent MS patients,[38,40] been in the context of clinical trials[47–49] or assessed individuals with prevalent spasticity.[50]

1.4 MS pathogenesis: Immune system, prodrome and risk periods

MS is thought to be an immune mediated disease. It is marked by inflammatory markers in the various clinical courses[51] with infiltration of immune cells into the CNS through the blood brain barrier (BBB). This infiltration causes the axonal degeneration and loss, inflammation, and demyelination resulting to disruption in signalling and increases in disability.[24] MS is often considered an autoimmune disease, as lymphocytes launch attacks against CNS autoantigens, resulting in demyelination.[24] Lymphocytes involved in the pathogenesis of MS are both T- and B-cells, and although MS is typically described as a T-cell mediated disease,[16] mounting evidence of B-cell involvement is shifting this view in light of effective therapies for treatment of MS targeting B-cells.[52]

Lesions, or focalised plaques within regions of the CNS such as white and grey matter of the brain, spinal cord and the optic nerve, are zones in which demyelination has occurred due to breakdown of the BBB.[16,24] Specific mechanisms as to why the breakdown of the BBB occurs are incomplete, however evidence suggests that there is a strong dysregulation of the cells involved in maintaining the BBB, possibly due to increases in inflammatory cytokines.[16] This disruption results in the barrier becoming permeable, allowing migration of various activated T-cells, B-cells and macrophages into the CNS that normally should not be present.[24] Their presence and coordinated attack on self-tissues cause additional inflammation, demyelination, and axonal loss within the CNS, resulting in clinically observable MS. However, the BBB does not have to be disrupted in order for T-cells to cross, but they must have the necessary activation markers.

The activated T-cells that become present within the CNS are thought to become activated in the periphery of the CNS, although hypotheses exist about their ability to become activated also within the CNS[24] and are called the so-called outside-in versus inside-out hypotheses. Evidence of this peripheral activation has been observed in animal models of MS, such as the mouse model of experimental autoimmune encephalomyelitis (EAE).[53] In order to induce mice to exhibit EAE, an injection of activated T-cells into mice is performed. Odoari et al showed that despite these injected cells being activated, they do not directly cross into CNS and instead migrate first to the lung tissues such as the bronchus-associated lung tissues of the respiratory system.[53] Here, the T-cells remain

until they are reprogrammed to have a migratory profile that will allow them to cross the BBB resulting in MS-like lesions and activity.[53]

1.4.1 Prodromal MS

A prodrome is defined as a period in which an individual experience a combination of signs, symptoms, diagnoses or other findings that occur prior to onset of a disease. The term "prodrome" is more widely accepted in the context of MS today, as the widely accepted fact in the 1990's was that MS had no prodromal period.[54] The term prodrome would only come to be frankly used in conjunction with MS in 2010.[4] Especially in the past six years, the MS field has expanded as consistent evidence in a variety of countries show a measurable, marked prodromal period in the context of development MS.[55–57]

The prodromal period is proposed to be five to ten years prior to MS onset as there are changes to the health-related behaviour of future MS patients, such as increased visits to doctors, hospitals and pharmacies resulting in higher rates of hospitalizations and filled prescriptions as compared to a general population.[55] A variety of conditions have been reported, varying greatly individually, much like the MS symptoms themselves. Cognitive impairments, dermatological visits, a variety of mental health problems such as depression, increased pain, bowel issues, sleep issues, and even changes to birth control measures have been observed in the five to ten years prior to MS onset, within the prodromal period.[54,58] The pathogenesis of MS has already begun in the prodromal period, driving the appearance of initial symptoms.

1.4.2 Defining a risk period for MS

Given the pathogenesis of MS is underway during the prodromal period, it becomes crucial to consider risk factors outside of this period. Traditionally, many risk factors for MS were determined in the period directly prior to MS diagnosis or onset. In light of the MS prodrome, it is possible that many of the factors described as risk factors were assessed within the prodromal period or may even characterize the prodromal period.

A true risk period should occur prior to disease initiation, and thus prior to the beginning of prodromal MS and most certainly prior to the onset of MS (Figure 4). Given that MS has a variety of signs and symptoms that vary individually, including the length of time from initiation to clinically diagnosable MS, pinpointing a true risk period is difficult.

Several recent studies have taken approaches to minimize the cross-over of the risk period and possible prodromal period, and to also reduce the risk of reverse-causation (the fact that a risk factor may be associated to MS only due to prodromal MS, or that MS onset that is causing the symptom, rather than the symptom causing MS to develop). These studies have left a gap between the exposure period and MS diagnosis or onset in order to account for both the prodromal period and onset period of MS.[59–61] Additionally, the usage of register-based data that is not subject to recall bias of the individuals

participating in a study asked to recall information prior to their MS diagnosis also helps to better identify possible periods at risk.[54]

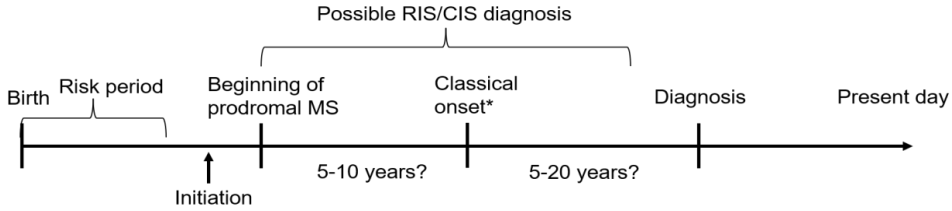


Figure 4: Major timepoints in the risk, initiation and development of MS. Abbreviations: CIS= clinically isolated syndrome; RIS=radiologically isolated syndrome. *Classical onset is what is considered to be the typical clinical onset in medical practice today. Onset year or date is often decided upon by the treating neurologist after discussions with the MS patient about their possible previous signs or symptoms of MS. The question marks and time periods denoted are estimations.

1.5 Risk factors for MS development and progression

Several known risk factors of MS have been determined, such as vitamin-D deficiency,[62] low sun exposure, smoking,[63,64] and infections, including infectious mononucleosis (IM) primarily due to Epstein-Barr virus (EBV).[60,65–67] MS is well known to disproportionally affect females three time more than males,[3] a pattern that begins to emerge in adolescence[14] with risk factors possibly affecting females to a larger degree than males such as vitamin-D deficiency and sun exposure,[68] although no definitive female-specific environmental or lifestyle risk factors are known. There are few sex-specific genetic reasons behind the disparity,[4,69] which points to the hypothesis of MS being caused by a combination of genetic, environmental and lifestyle factors. Specific risk factors, and the timing of risk factors relevant to this thesis are presented below.

1.5.1 Evidence of a critical risk period

Research in in recent years suggests the period of adolescence is emerging as a critical age period as some risk factors acting within this time greatly increase the risk of MS as compared to risk factors prior to, or after this developmental period. This key risk period, and in particular the ages between 10-20 years, seem to be especially of importance when considering individual susceptibility. A number of studies have identified a variety of environmental or lifestyle risk factors within this period.[60,70,71] Concussions in this period of susceptibility also have been shown to increase the risk of MS later in life by 200%.[71] Obesity in adolescence is also a known risk factor for MS,[70,72] with increases to body mass index (BMI) even within normal ranges that increase the risk of MS.[70] Adolescence obesity has also been shown to interact with underlying genetic susceptibility for MS.[73]

Furthermore, studies have shown that immigrating to a country that has a higher risk of MS, such as Sweden or Canada, as compared with countries closer to the equator increases the risk of MS, but only if migrating prior to adolescence.[4,74] Puberty and changes

in sex hormone levels have also been proposed as risk factors for MS, in particular the age of menarche.[9]

1.5.2 The role of infections

There have been long-standing hypotheses and evidence of infections being part of the cause of MS development. Viral infections in particular have been investigated for their role in MS development and include cytomegalovirus, EBV, human herpes virus-6 among others.[24] EBV in recent years has gained traction as a possible main cause of MS.[67] Though an infection of EBV alone is not a sufficient cause of MS, meaning other triggering factors in combination with EBV are necessary.

However, infections also play a possible role in MS progression as people with MS have a higher level of a variety of infections in general, but also those that require hospitalization. People with MS are more likely to have respiratory infections, urinary tract infections (UTI), skin infections and other infections[75] and infection burden among MS patients have been shown to be associated with disability and disease progression.[76] Respiratory infections for example, may increase relapses through similar mechanisms that affect MS risk by reprogramming immune cells to cross the BBB and increase CNS inflammation.[77]

As infections can be both causes and consequences of MS, particular attention should be paid attention to assessing these factors in the possible prodromal period of MS.[54]This thesis has tried to consider the timing of exposures surrounding diagnosis in order to establish a possible risk period outside of a prodromal period.

1.5.3 Genetic factors

Many studies have determined that there are genes influencing the development of MS. Clustering of MS is often observed among families: studies of families show that the majority of risk is observed among first degree relatives, although studies report conflicting magnitudes and ranges of relative risk estimates. Studies had shown that the children of parents affected by MS were 10-25 times more likely to develop MS than the general population,[4,78] but this was disputed by a more complete Swedish study. The large, national, matched cohort study in Sweden showed[4,82] that the relative risks were much lower, with parent-child relative risks lower than 10 times the risk of the general population.[79] Studies can agree that this risk sharply decreases with each increasing degree of relatedness, meaning that the greater the number of genetic factors shared with someone who has MS increases the risk of also one developing MS. This supports the involvement of inherited, genetic factors in MS development. It is unknown however, as to whether there is a larger maternal or paternal contribution in the transmission of MS, as studies have had conflicting results.[79]

Despite the familial aggregation observed in MS, MS has been shown to be only a partly heritable disease. Twin studies of MS are an effective way to determine heritability as by

studying the occurrence of MS among both monozygotic and dizygotic twins, the total proportion of MS risk that can be determined by genetic variation (heritability) can be assessed. Approximately 50–70% of MS risk can be explained by genetic variation,[79,80] the rest of the risk remains unaccounted for by genetic factors.

Familial studies and twin studies are only some of the methods that have been used to determine genetic influences on the risk of MS. Large-scale, multi-national genome wide association studies (GWAS) have also been conducted to identify specific MS risk genes. The most recent study to date nearly doubled the number of known risk variants by identifying a total of 233 MS susceptibility risk genes, including the first known variant on the X chromosome, the female sex chromosome.[69]

A combination of linkage and GWAS have shown that genes within the major histocompatibility complex, particularly genetic variants within the *human leukocyte antigen* (*HLA*) region, have the highest magnitude associations with MS risk. Evidence for the HLA region's involvement in MS is longstanding.[81,82] Individuals carrying *HLA-DRB1*15:01* have been shown to have nearly four times the risk of MS as compared to non-carriers, and individuals with *HLA-A*02:01* variants having a third lower risk of developing MS.[83,84] Overall, various genetic influences within the major histocompatibility complex suggest possible immune causes of MS; including the well-known involvement of various T-cell subsets and as well as more recently B cells, which may explain why some of the newer DMTs, such as ocrelizumab and off-label use of rituximab, that act to deplete B cells may be so effective.[52] However, it is not only immune-related systems that are involved. The majority of MS risk variants known to date are intergenic and are thought to have a regulatory function in genes with immune function or downstream effects in the immune system,[69] which demonstrates the large number of possible biological mechanisms that may involved in the development of MS.

Although a remarkable number of genes are associated with MS, they each confer a very low proportion of the total risk for developing MS. *HLA* genes confer the main genetic risk, but many other additional genes contribute to MS risk. This likely occurs in conjunction with lifestyle or environmental factors that increase underlying susceptibility, rather than one or two risk alleles or risk factor conferring the majority of the risk. This is also true of any one lifestyle or environmental risk factor, but many gene-environment, gene-lifestyle as well as gene-gene and environmental-lifestyle interactions may take place.[3,85]

Furthermore, no MS risk variants seem to impact MS progression or severity. Recently, a GWAS approach was used to identify associated variants to MS severity in more than 22,000 MS cases, which is the largest cohort to date. This GWAS also replicated their results at two loci, and found in particular that one variant was associated to MS severity as measured by ARMSS (author manuscript).[86]

1.5.4 Melanocortin-1 receptor gene, sunlight and vitamin-D

It is well known that the prevalence of MS worldwide varies among countries and regions. Countries such as Sweden and Canada have some of the highest prevalence rates of MS,[5,23] with countries within Sub-Saharan Africa and East Asia having very low rates.[6,87] The difference in rates may in part be due to underestimation caused by a frequent lack of diagnosing equipment and neurologists.[88] However, for many years this variation in distribution has been termed a "latitude gradient" corresponding to higher MS risk with increasing distance from the equator. In Sweden, a north-south gradient has been shown in MS[89] but also been shown in other countries[3] although whether it exists is debated.[6]

Individual ancestry as a risk factor has also been frequently discussed as individuals with Northern European ancestry have increased risk of MS. However, MS affects individuals of many populations, including individuals living at lower latitudes not of Northern European ancestry.[6] Large differences in living conditions, patterns of infections and ultraviolet radiation (UVR) between countries exist, possibly contributing to differences in risk and progression that are not easily generalizable by latitude. Additionally, the latitude gradient exists when considering prevalent cases, but not incident cases.[6] This implies latitude could not be a cause of MS as it follows a pattern of prevalent cases, whereas UVR and incidence of MS seems to have a stronger correlation.

Apart from the latitude hypothesis, and possible northern European genetic influences, alternative reasons for the distribution of MS worldwide is often discussed in the context of sunlight and vitamin-D. Decreased UVR is the most widely accepted hypothesis of the driving factor for the latitude gradient. UVR is an established risk factor for MS as it directly affects vitamin-D levels. Although both are risk factors for MS and are biologically connected, they have been shown to be independent risk factors, and [62,90] UVR has also been shown to suppress EAE.[91] Sunlight controls the production of vitamin-D within the body, and low vitamin-D has been repeatedly shown to be a strong risk factor for MS through EAE models,[92,93] but also through robust studies including mendelian randomization studies.[90,94]

One possible genetic explanation in risk differences among individuals could be due to differences in human pigmentation genes, such as the melanocortin-1 receptor gene (*MC1R*), and timing of exposures such as UVR. The MC1R is a receptor expressed on melanocytes in the skin controlling the production of melanin and therefore a person's tanning response to UVR (burning versus tanning). In the presence of UVR, a signalling cascade activates the switch from production of pheomelanin (lighter melanin) to eumelanin (darker melanin), resulting in the darkening of the skin for higher UVR protection and reduced skin reactivity to sunlight.[95,96] MC1Rs have been shown to also be expressed on various immune cells such as leukocytes and mast cells.[97,98]

Recently, studies have shown associations of *MC1R* single nucleotide polymorphisms (SNPs) with MS development and risk[99–101] and in MS progression.[102,103] Additionally, the MC1R has been shown to be neuroprotective in mouse models against neuro-inflammation and as a possible therapeutic target,[99] as well as could be responsible for protection against white matter damage seen in MS.[100] This relationship is conflicting as some evidence shows that *MC1R* SNPs contribute both to MS disease progression[104] and delays in disease onset.[105] Typically, people with lighter pigmentation have *MC1R* SNP variants responsible for their sun-sensitivity, however, individuals of darker pigmentation can also have these variants without having apparent sun-sensitivity.[106] Females also have more skin reactivity to UVR.[107]

The *MC1R* is highly polymorphic and SNPs in the *MC1R* can result in loss of function in MC1Rs,[108] resulting in UVR sensitive skin[97,109] which may potentially dysregulate the immune response, increasing individual risk of MS. Specifically the MC1R has been well-studied in the context of melanoma, and in particular SNPs in the *MC1R* called red hair colour (RHC) variants, are associated with increased UVR sensitivity.[106,109] People carrying these RHC variants are also likely to have red hair colour, however people may also carry the variants without having red hair.[95,97,108] People may also carry RHC variants that only have highly or somewhat reactive skin to UVR.[106] However, the relationship between *MC1R* and MS is complicated by the fact that the RHC variants are found in a possible overlapping region with another gene downstream of the *MC1R*, the tubulin beta class III (*TUBB3*) gene. Alternative splicing occurs producing a chimeric MC1R-TUBB3 protein[110] that can still be expressed. TUBB3 is a structural protein responsible for axonal and dendritic growth in the CNS.[111,112] Although *TUBB3* has not previously been associated to MS, its function in the CNS and overlap with *MC1R* warrant further investigation. Therefore, variants within *TUBB3* alongside *MC1R* have been included in this thesis.

2 Research aims

The overall aim of this PhD project was to further understand aetiological pathways in MS risk and complications. Each study's specific aims as well as a short summary of the background to each study are presented.

Study I – Respiratory infections and subsequent MS risk

This study aimed to investigate hospital-diagnosed infections of pneumonia as a risk factor for MS.

Rationale: Determining if there is an association of pneumonia with MS risk was motivated by the association of respiratory infections post-MS diagnosis that seem to exacerbate MS relapses. As increase in inflammatory activity in the lungs may be responsible for triggering an MS relapse, similar inflammatory activity may be a triggering factor for MS pathogenesis. Additionally, other infectious causes of MS have long been proposed, with EBV, cytomegalovirus, and herpes-zoster all possible infectious candidates hypothesized to act through molecular mimicry. To the best of our knowledge, this was the first study to specifically examine pneumonia infections as a possible risk factor for MS, and in particular to identify if there are sex-specific associations, or age-specific periods in which associations with MS are elevated.

Study II – Risk factors for spasticity treatment in people with MS

The aim of this study was to determine associations with spasticity treatment initiation and discontinuation among people with newly diagnosed (incident) and previously diagnosed (prevalent) MS. Baclofen is the most specific pharmaceutical treatment prescribed for spasticity among people with MS in Sweden, and here it is used to measure spasticity treatment patterns.

Rationale: Spasticity is a common occurrence among people with MS. Typically studies of spasticity have been conducted in the context of clinical trials for efficacy or safety of treatments, or in cross-sectional studies in which characteristics of patients requiring spasticity treatments were typically progressive-course patients and mainly males. Overall, characteristics of patients who receive spasticity treatment are not well understood, and in particular when MS patients receive spasticity treatment. Our study adds clinical understanding of the characteristics and timing of spasticity prescriptions among MS patients.

Study III –Melanocortin-1 receptor gene and MS risk

The specific aim of this study was to determine whether *MC1R* variants are associated with MS risk, and whether associations differ among males versus females. Additionally, given *MC1R* is directly involved in the skin's response to sunlight, we aimed to determine

whether any associations of the *MC1R* with MS risk are independent of vitamin-D, and sunlight. Furthermore, the shared region between *MC1R-TUBB3* was investigated to assess if one gene in particular is involved in MS risk.

Rationale: Sunlight, vitamin-D, and female sex have been shown extensively to be associated with MS risk, however mechanisms as to how these factors may be linked are lacking. Therefore, the focus of this study using a hypothesis-driven, genetic-based approach involving skin pigmentation genes and a combination of these above-mentioned factors was to further understand their possible role in MS risk.

Study IV – Melanocortin-1 receptor gene variants and MS severity

The aim of this study was to determine associations of the *MC1R* with MS severity as measured by ARMSS at different time-points. Furthermore, given sex-specific differences in progression and severity of MS, we aimed to determine sex-specific associations of *MC1R* with MS severity.

Rationale: This study was an extension of the *MC1R* risk study as the *MC1R* has been previously implicated in both MS risk and progression. Thus, expanding the risk study to the other side of the MS diagnosis was a natural continuation to understanding how *MC1R*, sunlight and vitamin-D may be involved in the MS disease process.

3 Methods

3.1 Research in the Swedish context

Researching MS in the Swedish context allows for a depth of data linkage that is expansive and can be used to understand MS from a multi-disciplinary perspective. The work of this PhD thesis was especially facilitated through Sweden's propensity to collect and record data on its residents, and of the tolerance and willingness of people to participate not only in register-data collection, but also in various studies. A combination of data including administrative and specialised population-based register data, individual lifestyle and environmental factor self-reported questionnaire responses, and genetic samples were used. This breadth of data is linked together through use of a personal identification number assigned to each resident of Sweden.

In this thesis, the focus of investigating causes and consequences of MS required assessing information about individuals before and after MS diagnosis. In this thesis, we examined factors and characteristics from birth to long after the MS diagnosis to assess complex questions. In this context, the ability to include a wealth of information over a long period of time is tremendously beneficial.

Register-based data is especially valuable as nearly the entire Swedish population is available for some measures. This is because Sweden has a universal tax-funded healthcare system that is accessible and affordable to residents. It provides coverage for all residents including in- and out-patient hospital visits, and pharmacological treatments, among others. What costs residents incur are reimbursed at different levels, helping to ensure that when health care is required, people do not abstain from seeking care. This is especially relevant in the context of studying MS, as signs and symptoms of MS can be observed long before clinical detectability and diagnosis. Healthcare access, and in turn the recording of health care visits in registers, goes hand-in-hand when studying complex diseases such as MS.

3.2 Data sources

Three types of data sources were included: register, self-reported questionnaires, and genetic. Several administrative and quality registers, meaning specialized disease-specific registers comprise the register-base data. Both the questionnaire and genetic data were obtained through several epidemiological studies of MS. An overview of the data used is shown in Figure 5.

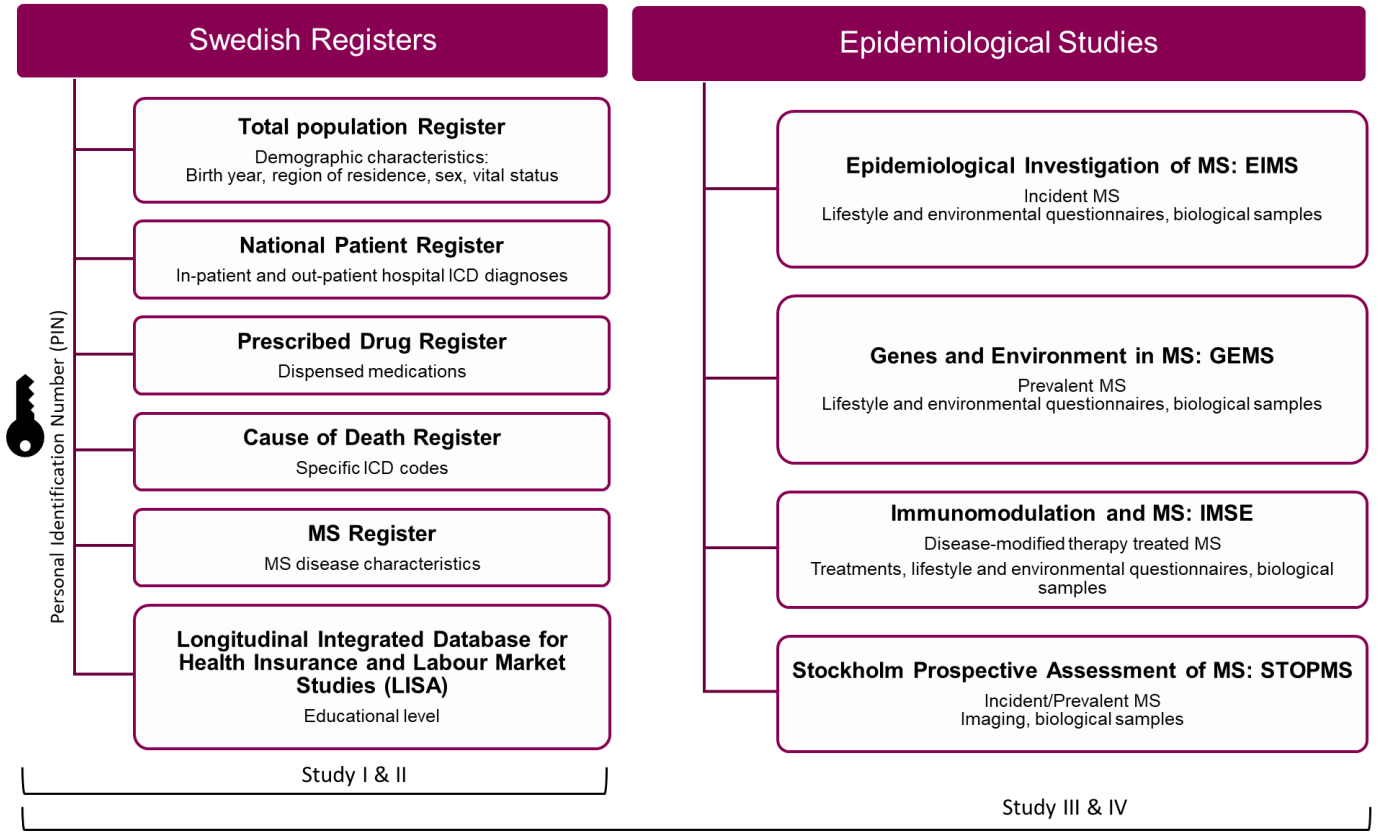


Figure 5: Sources of data included in Studies I–IV, all of which are linkable using the personal identification number. Study descriptions: Study I: Pneumonia as a risk factor for MS; Study II: Spasticity treatment patterns among people with MS; Study III: Pigmentation genes and MS risk; Study IV: Pigmentation genes and MS severity.

3.2.1 Register-based data

In particular, Swedish register-based data have been key to determine various exposures, outcomes and additional variables. Using a combination of information including hospital-based diagnoses, dispensed prescriptions, demographic information, and especially specific MS characteristics, has allowed study of exposures among a large number of people, over long periods of time.

The three key registers used in this thesis included the Swedish MS Register, the National Patient Register, and the Prescribed Drug Register and are described in detail below. Register-based sources of demographic data are also described.

The Swedish MS Register

The Swedish MS Register was developed specifically to monitor and track MS patients and treatments in the clinic and for research. The MS Register's goal is to optimize health care for MS patients.[113] Informed consent is required for inclusion, and both treating neurologists and patients can voluntarily report data, though patients can only report restricted measures. Participation in the register is not mandatory and data are not automatically collected from clinical charts. Approximately 18,000 people, a total of 85% of the population with MS in Sweden, are registered and data include diagnosis and onset dates, disease course, disease severity, DMT use, relapses, and other clinically relevant data since 1996. It first covered patients in specific regions such as Stockholm and Gothenburg prior to receiving public funding and being formalized as the Swedish MS Register in 2001.[114]

Despite its clinical and research-based usefulness, data have only been recently validated on a large-scale basis. A validation study found that the measures used in this thesis such as DMTs, relapse and EDSS had high accuracy as 91-99% of these values were confirmed between the patient's medical records and the register values.[115] The amount of missingness was higher for relapses, and EDSS, but moderate for DMT uses (35%, 14%, and 5% respectively). Missingness overall was greater for patients using older therapies such as glatiramer acetate or interferon-beta, but when updates were completed for EDSS, the overall mean values for EDSS remained similar as the mean prior to the update. This suggested missingness at random rather than systematic differences.

In this thesis, the MS Register was used to identify the first signs and symptoms of MS, as well as to effectively determine consequences of MS, including spasticity and disease worsening, measured using increases in EDSS scores. The MS Register data, including dates of onset and diagnosis, was used in combination with hospital-based visits where MS was diagnosed, to be able to establish the earliest signs of MS. This was in order to be able to assess possible risk factors for MS outside of a potential prodromal period of MS.

National Patient Register

The National Patient Register captures in-patient hospital visits from 1964 with national coverage beginning in 1987. Out-patient visits to specialised care were added to the register in 2001 and to date the estimated coverage is 99%.[116] The register records primary and secondary discharge diagnoses, using Swedish International Classification of Disease (ICD) codes. The register has been previously validated, with a variety of diseases having a high positive predictive value and overlap with specialised disease registers.[116] Some diagnoses, in particular psychiatric diagnoses, were added only from 1973.

The National Patient Register is complementarily used in several of the studies of this thesis in order to capture as many people with MS as possible. As participation in the MS Register is voluntary, it covers approximately 85% of the MS population, the remaining individuals can be identified by using ICD codes specific for MS. The National Patient Register is an administrative, obligatory register, and therefore individuals receiving hospital care who have MS can be identified through discharge diagnoses. Additionally, there is sometimes discrepancy between the MS Register and Patient Register with respects to diagnosis date, or missing diagnosis date in the MS Register, which can be complemented by the Patient Register. Not only can the Patient Register be used for MS diagnoses, but also for exposures including pneumonia or control diseases such as UTIs in Study I, or for comorbid diseases as possible influencers of spasticity treatment in Study II.

Prescribed Drug Register

Prescribed and dispensed medications in all pharmacies in Sweden are recorded in the Prescribed Drug Register from July 2005.[117] Data are classified according to Anatomic Therapeutic Classification (ATC) codes and includes all pharmacological treatments with the exception of those administered in hospital such as DMTs. Delays in data registrations in the first year after the register's establishment occurred and therefore it is key to have a waiting period of several months after the start of the register to ascertain new dispensations. Additionally, it is difficult to separate prevalent and new prescriptions within the first year of the register, therefore a look-back period typically of one year is used to determine if individuals have a treatment within this period to reduce the possibility that prevalent prescriptions are classified as new-use.

The purpose of this register in the Studies of this thesis was to assess specific pharmacological drugs used spasticity treatment in Study II, as well as treatments for diseases possibly influencing spasticity including Parkinson's disease and depression. The prescription is only recorded in the register once dispensed. This means it is not possible to see what drugs are prescribed that have not been dispensed. Typically, conditions diagnosed and treated within primary care are captured through the Prescribed Drug

Register, as a nationwide primary care register is currently not available. Some conditions such as spasticity or depression are often encountered within the primary care setting that would not be elsewise measurable without the Prescribed Drug Register. Very serious cases may be captured by hospital-records, but this would underestimate these conditions.

Registers for demographic data

The Total Population Register contains demographic information such as sex, region of residence and vital status of the entire resident population in Sweden. Data has been recorded from 1968 onwards.[118] The Total Population Register is also a source from which individuals from the general-population can be identified and matched as controls for MS patients, as was done in Study I (pneumonia) and III (pigmentation genes and MS risk). Migration and date of death information is also available in this register, which is of value when determining follow-up time or time-specific exposures or outcomes in all of the studies. Date of death can also be complemented by dates recorded in the Cause of Death Register which started in 1961.[119]

The Longitudinal Integrated Database for Health Insurance and Labour Market Studies (Swedish acronym LISA) records the highest achieved educational level per year for the Swedish population.[120] This register officially began in 2001, but information is available from 1990 as this data had already been collected using censuses. Highest attained education was used as a proxy in Study I and II for socio-economic position, as this could have influenced the association between pneumonia and MS, as well as factors influencing receipt of spasticity treatments. Highest attained education could be an indicator of the family-home life, as children of more highly educated parents tend to also be more highly educated. Educational level could influence health-seeking behaviours with respects to hospital-treatment for pneumonia, as well as being diagnosed with MS, or even seeking treatment for spasticity.

3.2.2 Epidemiological study data

Specific studies of the epidemiology of MS that have been undertaken at Karolinska Institutet have been an integral part of this thesis. An overview of each study is provided in Table 1 and includes the Epidemiological Investigation of MS (EIMS), the Genes and Environment in MS (GEMS), the Immunomodulation and Multiple Sclerosis Study (IMSE) and the Stockholm Prospective Assessment of Multiple Sclerosis (STOPMS). The purpose of EIMS, GEMS and IMSE is to study risk factors for MS, genetic risk factors for MS, and the effects of immunomodulatory drug treatments for MS, respectively. The fourth study, STOPMS, consists only of people with MS or possible MS. Its purpose is to identify methods for early diagnosis, determine prognosis and treatment effects through biomarkers. Specific details of each epidemiological study are provided below.

Table 1: EIMS, GEMS, IMSE and STOPMS - Epidemiological study designs at a glance.

Study	Purpose	Time-period	Case recruitment	Control recruitment	Information collected
EIMS	Identify risk factors for MS	2004-2021	*Incident MS*; included within two years of MS diagnosis. Enrolled at university hospital-based neurology clinics in Sweden.	Population-based controls: Two randomly selected individuals from the Total Population Register matched by age (5-year intervals), sex, vital-status and region of residence to cases at *MS diagnosis*.	Blood sample, lifestyle and environmental exposure-based questionnaire asking for information 5 years prior to MS diagnosis.
GEMS	Determine genetic aspects of MS	2009-2011	*Prevalent MS*: selected from the MS Register.	*Population-based control*: One randomly selected individual from the Total Population Register matched by age, sex, vital-status and region of residence to cases at *MS onset*.	Blood sample, lifestyle-based questionnaire similar to EIMS, but with broader time-periods for lifestyle and environmental exposure information.
IMSE	New DMT effects	2006-Current	*Incident and, prevalent MS initiating new DMTs*. First recruitment DMT was natalizumab, expanded to other DMTs as they became available.	*Primarily patients with MS*: later expanded to a subset of MS patients matched 1:1 with population-based controls (age, sex, vital status and region of residence matched).	Blood samples, and for select individuals, lifestyle-based questionnaire similar to GEMS.
STOPMS	Identify biomarkers for prognosis and treatment effects	2009-Current	*Newly diagnosed with MS, CIS, or new DMT initiators*. Living in Stockholm Region.	*Primarily patients with MS;* later expanded to include a subset of neurological-based controls (individuals undergoing investigation for other neurological symptoms unrelated to MS).	Blood sample, cerebrospinal fluid sample, MRI, clinical tests.

Study abbreviations: EIMS: Epidemiological Investigation of MS; GEMS: Genes and Environment in MS; IMSE: Immunomodulation in MS; STOPMS: Stockholm Prospective Assessment of MS.
Other abbreviations: CIS: Clinically isolated syndrome; DMT: Disease modifying therapy; MRI: Magnetic resonance imaging.

EIMS: Epidemiological Investigation of Multiple Sclerosis

EIMS is a case-control study that recruited consenting, newly diagnosed people with MS within two years of their MS diagnosis aged 16-70 years from hospital-based neurological clinics in Sweden from 2004-2021. Cases are individually matched to two randomly selected population-based controls from the Total Population Register by age (in five-year age strata), sex, vital-status and regions of residence at the time of MS diagnosis. If the control did not consent to participate or did not respond, another control was randomly selected for the case within a year of case identification. Cases and controls were still kept within the study regardless if the matched case or control were excluded at a later stage.[121]

Consenting participants were asked to leave a blood sample, and fill out a questionnaire on various lifestyle and environmental factors. Cases received the questionnaire at the neurology clinic where they were diagnosed and were asked to provide a blood sample at the same time. Controls received questionnaires, a blood sample kit, and a return envelope by mail. All questionnaires were returned by mail. Participants returning uncompleted questionnaires were asked to complete them by mail or telephone (if they wished to do so), and up to four reminders were sent by mail to non-responders. MS diagnosis information was updated in study data using the MS Register.

Data were manually entered into the study database for all participants by research coordinators for earlier versions of the questionnaire (as there were three versions of paper questionnaires used) and subsequently were returned digitally (two additional questionnaire versions, updated from the paper versions). Participants who wished to fill a paper-copy could request to do so.

GEMS: Genes and Environment in Multiple Sclerosis

GEMS, like EIMS, is also a case-control study. GEMS recruited prevalent people with MS that had not been included in EIMS between 2009 to 2011. People with MS 18 years of age or older who had consented to participate in the MS Register were selected. One individually matched, general-population control was selected from the Total Population Register for each case. Cases and controls were matched similarly as in EIMS (five-year age strata, sex, vital status, and region of residence) though at MS onset. Selected participants were asked to participate and received a questionnaire and blood sample tubes by mail, and were returned by mail. Reminders to participate were sent by post, three times for the questionnaire and twice for the blood samples.

The completed questionnaires were machine-scanned and stored in the study database. Only one questionnaire version was used in GEMS. Recently a quality-check of all questionnaire data stored electronically were checked against respondents' responses

from the returned paper questionnaire. Any and all errors were corrected to match respondents' answers. This updated questionnaire data was used in Study III and IV.

IMSE: Immunomodulation and Multiple Sclerosis

IMSE first started in 2006 as a post-marketing surveillance study following people with MS who were treated with natalizumab. Biological samples including blood and cerebrospinal fluid were collected. Originally, comparators were other individuals with MS with a more benign course.[122] From 2011, IMSE has since been expanded to several DMTs as they became available, as well as including an IMSE questionnaire (two separate versions used within IMSE) for select participants. The IMSE questionnaire study was similarly designed as EIMS and GEMS with population-based controls for each case and similar lifestyle exposure questions. Inclusion criteria for IMSE was any individual starting a study drug of interest. Individuals could be included in IMSE several times if starting a new therapy, but individuals were only asked to fill a questionnaire once. Cases who had not already responded to EIMS or GEMS were preferentially selected to complete an IMSE questionnaire.

STOPMS: Stockholm Prospective Assessment of Multiple Sclerosis

Individuals with MS or possible MS from the Stockholm Region were asked to participate at several participating neurological clinics. Individuals included are either: newly diagnosed individuals with MS; people with clinically isolated syndrome; or new DMT initiators. Individuals consenting to participate have specifically scheduled clinical check-ups, functional tests, imaging tests such as MRI, as well as collection of blood and cerebrospinal fluid. This allows for analysis of imaging and biomarker findings in relation to clinical progression and severity score changes. Several years after the start of STOPMS, control individuals were added to the study. These control individuals were individuals undergoing cerebrospinal fluid collection as part of a clinical investigation for neurological disorders other than MS.

3.2.3 Questionnaire data from EIMS, GEMS and IMSE

Questionnaires for EIMS, GEMS, and IMSE were based on the original EIMS questionnaire from 2004. As noted above, the questionnaires over time have had several versions. Additionally, although the questionnaires are similar to one another from EIMS, GEMS and IMSE, the time-periods in which individuals were asked to self-report their data, varied. For example, the majority of the questions in EIMS were designed so individuals provided exposure information in the five years prior to MS diagnosis, at the time of their MS diagnosis. At the time of study conception, the prodromal period that is now widely accepted in MS, was thought to not exist. Additionally, evidence of adolescence as a critical-risk period in MS susceptibility also emerged in later years and not widely described

in 2004. In more recent versions of the EIMS questionnaires, additional time-periods were added to account for the changes in understanding of MS. In GEMS and IMSE questionnaires, age-groupings or an age-specific reporting periods were used instead, as these study questionnaires were constructed in 2009 or later and could capitalize on the changing research landscape of MS factors influencing MS susceptibility.

However, due to the multiple versioning of the questionnaires, differences in questions are not only found between studies but also within studies from version to version. Time-frames from MS diagnosis to filling out questionnaires also vary and present a wide-range of challenges when harmonizing the questionnaire data. Originally planned was the inclusion of variables such as individual sunlight exposures, and vitamin-D supplementation. Typically, other studies using this data have separately analysed EIMS and GEMS data for example, or used it among only EIMS participants. As we involved SNPs with small effects, it was necessary to include all individuals. When we began to harmonize data, and in particular for the sunlight exposure questions, large differences in how questions were asked made it difficult to reduce the complexity of the data as some questions were more in-depth and comprehensive than others in different questionnaire versions. Eventually, a decision was made to use more objective measures such as a vitamin-D genetic risk score in Study III. Other questions encompassing items such as depression or mood disorders, were replaced with data from the National Patient Register or Prescribed Drug Register or educational information from LISA. This made sure that variables among included individuals were assessed in a uniform way. The ability to combine and use different types of data demonstrates the uniqueness and power of these large-scale data linkages encompassing both questionnaire data and register-based data in order to complement one another in providing as complete and complex information at an individual level.

One must be careful when interpreting exposure status when using self-reported questionnaires. Recall bias is a type of systematic error that can be introduced especially if participants do not remember events, or their likelihood to remember is influenced by their disease status. Potential differences that arise in recalling exposure events or possible confounders among people with and without MS can influence possible associations between exposures and MS. Therefore, the most objective questions in the questionnaires were used in this thesis work, which included BMI reporting, age, possible date of MS diagnoses, and immigration information.

3.2.4 Genetic data from EIMS, GEMS, IMSE and STOPMS

This section provides an overview of the genetic data used in Study III and IV. Table 2 provides an overview of relevant genetic terminology.

Table 2: Overview of genetic terminology.

Term	Definition
Allele	Variations of the same sequence of nucleotides (or a single nucleotide) in a specific location in the genome. At a specific location in the genome, there are two alleles (one inherited from each parent). If the inherited alleles are the same, they are homozygous. If different from one another, heterozygous.
Hardy-Weinberg Equilibrium	Considered a principle of genetic variation. States that genetic variation in a given population will remain stable over one generation to the next due to random mating. Genotype and allele frequencies will be constant in the absence of evolutionary forces such as mutations, selection or non-random mating. This is an ideal state, and departures can still occur.
Haplotype	A group of genes, or stretches of DNA, that are inherited together from a parent.
Linkage disequilibrium	The tendency of two alleles of separate genomic regions (loci) are found more commonly together than would be expected by chance. This means that alleles of one locus may not be independent of another.
Locus (singular) or loci (plural)	A specific location on a chromosome in the genome.
Minor allele frequency (MAF)	How frequent the least common allele is present in a specific population. An allele appearing less than 5% are considered to be uncommon, and less than 1% are considered rare variants.
Principle component (PC)	See principle component analysis.
Principle component analysis (PCA)	A statistical method used to identify population structure of individuals' genotypes. For example, genetic variation can be present due to geographical variation or ethnic background that can be captured in different principle components (PCs).
Single nucleotide polymorphism (SNP)	Variation in the genome of one base-pair. Gives rise to different alleles at a locus.

3.2.4.1 Genotyping

The biological samples collected from each of the above-mentioned studies were used to extract deoxyribonucleic acid (DNA) and subsequently genotype all individuals, which was necessary for determining the different SNPs used in Study III and IV (pigmentation genes and MS risk or severity). For the purposes of Study III and IV individuals had been genotyped using a GWAS specific array: the Illumnia Human OmniExpress beadchip array. This array has been used in GWAS of MS[86] and been previously extensively described.[123]

A GWAS specific array means that is has been designed to capture SNPs across the genome, typically spread out evenly at "tagging" locations in order to adequately capture variations. These tags in the genome are specifically chosen for even coverage, but also in

order to take advantage of linkage disequilibrium. Thus, tagging one SNP that is in high linkage disequilibrium with another portion of the genome will also capture that un-tagged variation and allow for high-quality imputation. The OmniExpress tags a total of 715,322 SNPs, of which 695,789 are autosomal variants.[124] Together, these tagged SNPs capture several million SNPs covering the linkage disequilibrium region, and together the whole genome. This large number of variants provides a good backbone on which imputation of un-tagged variants can take place.

Once genotyping was finished, extensive additional quality control was necessary.[125] As the differences between MS cases and non-MS controls with respects to their allele frequencies are of interest to determine possible associations with disease, artefacts resulting from quality errors can create spurious results.[126] The overall quality of the array helps inform the quality and coverage of the genotypes that are produced, but the specific markers must be robustly quality checked to exclude abnormalities. Quality control for each sample was performed, and included the overall sample quality in which genotyping success was greater than 98%, detection of possible contaminated samples by assessing heterozygosity, and assessment of genetic relatedness through identify by decent. Marker quality control was also assessed and encompassed these aspects: marker missingness was less than 2%, variants were removed if they deviated from Hardy-Weinberg Equilibrium with a significance threshold higher than 10^{-6} or if minor allele frequency (MAF) was less than 1%. Finally, to ensure comparability among cases and controls, differential missingness of specific variants was assessed for the variants of interest.

3.2.4.2 Imputation of genotypes

Genetic imputation capitalises on the inherent linkage disequilibrium structure and haplotype nature of the genome in order to calculate the probability of a specific genotype at a specific location. As long stretches of DNA can be inherited (haploid) in a given ancestral population, imputation takes advantage of this to predict non-tagged variants to different degrees depending on the strength of the linkage disequilibrium between tagged and untagged markers.[127] In our Swedish population, a reference panel from the Haplotype Reference Consortium is used in order to reliably impute.[128] This means that stretches of DNA in an individual are compared to this reference panel. Genotyped SNPs are matched to the reference in order to calculate the probability of imputed SNPs in the regions of interest. The stronger the linkage disequilibrium between markers of alleles, the better the prediction. A quality score is produced indicating the reliability of predicted SNPs.

3.2.4.3 Accounting for population stratification effects

Importantly, prior to imputation, underlying genetic population stratification must be considered. This stratification occurs when there is a difference in the frequency of alleles

between cases and controls that is due to ancestry or geographic location rather than any true gene-disease association.[129] This may cause spurious associations if not removed. A principle component analysis (PCA) is used to identify population stratification effects and calculate principle components (PC). This statistical method has applications in other fields, but in genetics it is used to identify variability within the data, mathematically reduce this stratification and calculate it into PCs.[130]

Another method for estimating population stratification is calculation of a genetic relatedness matrix (GRM) that is used often in mixed-linear model approaches[131] (discussed further below). The GRM is partially related to the PCA. The GRM is calculated based on imputed data, and can be calculated for an overall population, or specific populations such as males and females. The GRM calculates on an additive scale how genetically related individuals are within the study population in a pairwise manner and accounts for this in the context of a disease.[132] It compares individuals using a set of SNPs that are prespecified, in Study III and IV, markers were used across the genome. If individuals are very genetically similar (i.e. related), a GRM will help to identify these individuals for their removal, but also allow for examination of specific genetic differences. The purpose of the GRM is to adjust the analysis to prevent confounding due to population stratification, similarly to correction using PCs. Individuals who are distantly related or from diverse ancestries introduce complexity into the GRM; as these genetic differences could be attributed to population structure. Associations produced could be spurious, as they potentially will arise due to the underlying population structure rather than an association with disease status. Some stratification effects may arise from immigration, as ancestry-specific genetic variants contribute to differences, but Sweden also has an indigenous population that is genetically distinct from other populations.[133] These effects have been accounted for in this thesis work in Studies III and IV.

3.2.4.4 Genetic risk score

Genetic risk scores (GRS) are a useful tool to combine the information from multiple independent loci that may be involved in disease weighted by their effect. Often, GWAS methods identify a wide-range of SNPs in different loci of the genome that contribute a small individual effect in the risk of a disease. Usually, the top SNP representing each locus is chosen and these SNPs can be used in a GRS to summarize an individual's underlying genetic risk of a disease. The cumulative contribution of different SNPs is calculated by summing the number of risk alleles (0, 1 or 2 also called dose or genetic load) for each disease-associated SNP. Each SNP can also be multiplied by the estimated genetic effect from a GWAS to weight its importance in the disease, also called a weighted GRS.[134] In Study III, a weighted GRS was used to calculate underlying genetic risk of low-serum vitamin-D levels.[135]

3.3 Study specific methods and statistical methods

The specific methods and study designs including exposures and outcomes used in this thesis are summarized in an overview in Table 3.

3.3.1 Study I: Pneumonia as a risk factor for MS

3.3.1.1 Overview of case-control study design

This study (along with Study III described below), used a case-control design. Case-control studies can improve study efficiency and power, especially in the context of MS. Given MS has a long pre-clinical period and takes a number of years to manifest, following a number of people over time to see if they develop MS may be impractical. This can be mitigated in part by taking a sufficiently large sample of people to ensure that there is a possibility that some will develop MS, and when using registers this is often possible. Given MS is an uncommon occurrence despite its higher prevalence in Sweden, this still may not be sufficient to capture people with MS. A solution is then to identify people with MS, and include them into the study while also recruiting controls for comparison. In recruitment of study participants, how and where they are recruited matters to avoid selection bias.[136] Using the Swedish registers alleviates this selection bias, as nearly all possible individuals with MS in Sweden are able to be selected, especially when using the MS Register and the National Patient Register together as further described below. Individuals who did not consent to inclusion in the MS Register still are eligible for selection if visiting a hospital from 1964. Although Sweden has hospitals across the country, regional differences in access to hospital are present. In particular, people living in the Northern areas of Sweden have further to travel to reach a hospital.

Selection of controls is equally as important in a case control study, and randomly selected, general-population-based controls as used in both Study I and III are highly desirable. Due to the large sex-discrepancies in MS, regional differences with respects to access to healthcare and frequency of MS, and that MS is often diagnosed between 20-40 years of age, completely randomly selected individuals from the general population may not adequately reflect the distributions of the possibly confounding demographic characteristics of the MS population. This has been mitigated in Study I and III by matching on these characteristics. The role of matching is to ensure similarities on demographic characteristics, but also to gain statistical efficiency in analysis of the possible exposure-outcome relationship.[137]

Table 3: Thesis study design aspects including exposure and outcome definitions, data sources and statistical analyses.

			Main exposure		Main outcome			
Study	**Design**	**Study period**	**Definition**	**Assessment**	**Definition**	**Assessment**	**Data sources**	**Statistical methods**
I: Pneumonia as a risk factor	Matched case-control	1968-2012	Pneumonia diagnosis in childhood or adolescence	Categorisation of age-groups: 0-5, 6-10, 11-15, 15-20. Any ICD codes of the following: ICD 7: 490-493; ICD-8 and 9: 480-48); ICD-10: J12-J18.	MS diagnosis after age 20 years.	Earliest date of MS diagnosis (as recorded in MS Register, or National Patient Register diagnoses using ICD-8 or 9: 340, ICD-10: G35)	*Exposure*: National Patient Register *Outcome:* Diagnosis date from MS Register or National Patient Register	Conditional logistic regression
II: Spasticity treatment patterns	Cohort	2005-2014	MS diagnosis	Earliest date of MS diagnosis (recorded in MS Register or National Patient Register using ICD-8 or 9: 340, ICD-10: G35). At least two MS diagnosis required in National Patient Register.	Baclofen initiation and discontinuation	Baclofen initiation: dispensed prescription after look back period: Baclofen ATC code: M03BX01 Baclofen discontinuation: no dispensation after 90,120, or 180 days.	*Exposure:* EIMS, GEMS, IMSE, STOPMS participants *Outcome:* Prescribed Drug Register	Time-to-event using Cox regression models with time-varying covariates and failure functions
III: Pigmentation genes and MS risk	Matched case-control	2004-2018	*MC1R* SNPs, stratified by sex	Direct genotyping and genotype imputation	MS diagnosis	Any MS diagnosis (as recorded in individual study data, MS Register, or two or more National Patient Register diagnoses using ICD-10: G35).	*Exposure:* EIMS, GEMS, IMSE, STOPMS blood samples *Outcome:* EIMS, GEMS, IMSE, STOPMS study data, MS Register	SNP analysis: Mixed linear models adjusting for genetic relatedness; SNP independence: logistic regression.
IV: Pigmentation genes and MS severity	Cross-sectional	2004-2018	*MC1R* SNPs stratified by sex	Direct genotyping and genotype imputation	First or last ARMSS score of people with MS	Derived from age at EDSS and score	*Exposure:* EIMS, GEMS, IMSE, STOPMS blood samples *Outcome:* MS Register	Mixed linear models adjusting for genetic relatedness

Abbreviations: ARMSS: Age-Related Multiple Sclerosis Severity Score; ATC: Anatomical Therapeutic Chemical; EDSS: Expanded Disability Severity Score; EIMS: Epidemiological investigation of MS; GEMS: Genes and Environment in MS; ICD: International Classification of Disease; IMSE: Immunomodulation and MS; *MC1R:* Melanocortin-1 receptor gene; SNP: single nucleotide polymorphism; STOPMS: Stockholm Prospective Assessment of MS.

3.3.1.2 Study population and assessment of exposure and outcome

This study matched cases and controls in a ratio of 1:10 MS cases to controls matched by age, vital status, sex and region of residence. Although statistical efficiency is typically not increased after a matching ratio of 1:4,[138] in our study a larger matching ratio was necessary. This is because we had a high number of inclusion criteria, were interested in sub-groups of individuals, and wanted to perform sensitivity analyses to ensure the robustness of the results. MS cases were identified through a combination of individuals recorded in the MS Register (any record) and the National Patient Register using repeated occurrence of MS-related ICD codes (Table 4). The earliest date of MS diagnosis from either register was used as the matching year. MS cases were included so long as they were diagnosed after age 20 years for the primary analyses, and over 30 years of age for the secondary analyses. Individuals were included if they were born after 1964 to have complete information from birth when determining diagnoses given the Patient Register began in 1964. A flow chart denoting included/excluded individuals can be found in the Study I, Figure 1.[139]

The main exposure was a diagnosis of pneumonia in different age-groups throughout childhood and adolescence, with the outcome of MS diagnosed over 20 years of age. A control disease of UTI was used, as like pneumonia, UTIs are a known complication of MS.[140] A control disease is necessary as there may be a risk of reverse causation where prodromal MS or increased MS susceptibility prior MS diagnosis may increase the risk of infections. If this was indeed the case, then both UTIs and pneumonia may show associations to MS. There are primary and secondary discharge diagnoses recorded in the Patient Register, with typically the primary being the main reason for the hospital visit. Any diagnoses of pneumonia and UTIs were used in main analyses, but in order to reduce the possibility of misclassification, sensitivity analyses using primary pneumonia diagnosis were additionally used.

Possible confounding factors were carefully considered, and included highest attained education, as well as diagnoses of IM, an infection as a result of EBV, which is a strong infectious risk factor for MS.[67] Although sex was a matching variable, we wanted to also determine if its effects were sex-specific. In order to do so, a fully sex-stratified analysis was necessary, however there were too few of each sex in each exposure group to be able to draw robust conclusions.

3.3.1.3 Statistical analysis using conditional logistic regression

Conditional logistic regression was used to determine the association between pneumonia and MS, while adjusting for the possible confounding factors named above. Logistic regression using the maximal likelihood approach is preferred over other methods such as the Mantel-Haenszel approach as we are using a number of adjustment variables in addition

to the matching variables.[141] Logistic regression models the probability of an outcome occurring, such as MS, by estimating the log-odds of the outcome using a linear combination of independent variables. Odds are measured by the probability of the outcome divided by the probability of not having the outcome, and we can compare groups by using the odds ratio (OR). Conditional logistic regression is an extension of the logistic regression model in which matching characteristics are used as a stratum.[137] A term for matched stratum is introduced as a constant (provided there are no interactions), while estimating a coefficient for the variable of interest, and additionally conditioning over the number of cases in each stratum. This produces a log likelihood for each stratum of the disease, and summing over each stratum likelihood produces an estimation of the effect of the exposure on the outcome that maximizes the conditional log likelihood.[142] Unconditional models were also tested by including the matching variables, and no differences were observed in the results using conditional versus unconditional models.

Table 4: Study I design considerations with respects to exposure and outcome definitions, including additional variables in specific analyses.

	Main analysis	Sensitivity analysis 1	Sensitivity analysis 2	Sensitivity analysis 3
Exposure	Main: Any pneumonia diagnosis Control: Any UTI diagnosis Age-groups: 0-5, 6-10, 11-15, 16-20	Main: Primary pneumonia diagnosis Control: Any UTI diagnosis Age-groups: 0-5, 6-10, 11-15, 16-20	Main: Any pneumonia diagnosis Control: Any UTI diagnosis Age-groups: 0-5, 6-10, 11-15, 16-20, 21-30	Main: Any pneumonia diagnosis Control: Any UTI diagnosis Age-groups: 0-5, 6-10, 11-15, 16-20
Primary outcome	MS diagnosis 20+ Y Record in MSR or 1+ diagnosis NPR	MS diagnosis 20+ Y Record in MSR or 1+ diagnosis NPR	MS diagnosis 30+ Y Record in MSR or 1+ diagnosis NPR	MS onset 20+ Y Record in MSR
Secondary outcome	MS diagnosis 20+ Y Record in MSR or 2+ diagnoses NPR	MS diagnosis 20+ Y Record in MSR or 2+ diagnoses NPR	MS diagnosis 30+ Y Record in MSR or 2+ diagnoses NPR	MS diagnosis 20+ Y Record in MSR or 2+ diagnoses NPR
Additional variables	Highest attained education. Any diagnosis IM; age-groups same as exposure.	Highest attained education. Any diagnosis IM; age-groups same as exposure.	Highest attained education. Any diagnosis IM; age-groups same as exposure.	Highest attained education. Any diagnosis IM; age-groups same as exposure.

Matching factors of sex, age at MS diagnosis/index, and region of residence were used at strata variables. **Abbreviations:** IM: Infectious mononucleosis; MSR: MS Register; NPR: National Patient Register; UTI: Urinary tract infections; Y: Years of age.

3.3.2 Study II: Spasticity treatment patterns among people with MS

3.3.2.1 Overview of cohort study design aspects in a treatment context

A cohort study design was used in Study II in order to follow a group of MS patients throughout time to assess a variety of characteristics influencing whether or not they received a prescription to treat spasticity. This was an effective way to assess different

time effects such as age and length of MS disease on the risk of needing spasticity treatment. In particular, as spasticity has previously been primarily explored through cross-sectional studies or in randomized control trials of efficacy and safety, the specific timing of when MS patients required treatment had not yet been determined.

Of importance in a cohort study, is defining the period of time at which an individual is at risk of an outcome. A person should contribute person-time, or time at risk of the outcome, provided they are outcome free. In order to attribute different characteristics of MS to spasticity treatment, it was important to ensure that the person received a spasticity treatment after their MS diagnosis. Else, it is not possible to ascertain whether the spasticity treatment was prescribed as a result of MS. Therefore, time-at-risk for people with new diagnoses of MS started from their MS diagnosis and excluded individuals with a spasticity prescription prior to MS diagnosis.

Additionally, stopping time-at-risk was challenging in the context of prescription-use. Since a person dispenses a prescription for a specific purpose, this usually means that they already have the occurrence of the symptom prior to dispensing the prescription. In Study II, spasticity should already be present or being investigated if a treatment is dispensed. What we are measuring is then treatment patterns for spasticity rather than spasticity itself. Further adding complexity to assessment of the time-at-risk, is that the Prescription Drug Register started in July 2005. This means that some delays in prescriptions being recorded into the register can take place and that prescriptions prescribed prior to July 2005 could be registered or dispensed again from July 2005 onwards. Therefore, it was not possible to determine who had an existing prescription for spasticity treatment and receiving a refill of their prescription from a person who is receiving spasticity treatment for the first time. This necessitated use of a look-back period of time to exclude suspected prevalent users of spasticity treatments. A conservative period of one-year was defined as the look-back period.

Assessment of treatment discontinuation also presents challenges. In this case, time on treatment was considered the time-at-risk, and stopping treatment was the outcome. The difficulty here is not knowing how a person is using their prescription: taking regular doses, sporadically, taking it "just-in-case", or dispensing and not taking it at all. People can also wait different periods of time before refilling their prescription once empty, or they may not require an additional dispensation. Therefore, assessing the end of time-at-risk in this case was done by using different possible windows of time when people had the opportunity to refill a prescription. These windows were defined as 90 to 180 days after the date of the last dispensed prescription as different windows will influence discontinuation rates.

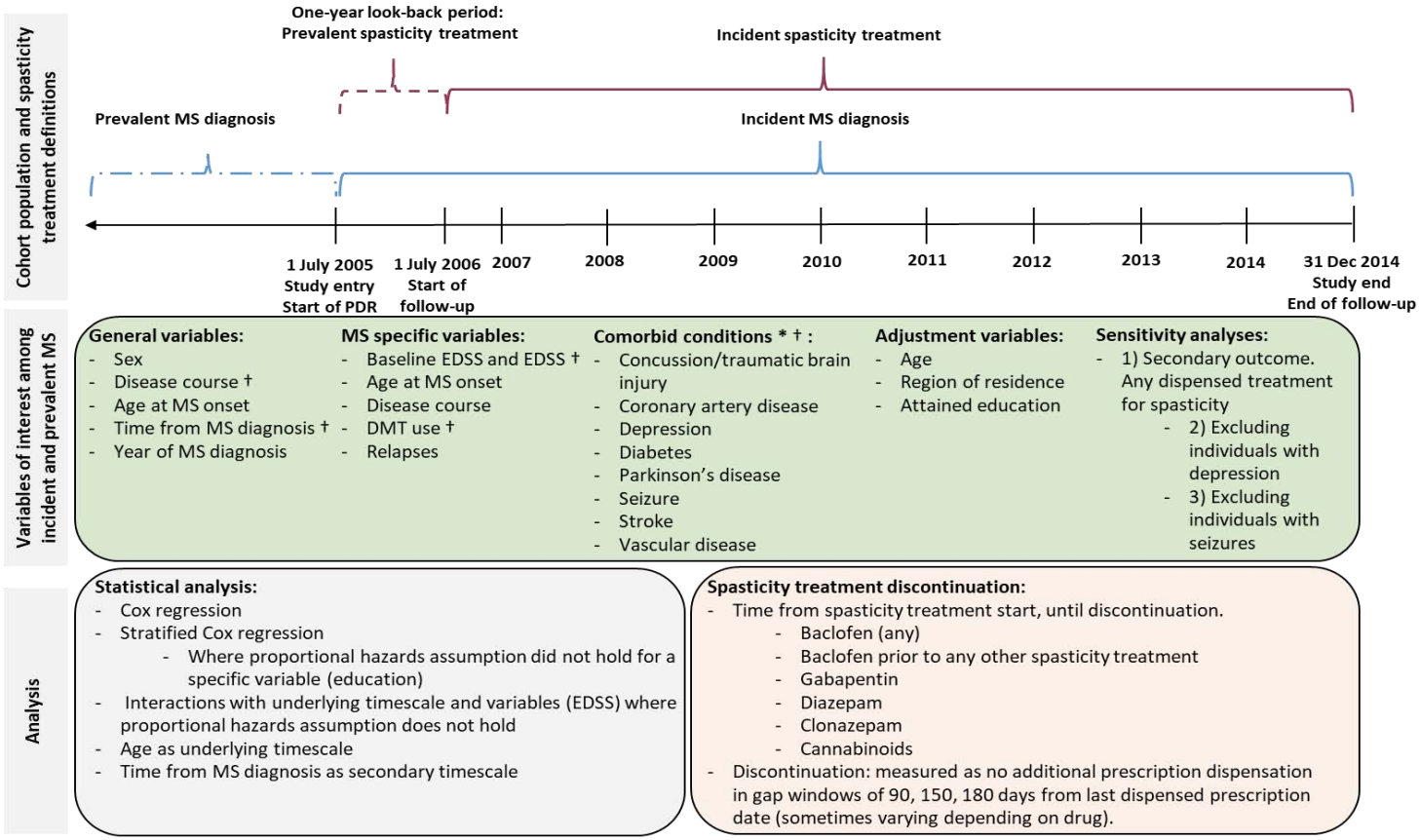
3.3.2.2 Study considerations: Study population, classification of spasticity treatment use

This study draws on a population of individuals recruited to EIMS, GEMS, and IMSE who provided blood samples and/or questionnaire data. Data from Swedish Registers was extracted and linked to these individuals. Diagnoses of MS were identified using the MS register, or if they had two or more ICD-10 codes for MS (G350, G359) in the National Patient Register at least 180 days apart. Individuals included as MS patients had diagnoses verified, and previously identified controls fulfilling the definition of MS were included as MS. Using at least two diagnoses reduced the possibility of recorded ICD codes for suspected disease rather than confirmed disease and this definition has been validated.[116,143,144] MS patients were further classified into two groups: incident if diagnosed on or after 1 July 2005; or prevalent if diagnosed prior to this date. This was necessary in order to further identify the specific timing of spasticity prescriptions among new patients with MS.

Baclofen is the most specific spasticity treatment used in Sweden and its dispensation recorded in the Prescribed Drug Register (ATC code M02BX01) formed the basis of the primary outcome in this study. New-use prescriptions of baclofen were of particular interest as the majority of studies examining spasticity used a cross-sectional approach, and were unable to determining the timing of spasticity treatment. Individuals with a spasticity prescription (Baclofen: M02BX01; diazepam: N05BA1; clonazepam: N03A301; gabapentin: N03AX12; or cannabinoids: N02BG10) during the look-back period were excluded. Additionally, persons with incident MS with a baclofen prescription prior to MS diagnosis were also excluded as it is not possible to determine if spasticity treatment was due to MS.

3.3.2.3 Variables of interest with respects to spasticity treatment

Specific MS variables were identified through the MS Register and included: disease course; DMT use classified as highly-effective therapies (alemtuzumab, ciclosporin, cladribine, daclizumab, dimethyl-fumarate, fingolimod, natalizumab, ofatumumab, rituximab and teriflunomide) or moderately-effective (interferons and glatiramer acetate). Relapse dates were extracted and erroneous dates, or dates three years prior to MS diagnosis were excluded. Relapses with 14 days of one another were considered the same episode. EDSS scores and dates of changes to the score between 0-2.5, 3.0-5.5 and 6+ were identified. Baseline EDSS was defined as the score closest in time to start of follow-up, provided it was not more than 90 days after start of follow-up. All variables of interest are described in Figure 6. Other possible comorbid disease conditions were also considered and their definitions using a combination of ICD codes and specific pharmacological treatments can be found in Appendix 0 of Study II.[145] These diseases were chosen as they have been previously shown to be associated with spasticity.



† Designates time-varying covariates. * Analyses of comorbid conditions was with any dispensed spasticity drug (secondary outcome).
Abbreviations: DMT: Disease-modifying therapy; EDSS: Expanded disability severity scale

Figure 6: Study II study design, variables of interest and analysis description.

3.3.2.4 Spasticity treatment discontinuation

Discontinuation of any treatment for spasticity was also of interest and was defined as the date when no additional spasticity prescriptions were dispensed after the previously filled prescription. Time-windows were used to determine this, and included a range of days: 90, 150, or 180 days. Although the prescription strength and the number of tablets dispensed are recorded in the Prescribed Drug Register, it was not possible to calculate a daily dose or the length of period the prescription was supposed to be taken. This is because the dosage prescribed does not always correspond to one pill, and it is not possible to know if an individual begins treatment, or adheres to treatment as prescribed. This same problem applies to trying to determine an approximate end-date of the prescription from dispensation date. Baclofen dosing also needs to be adjusted throughout time, also making a daily dose difficult to calculate. Therefore, individuals using baclofen are likely to refill their prescription, although in different time-intervals which required use of different lengths of gap-time. In some cases, such as for diazepam, prescriptions are filled more frequently as the drug is more tightly controlled, and thus re-fills of prescriptions must occur more frequently. Thus, some drug-specific gap-window variation was also used.

3.3.2.5 Statistical analysis: Cox regression, time-varying covariates and coefficients

Follow-up for incident people with MS started from either 1 July 2006 or MS diagnosis date, and 1 July 2006 for all prevalent people with MS. Follow-up ended was date of first baclofen initiation, 31 December 2014 (end of follow-up) or death, whichever occurred first. Analyses were performed for incident and prevalent patients with MS separately.

Cox regression models that produced hazard ratios (HR)[146] were used to assess the association of the variables described above with spasticity treatment. A hazard ratio quantifies the difference in the rate of the outcome (hazard rate) at any given time between two groups. In this case, time was modelled using age for the underlying timescale. This is because MS is often diagnosed at different ages and can influence the course of the disease. Spasticity may be both a result of age-related changes in the body in conjunction with MS-specific changes. To include the effect of having MS over time, time from MS diagnosis was included as a secondary timescale. As a variable can be both varying over time in terms of its effects on an outcome (time-varying coefficient), but also the value of a variable can change (time-varying covariate),[147] both aspects were considered when modelling. Therefore, when considering the proportional hazards assumption, (meaning that the effect of a variable can take on any magnitude or direction of effect, but that the association between a given exposure and the outcome is a constant function over time[148]) departures from proportional hazards were investigated in several ways. First visually by

plotting rates on a number of scales, secondly through plotting of Schoenfeld residuals,[149] and thirdly by introducing an interaction term with the underlying timescale.[148]

Variables that did not have proportional hazards over time and that were not of main interest, such as education, were modelled using a stratified Cox model.[150] Time-varying covariates were used when modelling the effects of comorbid diseases before and after their diagnosis on the rate of spasticity treatment. DMTs were also used as time-varying covariates as individuals can delay treatment start. EDSS's influence on spasticity treatment were investigated as a baseline characteristic, but also as a time-varying covariate as over time changes to EDSS scores are expected, and are not fixed over time.[151] As the effect of EDSS was not constant by age, an interaction between EDSS scores and age was introduced.

Spasticity treatment discontinuation was also modelled using a time-to-event approach. As the specific time to discontinuation was of interest, failure functions were estimated which model the probability of failure at a given time conditional on being in the risk-set until the specific time point. Failure functions were stratified by specific MS characteristics such as disease course and EDSS score at time of starting spasticity treatment, as these two factors have been previously discussed to have an influence for starting spasticity treatment. Cox regression models were also explored with respects to factors affecting discontinuation, such as sex, and MS characteristics, but no factors other than EDSS and disease course affected the rate of discontinuation and were not further considered.

3.3.3 Study III & IV: Pigmentation genes and MS risk and severity

3.3.3.1 Study population, genotyping and classification of MC1R-TUBB3 SNPs

Both Study III and IV included all individuals from EIMS, GEMS, IMSE and STOPMS who had been genotyped using the Illumina Human OmniExpress array, had imputed genotypes and passed quality controls as described above in genotyping and imputation. Three study populations were derived: the first included cases of MS and controls for the *MC1R-TUBB3* SNP association with MS risk analysis; the second included the cases of MS and controls who also had questionnaire information and data linked from Swedish registers for inclusion into the environmental/lifestyle analysis; and the third was individuals with MS who had first and last EDSS information from the MS Register for the *MC1R-TUBB3* SNP association with MS severity analysis. Study populations two and three were nested within the first population. An overview of the study design is provided in Figure 7.

The MS diagnosis of study population one was verified using individual study data, National Patient Register and MS Register data. For the second population, individuals were included only if they: had no missing year of birth or index year (date of MS onset/diagnosis or

matching); were born in Sweden; and who did not migrate between birth and index year. Migration and birth outside of Sweden were in order to ensure that individuals had similar possibility of specific exposures such as sunlight and other factors in Sweden. The third study population included only individuals with MS who additionally had a first or a last EDSS score and an accompanying date of assessment in the MS Register. Additional details for study populations one and two are described in Study III's manuscript and study population three is described in Study IV's manuscript.

3.3.3.2 MC1R-TUBB3 SNP associations to MS risk using MLMA-GREML

The *MC1R* and *TUBB3* are located on chromosome 16 in genome assembly GRCh37 (hg19) at base pairs 89,978,527-89,987,385 and 89,985,733-90,002,500 respectively. All variants from these positions were extracted. Additionally, in order to consider possible regulatory element effects, variants one kilobase upstream of *MC1R* and downstream of *TUBB3* were also included. Within these positions, five known RHC variants were available: rs1895005, rs2228479, rs1805007, rs1805008, and rs885479. We analysed RHC variants, as well as expanded the analysis to all available variants within the *MC1R-TUBB3* region. This was to be able to infer from possible significantly associated SNPs as to whether the *MC1R* or the *TUBB3* regions were associated to MS. Analyses were first performed including all individuals, then stratified by sex. Given our cases and controls are sex-matched, a fully stratified analysis is necessary. P-values were corrected using both a false discovery rate correction and a Bonferroni correction.

Analyses of the association between *MC1R-TUBB3* and MS risk were performed using mixed-linear model association-genomic relatedness based restricted maximum-likelihood models (MLMA-GREML).[131,132,152] This model uses a GRM in order to control for population stratification, as was described above. This analysis could also have been performed using logistic regression and correcting for PCs in a single SNP analysis approach, but the MLMA approach was chosen instead for several reasons: the models are better able to detect associations if they exist due to better power, especially important when considering the low-frequency variants with *MC1R-TUBB3*; and they avoid false positives when there is an underlying population structure by applying a correction specific to the population. In contrast to a single SNP association model, the MLMA fits all SNPs of interest as a random effect, which accounts for linkage disequilibrium between the SNPs. Despite a MLMA model being a linear-type model with typically a continuous outcome, it can also be applied to case-control data through use of a linear transformation.[132] We applied this transformation in order to report the effect estimate, ORs, for associated SNPs.

When using the MLMA model in the case-control setting, it is imperative that the genetic data are thoroughly quality checked, as is performed in our data, in order to avoid genotyping errors that may lead to cases being more similar to one another. This is less

likely when an outcome is continuous, as it is less likely to have associations in a specific direction associated with a SNP. It is more likely to occur with a binary outcome.

This type of SNP specific analysis is of relevance, even in the era of GWAS analyses. GWAS are extremely powerful tools, provided there is sufficient sample size in order to detect small effects. As GWAS sample sizes and international consortiums work together to analyse SNPs, more SNPs are being shown to be associated with MS. Where a GWAS falls short is in the case of SNPs with a small effect or of a low frequency as these effects will likely not be detected in a GWAS analysis due to the large number of SNPs analysed, and the heavy correction applied to the p-value in order to avoid false positives. This is where a MLMA model for analysing specific SNPs of interest, or in this case a specific locus, is useful. Additionally, determining sex-specific effects through stratification further reduces the power to detect associations in a GWAS. By using a hypothesis-based, SNP-based approach and a MLMA model we are more likely to be able to detect an effect over traditional GWAS analyses.

3.3.3.3 MC1R-TUBB3 SNP associations to MS severity using MLMA-GREML

3.3.3.3.1 Overview of cross-sectional study design

A cross-sectional study design was employed in Study IV. All available people with MS from the epidemiological study data (EIMS, GEMS, IMSE, STOPMS) were included provided they had a measurement of their disease severity as measured by EDSS. People with MS were further stratified based on their *MC1R* SNP status, and the outcome was their disease severity. Although a cross-sectional approach is not the most ideal study design in the context of repeated measurements in MS severity, it allows for examination of initial associations and influences of variables on a possible association to be identified. Although time effects cannot be explored, some measures such as weighting the outcome based on age, can provide an idea of possible timing aspects with respects to exposure and outcome, as was done in this study.

In order to test the association of RHC *MC1R-TUBB3* variants with MS severity, as similarly described above for MS risk, first or last EDSS scores were converted to the ARMSS score. ARMSS score was log-rank inverse transformed in order to produce a normally distributed score. This transformed score was used as the outcome for either the first or last ARMSS. Note that not all individuals had both a first or last score, therefore if only one score was recorded, it could be repeated for both analyses. As ARMSS has been shown to be robust in classifying individuals' MS severity when measured at any time, this is acceptable.

A MLMA-GREML approach was also used, with first and last log-rank inverse transformed ARMSS used as a continuous variable in separate analyses. Fully sex-stratified analyses

were also performed, as patterns of severity and progression have been shown to be different among males when compared with females.

3.3.3.4 Top associated MC1R-TUBB3 SNPs and individual environmental and lifestyle MS risk factors

SNPs identified as associated with MS risk in the *MC1R-TUBB3* SNP analyses were included in this subsequent analysis, which investigated whether associations in *MC1R-TUBB3* were independent. Logistic regression models, adjusting for the first five PCs and matching variables (age, region of residence and sex) also included other relevant, known risk factors for MS. Analyses were stratified by sex to determine sex-specific effects, and multiplicative interactions between the known MS risk variable and *MC1R-TUBB3* SNP were investigated.

Specific MS risk variables that were included were in three categories: 1) Vitamin-D and sunlight; 2) underlying genetic risk of MS, and 3) established lifestyle risk factors for MS.

1) Vitamin-D levels could not be assessed directly, however a GRS for the likelihood of having low vitamin-d blood serum levels was used as a proxy using five of six risk variants that were in the GWAS by the SUNLIGHT consortium for serum vitamin-D levels.[135] This SUNLIGHT consortium's variants were chosen as the six variants explained still a good portion of variance in serum vitamin-D levels as was found in a later, larger GWAS that included many small-effect variants (a total of 143 variants)[153]. The specific SNPs used are found in Supplementary Table 1 of Study III. The GRS was calculated as described above. Individuals with missing SNP information were assigned the major alleles among controls. Sunlight was measured by region of residence as a proxy, as categorized into three main regions of Sweden: Norrland (northern, includes Umeå), Svealand (central, includes Uppsala and Stockholm) and Götaland (southern, includes Gothenburg and Malmö).
2) Self-reported data from the EIMS, GEMS and IMSE questionnaires was used to assess whether individuals had first-degree relatives with MS. Thus, the underlying genetic risk of MS was measured by proxy using this measure.
3) Two main lifestyle risk factors for MS are smoking and BMI. Both of these were also assessed from the questionnaire data. Smoking was assessed prior index date (date of MS diagnosis or matching), through self-reported years or ages of smoking and was categorized as ever vs never. Height and weight were reported at age 20 years and BMI was calculated and categorized as <18.5, 18.5-24.99, 25-29.99 and 30+.

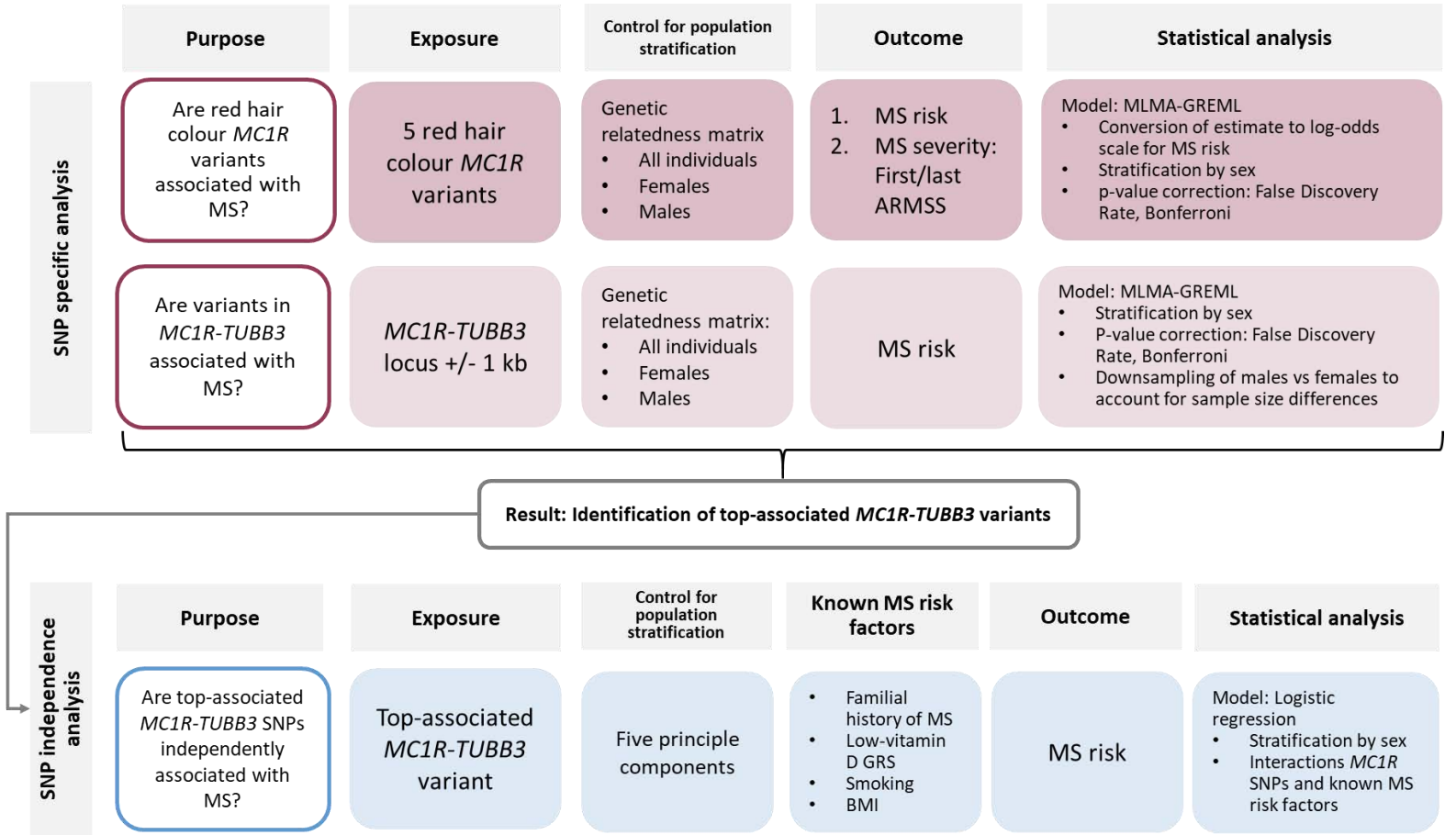


Figure 7: Study III and IV study design, variables used in exposure and outcome, and statistical analyses.

3.4 Ethical considerations

Research at its core should be composed of good biomedical ethics. Biomedical ethics traditionally is composed of four grounding principles: beneficence, non-maleficence, respect for autonomy and justice. Often thought of as separate topics, these intertwine throughout research. Two additional topics that fit within these principles, important to consider, are the researcher's duty as well as personal biases.

In this doctoral thesis, large, comprehensive datasets including personal medical history, specific disease history, demographic data, self-reported data about one's life as well as entire genotypes have been linked together. Much of these data do not require informed consent for collection, but for inclusion in specific disease registers such as the MS Register, or inclusion in EIMS, GEMS, IMSE or STOPMS, informed consent was required. I feel the duty of the researcher is both to the scientific community, but also to the individuals consenting to use their data and directly contributing data. This means responsible handling of the data as according to Swedish and European data-protection laws, but also respecting the decision of Sweden as a nation to collect and record this data, that the collected data would do no harm and that it will have a purpose. This means not only ensuring studies are feasible and do no harm, but also considering the relevance and the purpose of the study before extracting or collecting data. I also believe that it is the researcher's duty to remember that there are very real individuals who are behind the numbers in the dataset, and in this way, this allows respect for individual autonomy among individuals who consented to data collection. This is especially important in the context of register-based research not requiring informed consent. Recognizing persons included within the study as a person also helps to identify how the research can beneficial.

Personal-bias in research is not often discussed within a scientific manuscript. A number of other types of bias such as biases arising from data collection, or bias arising from disease classifications can reliably be found in a scientific article's strengths and limitations section, in particular in epidemiological studies. The question here I would like to consider is whether our own personal biases influence our research. We as scientists are trained to think objectively, without preconceptions and to be guided by established scientific evidence. Of course, this is not always possible. One decides how their will investigate a disease, as I have done in my thesis. My point of view is that MS is a neurological, immunological disease. If I am incorrect in this perspective, it is possible I could do more harm than good if results lead to a false conclusion. Societal bias and understanding may also influence personal bias or understanding. This can be seen throughout history as knowledge changes and is modified over time. For example, it was not long-ago that germs as we known them were thought not to exist, the benefit of hand-washing was a myth and disease was simply an imbalance in the body prior to Ignaz Semmelweis' disruption of this way of thinking only a short 160 years ago.[154]

Empirical science seems to have an answer to the personal bias or societal bias problem, if it does influence research. The answer is that empirical science has a self-correcting feature even in absence of personal bias. Evidence-based research methods allows for scientifically established facts to be changed and modified in the presence of new evidence. With respects to MS, even within the relatively short-time I have been directly involved in MS research, facts have been established and refuted. Some changes include the acknowledgement of a prodromal period not previously thought to be present in MS, the acceptance of smoking as an established risk factor, and evidence of both white and grey matter of the brain being affected.

Although I have done my best to not allow personal bias to influence my research, it is reassuring and daunting that this self-correcting feature of science may either correct, refute, or confirm the results of my studies. In this context, it is less important to consider if one is "right", but rather how far one might be from the "truth". In ensuring the usefulness of my research, in this context to people with MS and also other researchers in MS, I am then working to also form a small collection of evidence, which help to fit into the current knowledge of MS as a disease. This means it is my additional duty to contribute to evidence, and recognize the limitations of my own research simultaneously, in order to make room for new theories, information, methods and evidence to steer my research.

4 Summary of key results

This section will summarize the results from each of the studies included in this thesis.

Study I: Pneumonia as a risk factor for MS

Diagnoses of pneumonia, IM, and UTIs were rare, as fewer than 1% of either cases and controls had this diagnosis. Given few individuals had repeated diagnoses of pneumonia, it was dichotomized as ever versus never. Pneumonia as a primary or secondary diagnosis among individuals aged 11-15 years showed the highest magnitude association with MS risk overall even after inclusion of UTIs, IM, and education in models (Table 5). Using MS definition 1, pneumonia between ages 11-15 years showed a statistically significant association with MS with an OR of 2.00 (confidence interval [CI] 1.22-3.27). Considering individuals with at least two MS diagnoses from the National Patient Register or a record in the MS Register decreased the number of MS cases by 8% (n=517), but the association of pneumonia ages 11-15 years with MS remained consistent and only decreased marginally in magnitude (OR of 1.90, CI 1.12-3.23). UTIs showed no statistically significant association with MS risk, although they were more common in ages 16-20 years. Sex-specific associations were of interest but exposure numbers were too few to be able to stratify analyses. In general, UTIs were more frequent among females than males, however no specific pattern was observed with pneumonia diagnoses. Similar results, although of reduced magnitude, were observed for pneumonia in sensitivity analyses among people diagnosed with MS over age 30 years.

Table 5: Association of pneumonia with MS in childhood and adolescence with urinary tract infections as a control infection for reverse causation.

	MS definition 1: MS diagnosis over age 20 years from the MS Register or National Patient Register.			**MS definition 2:** 2+ MS diagnoses over age 20 years from National Patient Register or record in MS Register.		
	Cases n (%)	**Controls n (%)**	**OR (CI)**	**Cases n (%)**	**Controls n (%)**	**OR (CI)**
N	6109	49479		5592	43327	
Pneumonia						
Age (years)						
0-5	51 (0.83)	498 (1.01)	0.81 (0.61-1.09)	46 (0.82)	428 (0.99)	0.83 (0.60-1.13)
6-10	16 (0.26)	161 (0.33)	0.79 (0.47-1.33)	15 (0.27)	147 (0.43)	0.77 (0.44-1.43)
11-15	30 (0.33)	81 (0.16)	2.00 (1.22-3.27)	17 (0.30)	74 (0.17)	1.90 (1.12-3.23)
16-20	19 (0.31)	112 (0.23)	1.31 (0.80-2.13)	17 (0.30)	94 (0.22)	1.27 (0.73-2.19)
Urinary tract infections						
Age years)						
0-5	39 (0.64)	317 (0.64)	0.96 (0.69-1.35)	33 (0.64)	275 (0.63)	0.89 (0.61-1.30)
6-10	27 (0.44)	195 (0.39)	1.10 (0.74-1.66)	24 (0.43)	169 (0.39)	1.16 (0.75-1.79)
11-15	10 (0.16)	104 (0.21)	0.68 (0.35-1.31)	7 (0.13)	94 (0.22)	0.53 (0.24-1.14)
16-20	60 (0.98)	362 (0.73)	1.30 (0.98-1.71)	54 (0.97)	317 (0.73)	1.30 (0.96-1.75)

Abbreviations: CI: Confidence interval; Odds ratio: OR. Note: Odds ratios show the association of pneumonia to MS, conditioning on matching variables, adjusting for infectious mononucleosis, and highest attained education.

Study II: Spasticity treatment patterns

Of the 5345 individuals with MS included, 34% were classified as incident and 66% as prevalent MS. In general, people with prevalent MS were older, had a greater proportion with progressive course, and had higher disability as measured by EDSS than individuals with incident MS. Of these individuals, 10% incident and 24% prevalent MS cases received and collected a prescription for baclofen. When widening the definition of possible spasticity treatments, 33% and 50% of individuals with incident and prevalent MS collected a treatment. Among those with incident MS, dispensed treatments proportions were: 31% baclofen, 1% clonazepam, 17% diazepam, 50% gabapentin and < 1% cannabinoids. Dispensed treatment proportions individuals with prevalent MS included: 44% baclofen, 3% clonazepam, 20% diazepam, and < 0.5% cannabinoids. Overall, with increasing age, baclofen and other possible spasticity treatments increased in use. Prescribing reasons for each treatment are unknown, though baclofen is the most specifically used prescription for spasticity treatment among people with MS in Sweden.

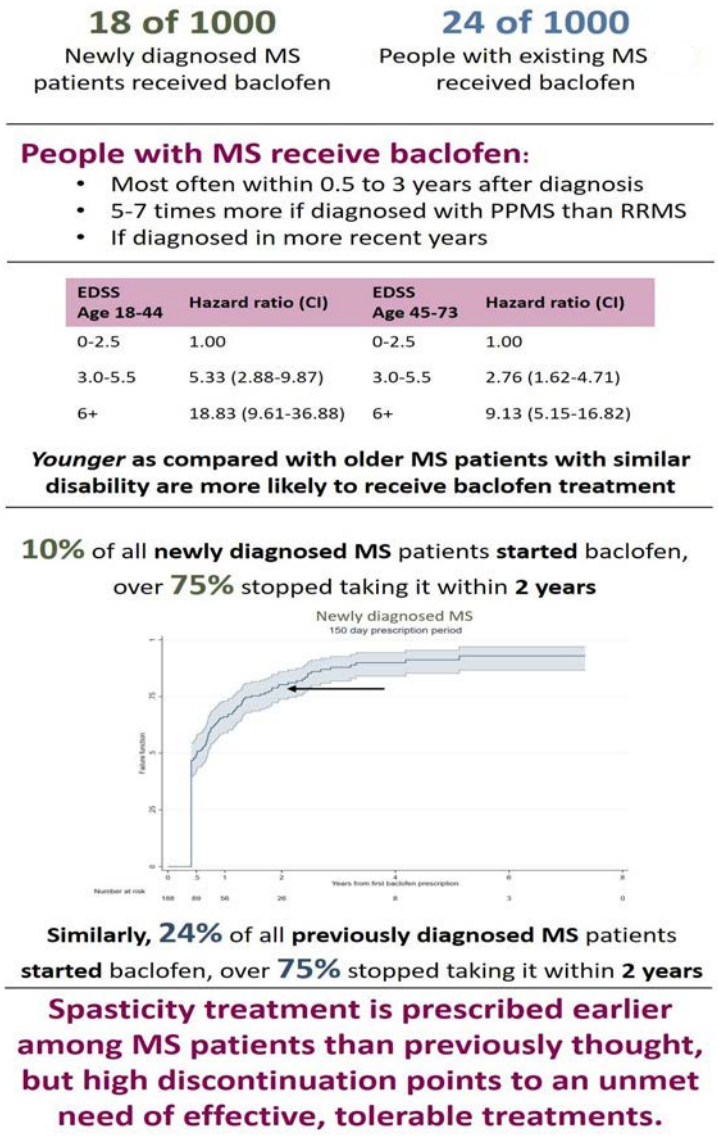
Strong sex-specific effects among both incident and prevalent MS were observed when considering models with disease course, time with MS, MS onset and year of MS diagnosis. Females were 25% less likely to receive baclofen than males. These sex-specific effects were attenuated when considering models where disability measured by EDSS were included. No interactions with sex were observed, pointing to disability being the more relevant indicator of baclofen treatment.

When examining disease course and EDSS separately, individuals with progressive disease courses (PPMS or SPMS) were 4-7 times more likely to receive baclofen than RRMS among either the incident or prevalent MS individuals. Though over time individuals are more likely to convert to a progressive course, length of time with MS showed no association with baclofen and though not statistically significant, rates of baclofen dispensation decreased with increasing time with MS. Age at MS diagnosis increased the rate at which baclofen was received with a one-year increase in age at MS diagnosis resulting in a 1% increase in baclofen prescription (HR 1.01, CI 1.00-1.02), seen among people with prevalent MS.

Associations of baclofen treatment with disability was measured using both EDSS at baseline and where it was allowed to vary over time. Baseline EDSS showed that increases in EDSS were also associated with increased baclofen treatment, though HRs for EDSS of 3.0-5.5 to 6+ were similar in magnitude. When allowing EDSS to vary over time, age- and disability-specific trends emerged both among those with incident and prevalent MS. Individuals at younger ages and with greater disability had 2-3 times larger HRs as compared with older individuals of the same disability (Figure 8, table), and were of larger magnitude among those with incident MS. These age- and disability-specific rates were observed even after controlling for sex, age at MS onset, disease course and DMT use.

Among individuals using baclofen, discontinuation rates were very high, with 75% of all individuals discontinuing treatment within two years of initiation (Figure 8). When stratifying by disease course or EDSS, specific trends emerged. Individuals with progressive disease course or higher EDSS remained on treatment approximately 0.5 to 1-year longer than individuals with RRMS or low EDSS (0-2.5), but they too discontinued within two years. People with RRMS and lower EDSS most often discontinued within 6-months of initiation. These trends were observed both among incident and prevalent MS.

Figure 8: Study II results at a glance.



Study III: Pigmentations genes and MS risk

A total of 12 466 individuals were included in the *MC1R-TUBB3* gene analysis, of which 6585 were cases and 5881 were controls. Examining only the RHC SNPs, rs885479 showed a statistically significant association with MS that remained significant after correction for multiple testing (n=5) (Figure 9: OR 0.87, standard error [SE] 0.013 p-value < 0.05). Stratifying by sex (Females N=9203: 4777 cases, 4426 controls. Males N=3263: 1808 cases, 1455 controls) showed sex-specific associations across two markers (rs2228479 and rs885479), of which rs885479 remained statistically significant after correction for multiple testing (OR 0.83, SE 0.015, p-value < 0.01). No association was observed among males for any of the RHC SNPs.

Examining the association of the entire *MC1R-TUBB3* locus including 1Kb flanking regions with MS showed that rs885479 remained the most strongly associated variant. No other variants associated with MS. After correction for multiple testing (n=144), the association of rs885479 did not remain statistically significant. Stratifying by sex showed possible sex-specific differences, as smaller p-values were observed among females as compared with males. The same analysis with down-sampled females to ensure equivalent sample sizes between sexes showed that although still the largest magnitude association with MS, this SNP did not remain statistically significantly associated among females with MS.

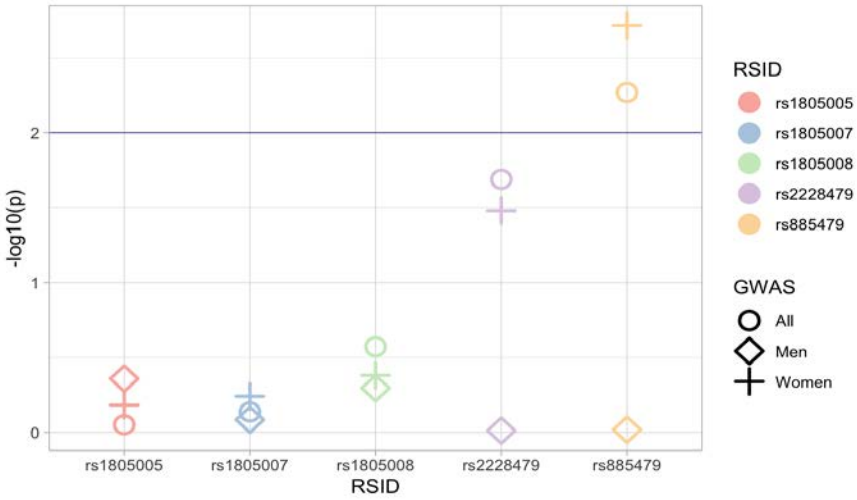


Figure 9: Association of *MC1R* red hair-colour SNPs with MS among all individuals and stratified by sex using mixed-linear model association methods. Models include correction for population stratification effects using a genetic relatedness matrix.

The gene-environment analysis used the top-associated SNP rs885479 in logistic regression models to test its independence from other well-known MS risk factors. A total of 10 582 individuals were included (5674 cases, 4907 controls) as they had questionnaire and register-linked data. The analysis was performed for all individuals, and stratified by sex. The main results of the female-specific analysis are in Table 6 and showed that none

of the MS risk factors influenced the association of rs885479 with MS among females. The association remained consistent even when step-wise inclusion was performed for including the vitamin-D GRS, first-degree relatives with MS, smoking prior to index (MS onset or matching date) and BMI at age 20 years. Furthermore, none of these variables showed any effect modification with rs884379 (multiplicative interaction). Inclusion of all of the MS risk factors in the model alongside rs885479 further increased the magnitude of the SNP's association with MS (OR 0.81, CI 0.71-0.93). Increases in the number of alleles of rs885479 also showed that homozygous carriers had a lower OR than heterozygous carriers and non-carriers (homozygotes OR 0.50, CI 0.27-0.95; heterozygotes OR 0.85, CI 0.74-0.97).

Table 6: Determining top-associated *MC1R* SNP rs885479-A independence from other known MS risk factors among females. Odds ratios were calculated using a logistic regression model and included principle components 1-5, cohort (EIMS, GEMS, IMSE), year of birth, and region of residence at index (index defined as year of MS onset or matching year). Risk factors included together in the model with rs885479.

FEMALES	**N=7134**	
	n Cases/Controls	**Odds ratio [95% CI]**
rs885479-A		
Non-carrier	3182/2976	1.00
Carrier	444/532	0.81 [0.71-0.93]
Vitamin-D genetic risk score	3626/3508	0.56 [0.26-1.20]
First-degree relatives MS		
No	3376/3461	1.00
Yes	250/47	5.33 [3.88-7.33]
Smoking prior index		
No	1560/1816	1.00
Yes	2066/1692	1.41 [1.28-1.56]
Body mass index (BMI) age 20 years		
BMI<18.5	487/437	1.00
18.5<=BMI<25	2744/2766	0.94 [0.81-1.08]
25<=BMI<30	295/250	1.14 [0.91-1.42]
BMI 30+	100/55	1.80 [1.25-2.60]

Study IV: Pigmentation genes and MS severity

Individuals with MS with an EDSS score were included (N=6113, females=4448, males=1665) with analysis of the first and last score. Individuals were mostly of relapsing-remitting course, and with a low EDSS score (0-2.5). Median ARMSS score at first measurement was 4. The first and last rank-inverse log-transformed ARMSS score tested the association of RHC variant to MS severity, and was performed among all individuals and stratified by sex. No association of RHC SNPs with ARMSS scores was observed once multiple-testing was accounted for using false discovery rate or Bonferroni corrections (Figure 10). There was some equivocal evidence to suggest a sex-specific effect for rs1805007 with last ARMSS among females (beta=0.12, SE 0.05, p=0.01), but this association did not pass correction for multiple testing using either false discovery rate or Bonferroni correction (tests corrected n=5, p=0.06).

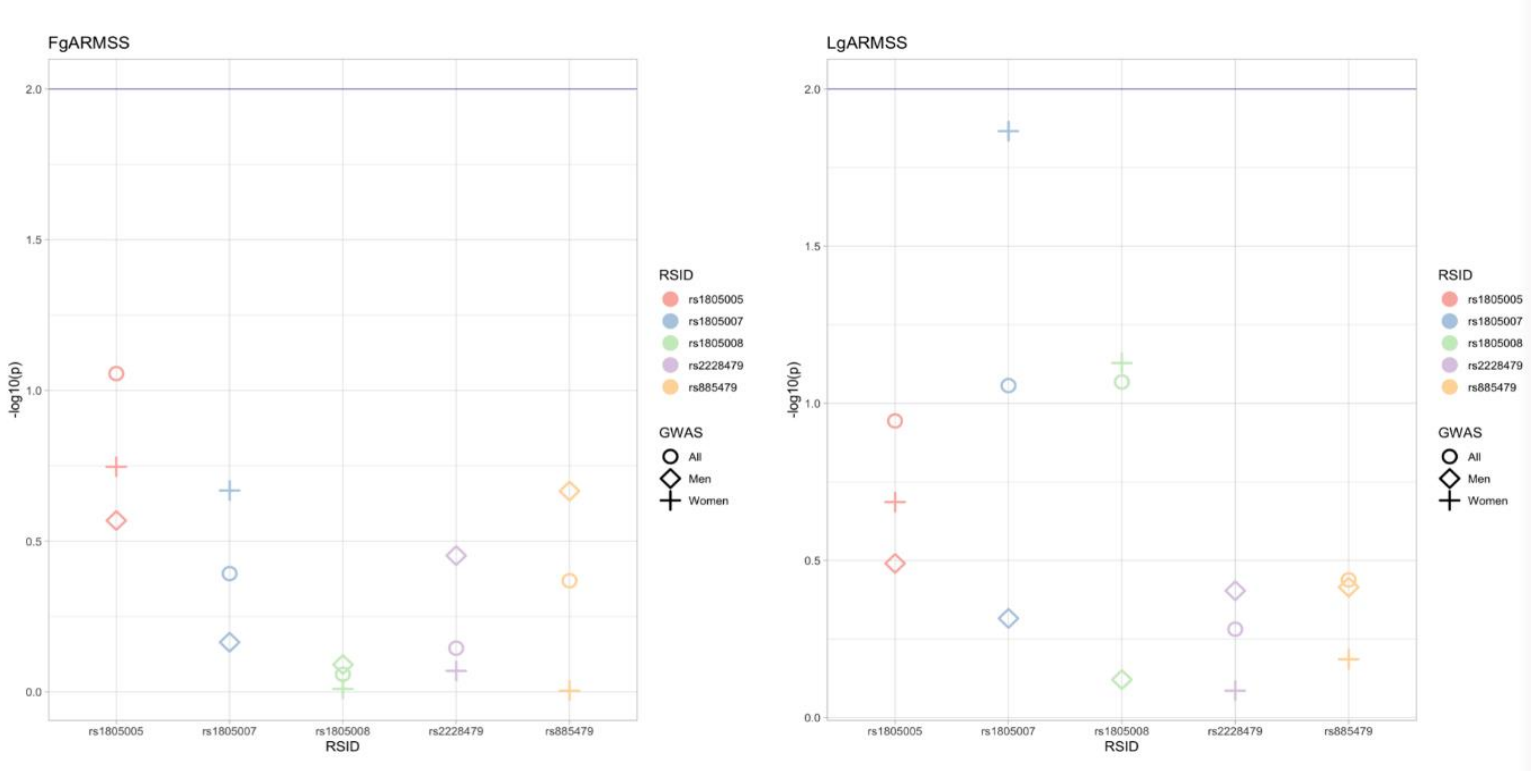


Figure 10: Melanocortin-1 receptor gene red-hair colour SNP associations with first and last rank-inverse log-transformed global Age-Related Multiple Sclerosis Severity Scale (FgARMSS, LgARMSS respectively). Analysis was performed among all individuals with MS and stratified by sex. SNP rs1085007 had the smallest p-value among females, but no statistically significant association remained after correction for multiple testing (n=5, p=0.06).

5 Discussion

This thesis explored aspects of MS aetiology and consequences through use of a combination of data sources and epidemiological methods. This thesis adds to a growing body of evidence of a critical risk period in MS risk; strengthens the importance of assessing time-related trends in the treatment of MS symptoms; shows that the same risk factors in the causes of MS may not be involved in the severity of MS; and investigated sex-specific effects. This section of the thesis will discuss aspects pertaining to each study, while also considering points of perspective and possibilities for future studies.

Study I provided evidence in support of a critical risk period, and highlighted the relevance of examination of risk factors outside of a potential prodromal period. Hospital-diagnosed pneumonia was associated with increased MS risk in early adolescence at age 11-15 years, but not in earlier childhood or late teenage years. To further evaluate the possibility of prodromal effects, we included UTIs as a control infection. No association with UTIs were observed, indicating that reverse causation is not a likely explanation for the observed association with pneumonia in adolescence.

Study II demonstrated age-specific and disability-specific associations with baclofen treatment use that to our knowledge is the first to do so. The use of a cohort study design, and inclusion of newly diagnosed people with MS made it possible to observe time-specific associations that to our knowledge, were not previously reported. Younger individuals with increased disability as compared to older individuals of similar disability received baclofen treatment at higher rates. Sex differences previously shown in other studies were attenuated once specific MS characteristics such as EDSS were included in models. This study demonstrates that assessing the timing of symptomatic treatments in MS are needed to better understand the possible treatment needs and appearance of symptoms among MS patients. This study also showed that additional tolerable and effective treatments may be required given high discontinuation rates of treatments.

Study III and IV in particular used a variety of methods and data sources in order to evaluate associations of the same risk factors prior to and after MS diagnosis. This was also in part the main source for the hypothesis of Study I as pneumonia occurs more frequently among MS patients. Study III showed a possible female-specific, genetic risk factor for MS. The results of Study III showed that the association with *MC1R-TUBB3* was not modified by known risk factors for MS such as genetic susceptibility to low vitamin-D, BMI or smoking prior to MS diagnosis. The same locus was considered with respects to MS severity, though no associations were found. Examining the same risk factors prior to and after the MS diagnosis can help to better determine the causes and consequences of MS to inform future study directions.

5.1.1 Critical risk period in MS development

Methodological limitations in studying risk factors for MS

Evidence of a critical risk period in the risk of MS is growing. Hypothesis surrounding adolescence as having a key role in MS pathogenesis have been alluded to, though few studies have investigated exposures in adolescence. The previous paucity of studies could be due to feasibility and practical reasons. A sufficient number of individuals who have had an exposure that is relatively uncommon, and also develop MS are difficult to recruit in a longitudinal study, taking years to achieve at great financial cost. An alternative method would be to include people with MS and ask them to recall whether they have had various exposures in adolescence, however it may be difficult or impossible to remember events in childhood and adolescence. Incorrectly remembering events can lead to misclassification of exposures, especially if both people with and without MS do not recall exposures in the same manner, causing recall bias. Nevertheless, even by collecting information from people with and without MS about possibly uncommon exposures may still not result in a sufficient number of individuals to be able to stratify by age-groups, or examine associations in additional groups. Another alternative is the use of register-based data, such as the data used in all Studies of this thesis, though register-based data are also not free from possible error. Data in administrative-based registers such as the National Patient Register, and specialized disease registers, such as the MS Register, capture a wide-range of disease diagnoses and specific clinical characteristics. Yet, there is still missing data in the MS Register, in particular for repeated measurements such as EDSS and relapses though large efforts have been made to reduce the missingness.[115] Furthermore, given the administrative nature of some of the registers, validity and accuracy of the diagnoses can be difficult to determine, though extensive work has been done to validate various diagnoses.[116,118] When considering time-sensitive exposures, register-based studies minimize loss to follow-up, are not subject to recall bias, and nearly the entire population of MS patients in Sweden are eligible for selection into a study.

Further evidence of a critical risk period

Though all studies have limitations, use of various study methods and data sources help to triangulate evidence of a critical period in adolescence for MS risk. Evidence first emerged of a possible critical risk period in MS development in studies from the 1960s-1970s. These studies were of migration from countries where the occurrence of MS was infrequent to countries where MS was more frequent.[155,156] Kurtzke et al. found that migration prior to the age of 15 years was associated with an increased risk when migrating from countries of low MS prevalence to high prevalence and vice versa.[155] These studies primarily hypothesized that reasons for the associations were due to infectious causes, in particular EBV exposure. However, the change in risk given migration from one country

to another also pointed to possible environmental and lifestyle influences due to country-specific differences. A more recent study in Canada by Rotstein et al. confirmed age-specific associations of migration with MS risk.[74] Additional risk factors in adolescence have since emerged, many of which have been studies in Sweden. Concussion is among one of the exposures investigated, as one concussion between age 11-20 was associated with MS, increasing MS risk by 22% (OR 1.22, CI 1.05-1.42), with repeated concussions further increasing the risk by 133% (OR 2.33, CI 1.35-4.04).[59,61] Increases to BMI in adolescence among conscripted males in Sweden also showed an association with increased MS risk.[70]

Additional associations with infections in adolescence in addition to Study I have also been observed. In another study that was an expansion of Study I, in which I was a co-author, a birth cohort was used to assess whether other types of infections were associated with MS.[60] Using a similar aspects of study design as Study I, infections including pneumonia, IM, CNS infections and general respiratory infections among others, were investigated in broader age-groups (0-10, 11-20 years of age). This study found that any infection in adolescence was associated with MS risk (HR 1.33 CI 1.21-1.46), but particularly CNS infections in adolescence were associated (HR 1.85 CI 1.11-3.07). Though Study I used a positive control of UTIs, and the CNS infection study used a birth cohort, it was debated whether the associations with infections were due to underlying genetic susceptibility to infections. Study I and Xu et al.'s birth cohort study also showed associations with IM and MS, in which adolescence had the highest magnitude associations, yet this association also could have been due to increased susceptibility to infection. Therefore, a sibling-based study comparing siblings concordant for IM, in which I was also involved, investigated the association of IM and MS.[157] A sibling-based study by design controls in part for environmental factors but also genetic factors. Both childhood and adolescence were associated with increased MS risk, with the highest magnitude associations in adolescence (childhood IM, HR 1.98 CI 1.21-3.23; adolescent IM HR 3.00 CI 2.48-3.64).

Further evidence of EBV's involvement in MS as a likely element in causal pathways of MS development was recently published in 2022.[67] Using data from individuals in active military service in the United States, the authors compared people with positive EBV to EBV negative serum individuals and showed a 26 times increased risk of MS (HR 26.5 CI 3.7-191.6). They also complemented this result by showing that serum neurofilament light levels were not elevated at the time of infection, indicating that biological mechanisms involved in MS pathogenesis were not detectible.

Relevance and points of perspective

Understanding the possible influences of MS development are relevant to identify the mechanisms of MS to ultimately prevent it from occurring. MS is not a one-cause disease:

it requires a cumulative effort of multiple exposures in driving pathogenesis. The resilience of the body and CNS to withstand, repair and remain healthy when encountering possible compromising environmental and lifestyle factors coupled with an underlying genetic susceptibility to develop MS is overwhelmingly positive. Though, understanding these complex intertwinements and timing of factors in which MS then begins to develop in the research context is rendered more challenging, especially when overall contribution of each risk factors has a small effect. There may also be a point in which possible initial pathogenesis is reversed due to CNS resilience, but with a repeated or additional occurrence of risk factors, as observed with concussion, this may tip the scales faster and more certainly towards MS development. Future research should seek to understand what combination of factors are interacting detrimentally. Initially planned in this doctoral work was to combine genetic factors with IM, pneumonia and various additional forms of trauma to the CNS that was not limited to concussion. However, all three studies were not possible due to nearly no diagnoses of any of these factors among individuals who also had genotypes in data we linked to Swedish registers. Among nearly 16,000 individuals (both MS and non-MS), counts of occurrence were well under 30, making additional stratification by genes of interest, or even a MS genetic risk score unfeasible. Likely cross-collaboration through large consortia or among several countries will be a relevant path to explore these possible associations.

5.1.2 Sex-specific differences in MS

All studies considered sex-specific effects in their initial study design. Examining sex-specific effects in Study I was unfortunately not possible given the low number of pneumonia diagnoses in general. Study II investigated whether sex-specific effects remained after inclusion of other variables in models that included more than only disease course. Study III and IV's main hypothesis was that there could exist sex-specific differences with respects to *MC1R*, sunlight and other lifestyle factors for both MS risk and severity. This section discusses possible sex-specific effects in further detail below.

Sex differences in infections involved in MS risk

Though Study I aimed to investigate sex-specific effects, stratification by sex resulted in several age-groups having very few occurrences of pneumonia, UTIs, and IM. This was a greater problem among males because of the sex-disparity in MS: fewer males are included as fewer have MS. Therefore, the lower number of infection occurrences among males could have been due to the lower proportion of males overall. If females had been down-sampled in this study, in a similar approach as to what was performed in Study III (see methods in the attached manuscript), it is possible that the already few occurrences of pneumonia and UTIs among females would have been further reduced. Nevertheless, sex-specific effects should be further investigated with respects to various infections. The finding of EBV as a plausible causal factor in the pathogenesis of MS[67] (as discussed

above) is an exciting finding. Within the study population included, a third were female which is interesting as often active military personnel data include only men. Effects were estimated among all individuals, as numbers in general were small. Therefore, it remains unclear as to whether there exists a difference among sexes with respects to EBV and the association with MS. Possible sex-differences could indeed exist. A study of EBV in the United Kingdom found that EBV seropositivity and manifestation of IM occurred more frequently among females, especially during adolescence defined as age 10-15 years.[158] In age-periods after 15, incidence of IM among males was greater than females overall.[158] The occurrence of an IM infection that may have sex-specific or age-specific variation could have an impact for MS, given the sex-disparity and evidence to support an age-specific critical period for MS risk. Sex-differences in immune responses have been widely discussed, given females as compared with males have been described as having a greater innate and adaptive immune response to pathogens.[159] Some evidence also seems to support age-specific effects acting in from birth to adulthood.[159] One study examining inflammatory responses among adults showed marked differences in inflammatory responses among males and females.[160]

Sex differences in symptoms of MS

Not only can sex differences occur with respect to MS risk, but also with respects to MS severity and symptoms. This was one of the reasons to investigate possible sex-effects in Study II with respects to spasticity treatment. Previous studies indicated that males were more likely to receive spasticity treatment.[38,161] Though initially an association with sex was observed in our study, once accounting for MS specific characteristics, sex differences were attenuated suggesting MS disability is a main influencer of spasticity treatment. It is important to be aware of possible symptomatic presentations between the sexes including their possible timing. Symptoms could be overlooked leading to treatment delay or attribution of symptoms to other causes. Differences could also arise due to variation in health-seeking behaviours between genders.[162] Men tend to delay seeking care, which could lead to initial observations of spasticity treatment being required among males with more severe MS for example, or early appearance of symptoms among females could be dismissed. Not only does sex influence possible health-seeking behaviour, but other social determinants of health such as work, family and overall resources also have influences. Social determinants of health act through complex interactions, and so far, have been not well documented in the context of MS.[163] Though a recently published study in Sweden found that two-thirds of MS patients report restrictions in their work, family-life, leisure-activities and contact with their social network.[164] These restrictions in particular were reported among women with MS, and may further influence health outcomes in MS or health-seeking behaviours. Poorer overall health can lead to additional symptoms of MS or a worsening of symptoms such as spasticity, and an increased burden of comorbid diseases.[18] Sex-specific differences in

the prevalence of comorbid diseases among MS patients, including occurrence of mental health-related diseases, have been reported in Sweden[18] which not only contribute to disease mechanisms of MS, but also complex feedback mechanisms in the social-determinants of health.[163] Additional studies of the timing of symptoms of MS and how they may be influenced by sex and other social determinants of health in conjunction with MS disease processes may help to better anticipate and meet both the pharmacological and non-pharmacological treatment needs of patients.

Possible sex-specific differences in MC1R-TUBB3

Study III demonstrated why sex-specific analyses are necessary, as the effects observed when stratifying by sex were only present among females. If only having performed the analysis without stratification, the effect observed in the joint analysis could have been interpreted as being present among both males and females in equal magnitude, which was not the case. Ensuring that causes and consequences of MS are considered separately among males and females can help to better understand MS, and why females are diagnosed with MS more often.

Possible mechanisms for the involvement of the MC1R in MS risk are still unknown though involvement of MC1Rs in various diseases was discussed in the manuscript of Study III. Typically, MC1Rs have been studied in the context of melanocytes within the skin.[96] Here, the reported biological mechanisms of MC1Rs will be discussed, along with some speculation about possible mechanisms in MS risk.

Females in general tend to have increased skin sensitivity to sunlight, and tend to burn rather than tan in response to UVR.[165] Two studies have found a difference in RHC *MC1R* mutations between the sexes, as only effects were observed among female carriers of RHC variants, though neither study had a larges study size.[101,107] Mechanisms for this difference between the sexes are unclear.

The expression of MC1Rs and the activation of subsequent pathways in the presence of UVR is a really interesting example of visible, direct interaction between gene and environment. In direct response to sunlight, the body upregulates the transcription of MC1Rs. UVR exposure helps trigger the cleavage of pro-opiomelanocortin protein (secreted by the anterior pituitary[166]) to adrenocorticotropin hormone, the direct precursor of the alpha melanocyte stimulating hormone (alpha-MSH), one of the main signalling hormones that binds MC1Rs.[95,109]The binding of alpha-MSH activates an intracellular increase of cyclic adenosine monophosphate (cAMP) triggering an increase in eumelanin production.[167] Eumelanin provides greater UVR protection in melanocytes as it visibly causes the skin to darken. In individuals with RHC *MC1R* variants, eumelanin production is impaired, and pheomelanin, a lighter pigment is produced.[167] This means that a lighter pigmentation is produced, providing less protection from UVR and causing skin

to burn in sunlight. The response of the melanocyte to activate melanin production is proportional to the amount of cAMP produced intracellularly.[109]

Not only are MC1Rs found within melanocytes and for UVR protection, but they have also been implicated in immune signalling within the skin.[168] They have been shown to reduce neuroinflammation after brain haemorrhage in rats[169,170] and slow EAE progression in mice[99] though binding of a synthetic alpha-MSH molecule, NDP-MSH to MC1Rs. alpha-MSH has been shown to be a ligand for not only MC1R, but also to other g-protein coupled receptors in the melanocortin family, MC2R (adrenocorticotropic hormone receptor), MC3R and MC4R (MC3R and MC4R are expressed mainly in the CNS).[171] These other receptors are involved in a number of functions, including adrenal functions, energy homeostasis, and food intake, sexual functions and exocrine function.[172,173] It remains unclear as to whether mechanisms upregulated after sunlight exposure also causes increased binding of ligands to other members in the melanocortin receptor family.

The possible hypotheses surrounding *MC1R* involvement in MS risk is further complicated by the fact that *TUBB3* is in close proximity to *MC1R* as it too sits just downstream of *MC1R* on chromosome 16. Alternative splicing causes formation of a *MC1R-TUBB3* chimer described by reviewed by Herriaz et al.[174] Increases in alpha-MSH has been proposed as a possible stimulator of the chimeric formation as increases in alpha-MSH concentration in vitro shifted transcription of *MC1R* to the *MC1R-TUBB3* chimer.[110] Their study described that the consequences of the chimer formation are possibly lower cell membrane expression, and a reduction in activation of cAMP production pathways as compared to non-chimers when stimulated by NDP-MSH. A similar reduction in activation was observed in the same study among several of the RHC variant MC1Rs as well.[110] Furthermore, if TUBB3 expression and function in various cells other than melanocytes is affected by the formation of MC1R-TUBB3 chimers, this could plausibly provide an alternative hypothesis as to the influence of *MC1R* variants in MS risk. TUBB3 is expressed mainly in the CNS within neurons and is involved in microtubule dynamics.[112] *TUBB3* variants have been shown to be involved in conditions such as epilepsy.[175]

TUBB3 expression has been shown to be sensitive to changes in neuronal activity, with downregulation of *TUBB3* influencing microtubule dynamics such as increase in growth. TUBB3 has been showed to be involved in microtubule dynamics, and responds to changes in neuronal activity.[112] Further hypothesis of TUBB3's involvement in MS could be due to its further role in microtubule dynamics as it has been shown to affect peripheral axon regeneration in dorsal root ganglion.[111] The same study showed that *TUBB3* knockout mice shown no apparent defects, further demonstrating possible compensatory mechanisms in upregulation of other tubulin forms in the absence or reduction of TUBB3.[111]

Though Study III found associations of *MC1R* with respects to MS risk only among females, whether it is the involvement of *MC1R* or *TUBB3* in MS development remains to be determined.

A note about investigating associations of sunlight with MS

Study III and IV initially planned to directly investigate UVR as a possible modifier of any associations observed with *MC1R-TUBB3*. Sweden specific UVR data was obtained through the Swedish Meteorological Agency as they have measured surface UVR in watts per metre squared at 10 active locations, spread over Sweden every hour since 1983.[176] Given the UVR coverage was not uniform in all regions of Sweden, the data from stations in each of three geographical regions of Sweden was combined into Norrland, Svealand and Götaland, similarly to how region of residence was categorised in Study III and IV. The idea was to calculate potential UVR exposure in different age-groups to be able to determine possible interactions with *MC1R-TUBB3*.

However, differences between the regions in terms of UVR were specific to each region with nearly no overlap when examining in cumulative UVR measures between regions. This meant that using the UVR variable versus the region of residence variable made no differences in models, and was likely measuring UVR exposure by proxy, as they were nearly 100 percent correlated. Thereby adjusting for region, one was effectively adjusting for possible UVR exposure. Additionally, given measurements of UVR began only in 1983 meant individuals born in or after 1983 could be included. This meant that there were too few individuals to adequately stratify by region in the risk analysis (given the matched nature of the data) to investigate sunlight effects, or in the severity analysis as only individuals with MS were included.

The possible specific mechanisms of how UVR or sunlight is involved in MS remains to be determined, though pathways of vitamin-D production are influenced by UVR exposure. UVR on average has increased 0.3% per year and sunlight 0.7% per year since 1985 in Sweden,[176] which may have implications for MS risk and progression. Changes in behaviour surrounding sun-protection have also changed over time as awareness of melanoma has grown over time, in particular among women who report increased sun-avoidance behaviours.[177] Increased sun-avoidance behaviour may be cancelled out, at least in part, by the increases in yearly UVR, and sun-avoidant behavioural effects in MS are unclear. Additionally, some of the key risk factors associated with MS risk such as smoking may impact sun-related behaviour. In particular, a Swedish study found associations of smoking with increased sun-exposure among women.[178] These are complex factors to tease apart, especially given the additional intertwinement of UVR directly influencing production of serum levels of 25-hydroxyvitamin D and their involvement in MS risk and progression.[179] The investigation of these factors in an observational study context is also

difficult, though mendelian randomization studies have shown that low vitamin-D are risk factors for MS,[90] which prompted its use in Study III.

Points of perspective in determining sex-specific associations in MS

Sample size issues in conducting sex-specific analyses may well be the reason for why a greater number of studies have not performed sex-specific analysis. Faced with a choice between presenting an association versus and no association if using stratification, the main association is presented. Though statistical significance does provide a type of standard to adhere to when interpreting and presenting results, this can result in negative results remaining unpublished. The clinical relevance of presenting non-statistically significant results when performing sex-specific analyses should not be ignored, especially in the field of MS when a clear sex-difference exists. One suggestion would be to provide sex-stratified analyses in supplementary material, or mention of sex-specific effects could be made a reporting requirement of studies. Sex is nearly always included as an adjusting factor or confounder. If the effects of sex are in opposite directions between sexes when including sex in the model, the result would likely be null, or if an effect is present only among one sex, it may render the effect not statistically significant. This was observed in Study III with the overall effect among all individuals was smaller than the effect among only females. The overall effect estimate was biased towards the null effect when including sexes together. However, as in Study I, the total number of exposures among males and females when stratifying analyses resulted in categorizations of fewer than 10 per age-group, and in some cases even fewer. Reporting possible associations here are unreliable, and it is unclear as to what their purpose would serve other than for speculation. A careful balance between reporting null or insignificant findings with scientific integrity should be weighed when investigating sex-specific results. Though the cross-collaboration of several countries may help to rectify the numbers problem, care should be made when considering exposures between countries with respects to their generalizability.

5.1.3 Considerations in understanding the severity of MS

Effort was made in the studies comprising this thesis to understand MS from both sides of the diagnosis. The causes of MS may not be analogous with the consequences of MS, though some overlap seem apparent, in particular with respect to factors increasing inflammation. Inflammatory activity is thought to be involved in initiation of MS pathophysiology as inflammatory lesions and markers are present[51] and include the likely involvement of pro-inflammatory markers in the breakdown of the BBB.[180] The involvement of the immune system and inflammatory activity in MS risk is evident when considering the multitude of immune-related genes that are associated with MS susceptibility. Yet, no susceptibility variants are associated with MS severity as shown in the latest MS severity GWAS (author manuscript).[86]

Study IV was unable to show an association of RHC *MC1R* variants to MS severity as measured by ARMSS. This result, although not initially hypothesized, may not be entirely unsurprising in light of the results from the MS Severity GWAS. If *MC1R*s speculated involvement in MS pathogenesis is via immune-mediated responses or inflammatory aspects, then it is possible it would not be captured in MS severity. The MS Severity GWAS found an association, and a suggestive association with ARMSS of genes involved in CNS resilience. This suggests possible neurodegenerative pathways over inflammatory pathways as immune-related genes were not associated to MS severity.[86]

It is also possible that some initial effects of identified MS susceptibility variants, or the *MC1R-TUBB3* variants as in Study III, may exert their influence in earlier years proceeding MS diagnosis, and possibly among individuals who have a relapsing versus progressive course. Increases in the severity of MS as measured by EDSS often occur during or post-relapse activity, and can return to normal, or near-normal baseline severity in the short-term, and for many years thereafter. When disability increases occur outside of short-term relapse activity, the changes are often not as drastically increased, therefore the exposure of interest and its influence for MS severity may be difficult to detect at a sufficient level for meaningful interpretation. This is especially difficult if the effect of genetic factor exerts only a small influence on MS severity overall. The ability to detect an effect is further compounded by the frequency at which an allele presents in a population. The less frequent the allele in combination with a modest effect size will make it increasingly difficult to detect associations. This combination, small exerted effect with possible low-frequency variants, is a reality of Study IV with respects to the RHC *MC1R* variants.

It may also be that EDSS, and derivatives thereof including ARMSS and the MSSS, may not be sensitive enough to measure smaller changes to MS severity, or measure various aspects of MS severity. MS leads to a multitude of additional symptoms, such as spasticity and cognitive problems, that are underrepresented in deriving EDSS. The natural variability of MS also means that many patients do not progress at a similar rate, or with a similar increase in severity when measuring EDSS. As found in one study, patients can also remain relatively stable in their EDSS scores for up to 15 years[181] though increases in severity are more rapid among PPMS patients than RRMS.[182] RRMS patients increased in mean EDSS scores by only approximately 1.5 points over a period of 15 years in another study and severity was higher among males as compared to females.[182] This is evidence of a rather small change over a long period of time that can be difficult to capture, even through use of clinically based register-data in the Swedish MS Register.

Furthermore, changes to diagnostic criteria over time means that when comparing individuals from their MS diagnosis over different time-periods, initial severity measurements could be very different. For example, today, a person would receive a MS diagnosis where 20 years ago the same person would have received a diagnosis of

clinically isolated syndrome.[19] Not only could initial EDSS scores be very different, but the disease trajectory of a current patient may remain stable for a greater number of years, especially in the era of highly-effective DMTs. Examining associations with a broader spectrum of measures of MS severity may prove to be relevant in determining the biological underpinnings of disease progression. Development of additional methods to capture aspects of MS severity may be necessary that are more sensitive to changes in both the short and long-term.

A note about GWAS in the context of spasticity

One possible measure of MS severity was thought to potentially be spasticity and GWAS of spasticity was originally planned as part of Study II. This was due to evidence that spasticity affects patients who have a progressive course or an increased disability as measured by EDSS, which we were able to partially confirm through associations with spasticity treatment in the first portion of Study II. Thus, we wanted to identify if there were genetic associations to spasticity, as it could be possible that spasticity is a marker of increasing progression. We initially believed we could use spasticity treatment as a proxy for spasticity as the outcome, measured through use of any possible spasticity treatment as defined in Study II (use of any of baclofen, gabapentin, diazepam, clonazepam or cannabinoids). In fact, we performed a GWAS that was presented at the 29th European Charcot Foundation Congress in 2021 where we showed that there were two possible suggestively associated SNPs with spasticity. However, the analysis after that presentation was expanded to try to better understand the reliability of our results, and considerations to how results were to be replicated were also discussed. We realized that the main outcome of any spasticity treatment was non-specific and overlapped greatly with possible pain and seizure treatments. We then modified the definition of spasticity treatment to the more specifically prescribed baclofen. The GWA was re-run, with results that were different from the first run with any spasticity treatment as the outcome. Stratification of the results revealed inconclusive results, given the reduced sample size which was likely not sufficiently large to be able to detect a statistically significant result. As a part of the main spasticity analyses, we also wanted to understand not only factors that influenced the start of treatment, but also how quickly individuals discontinued. Up until this second attempt at a GWA, discontinuation had not been completed. The striking, rapid decrease in spasticity prescriptions we observed raised questions about whether individuals truly had spasticity, or if they had mild spasticity for which the adverse side-effects of baclofen outweighed the benefits of continuing treatment. Through treatment patterns alone, this could not be deduced. We became additionally certain that we were measuring treatment patterns reliably, but not spasticity itself. The patients we could be most certain had spasticity were those that persevered taking baclofen, though they were extremely few. Consequently, the GWA was rendered impossible at this time, and was removed from further consideration. In order to be able

to appropriately conduct a GWAS in which we can be more certain with respects to whether individuals have spasticity, an alternative way to identify individuals with spasticity is necessary.

6 Conclusions

Through the studies of this PhD, we have contributed to the understanding of risk factors for MS aetiology and consequences in several ways. Study I and extensions thereof have shown increasing evidence for adolescence as a critical risk period in MS. Study II highlighted the importance of assessing time-related trends with respects to symptomatic treatment of spasticity among MS patients and identified that more effective or tolerable treatment alternatives may be necessary in the long-term management of MS. Study III found a female-specific genetic risk factor for MS in the *MC1R* that remained independently associated with MS after consideration of other known MS risk factors. Associations of the same *MC1R* variants were investigated after the MS diagnosis with MS severity, though no associations were found with MS severity. This highlights the relevance of examining the same factors before and after MS diagnosis to better understand their involvement in MS. Finally, all studies tried to consider sex-specific effects given the large sex-disparity in MS. Understanding whether causes and consequences of MS are different among males and females can help to better understand MS and why females are more affected.

Ultimately, central to MS research are the people affected by MS. Understanding the causes of MS can help in investigating new ways to prevent MS in the hopes of reducing the incidence of MS. Understanding whether the same causes of MS are acting to increase the severity and progression of the disease can help to treat and manage the symptoms of MS to limit restrictions to daily life. Ascertaining the timing of symptom presentation for better treatment management may help identify early signs of progression, and reduce disability.

7 Acknowledgements

It has been my absolutely pleasure in being a MS researcher these years. This process started back during my Master's thesis at Karolinska, and there are so very many people to thank that have helped me along the way. You all have very much made my time as a PhD student better in many ways. I sincerely have appreciated the support, the ideas, the fika breaks, the hiking, the knowledge sharing, the friendships, and even hugs and words of encouragement when I felt simultaneously overjoyed and very stressed at being in research. This period of my life is certainly one I will not forget.

It is such a privilege to be able to be where I am, researching MS. It would not have been possible without the contribution of the people living with MS in Sweden. Thank you for contributing your data in studies and the registers. Thank you to the research coordinators who have made it possible to do research.

To **Scott Montgomery**, my main supervisor. You have given me so much freedom and respect to contribute to research by encouraging and supporting me over these years. I came to you determined to do a PhD from the day we met, and look how far you helped me come. Not a single crazy proposal of mine did you actually call crazy, and though my many, many questions were never ending, you guided me all the way. I'm so immensely grateful you treated me as an equal. It really gave me confidence to accomplish what we have; I am so grateful you choose me to be your PhD student.

To **Pernilla Stridh**. Who know a change-up of supervisor's deep into the first year would make all the difference in the course of the PhD! You have really have been a key part of my development as a researcher, being a teacher of all things genetics, discussion of all possible mechanisms no matter how farfetched, and even a supportive ear when I've been really down. I'm so thankful for your support through these years.

To **Lars Alfredsson**. Lars, I pestered you all those years ago to have me in your research group as a Master's student, and you helped lead me to my PhD with Scott. I so wanted to do MS research, and you said yes. Without you, I'm not entirely sure where I would have ended up. Thank you for helping me to get where I am now.

To **Ingrid Kockum**. When I came to you rambling on about pigmentation genes and MS as a very enthusiastic Master's student, you helped me figure out how to do it. I can't imagine how my PhD would have taken shape without your help then, and throughout these years. Thank you so much.

Thank you to **Kyla McKay** for being an absolute remarkable person. Without you, I would have never known about the grant opportunities, had insider tips into a career as a researcher, had fewer fabulous dinners, and after-works. You're truly a golden spirit, and though you may not think you've done a lot for me, it has meant truly, so very much.

An enormous thank you to **Sarah Burkil**l, you absolutely saved my bacon in my first year as a research assistant. I miss our long wine evenings where we could just talk about anything. You had my back personally and professionally, words are simply not enough.

To **Chantelle Murley**. Who knew compliments about trousers would make us friends? But I'm so glad it did. The fika, the wine, the dinners and fun. The research, laughs, and everything in between. You've made the research process so much better.

To all my **old and new PhD friends**: you have made the PhD so much fun! **Peter Alping**, for never once saying you didn't have time for more of my coding questions or DMT questions, run buddy and being our hiking leader! **Renata Zelic** for all things Stata related, renovation conversations and late work discussions. **Andrei Barbulescu** for modelling complex time aspects, and knowing everything statistics. To **Matilda Morin** for always asking me how I am and our runs. To **Marina Dehara**, for your positivity and kindness, and the best hugs. To **Huiling Xu, Viktor Wintzell, Eleni Tsamantioti, Arda Yilal,** thanks for all the good moments.

Thank you to **Klementy Shchetynsky** for all your specialist input and hard work into all the steps leading to, and steps even within Study III and IV. They would not have been possible without you.

Thank you to **so many people at KEP**: Thank you to my financial guru, **Pernilla Appelquist**. On my worst of worst days, you have always made me feel new again, your words have meant more than you can know. You're positively fantastic. To **Helena Nord** for being simply the best HR and personal support one could ever dream to have. You're exceptional. To **Samir Hussein** who can fix any computer, even one covered in wine days before deadlines. Thank you to **Thomas Frisell**, my defence chair, but also a big help in pushing me to think outside the box way back in my Master's which lead to only a better PhD project, and making my first study better. Thank you to **Hela Westerlind**, my mentor for knowing all things administrative, the research process, and knowing what direction to point me when feeling lost. Thank you for **Michael Fored** for knowing everything PhD related, and being a listening ear with always a kind word. To **Johan Askling** for adopting me into his research group in the beginning until I found my way. There are **so many colleagues** that have helped me over the PhD, despite covid and not seeing one another so often for a while, I simply cannot thank you all enough.

Thank you to the members of **Ingrid's research group at CMM** and other **CMMers** for adopting me as an unofficially officially affiliated member of the group. **Alexandra Gyllenberg, Ali Manouchehrinia, Thomas Moridi, Xia Jiang, Qianwen Liu, Yuan Jiang,** and **Tomas Olsson**, **Fredrik Piehl.**

Thank you to **all the other collaborators and co-authors** over the years, you have been such a positive influence on my PhD.

Thank you to **Chantelle, Peter Matilda,** and **Hampus** for your hard work to proofread parts of my thesis. Thank you to the small army of practice opponents for my defence: **Matilda, Marina, Viktor, Eleni, Anton, Renata, Ian, Arda,** and **Kyla.**

Thank you to **Viia Rosén** who gave me a home away from home when I first came to Sweden. I would have been lost without you. To **Pääro Ustav** for making me so welcome, fun jokes and being the first person ever to request my published works. Suur aitäh!

To **Julia Kedrzyci**. My doubles partner in crime, my friend, and my support. Thank you for always checking in on me, and sticking with me all these years.

To **Claes-Henrik**, **Anita**, och **Jesper Ny**: Enormt stort tack för allt ni har gjort för mig under dessa år. Jag känner mig precis som en del av familjen. Ni är min svenska familj helt enkelt. Jag är så tacksam för er, och alla fina stunder. Ni har gjort det enklare att vara så långt ifrån min familj i Kanada.

To **Hampus Ny**. You are simply so special to me. You've truly been the best part of my Master's and PhD, without a doubt. I'm so grateful for you, our adventures, our fun, for hugs and laughter, and so much more to come.

And finally, **to my family**. **Mama, Dad, Ryan**. I could fill a whole book with all you have done to help me to become who I am. Endless support, endless love, endless patience, endless all. For letting me do all my crazy science experiments in the kitchen, for building literal mountains, for always answering my million questions, for helping me, teaching me, for everything.

8 References

1. Murray, T. J. & McDonald, W. I. *Multiple sclerosis : the history of a disease*. (Demos Medical Publishing, 2005).

2. Zalc, B. One hundred and fifty years ago Charcot reported multiple sclerosis as a new neurological disease. *Brain* **141**, 3482–3488 (2018).

3. Olsson, T., Barcellos, L. F. & Alfredsson, L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol* **13**, 25–36 (2016).

4. Ramagopalan, S. V., Dobson, R., Meier, U. C. & Giovannoni, G. Multiple sclerosis: risk factors, prodromes, and potential causal pathways. *Lancet Neurol* **9**, 727–739 (2010).

5. The Multiple Sclerosis International Federation (MSIF). Atlas of MS 3 rd edition Part 1: Mapping multiple sclerosis around the world key epidemiology findings. *The Multiple Sclerosis International Federation (MSIF), September 2020* 1–37 (2020).

6. Koch-Henriksen, N. & Sørensen, P. S. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol* **9**, 520–532 (2010).

7. Walton, C. *et al.* Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Multiple Sclerosis Journal* **26**, 1816–1821 (2020).

8. Brain, W. R. Critical Review: Disseminated Sclerosis. *QJM: An International Journal of Medicine* **os-23**, 343–391 (1930).

9. Krysko, K. M. *et al.* Sex effects across the lifespan in women with multiple sclerosis. *Ther Adv Neurol Disord* **13**, (2020).

10. Ernstsson, O. *et al.* Cost of Illness of Multiple Sclerosis - A Systematic Review. *PLoS One* **11**, 1–25 (2016).

11. Confavreux, C. & Vukusic, S. Chapter 15 – The clinical course of multiple sclerosis. in *Handbook of Clinical Neurology* vol. 122 343–369 (2014).

12. Kobelt, G. *et al.* New insights into the burden and costs of multiple sclerosis in Europe. *Mult Scler* **23**, 1123–1136 (2017).

13. Gyllensten, H. *et al.* Costs of illness of multiple sclerosis in Sweden: a population-based register study of people of working age. *European Journal of Health Economics* **19**, 435–446 (2018).

14. Wallin, M. T. *et al.* Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **18**, 269–285 (2019).

15. Burkill, S. *et al.* Mortality trends for multiple sclerosis patients in Sweden from 1968 to 2012. *Neurology* **89**, 555–562 (2017).

16. Filippi, M. *et al.* Multiple sclerosis. *Nat Rev Dis Primers* **4**, 1–27 (2018).

17. Scalfari, A. *et al.* Mortality in patients with multiple sclerosis. *Neurology* **81**, 184–192 (2013).

18. Smith, K. A. *et al.* Comorbid disease burden among MS patients 1968–2012: A Swedish register–based cohort study. *Multiple Sclerosis Journal* 1–13 (2020) doi:10.1177/1352458520910497.

19. Sorensen, P. S. *et al.* The apparently milder course of multiple sclerosis: changes in the diagnostic criteria, therapy and natural history. *Brain* **143**, 2637–2652 (2020).

20. Jarius, S. *et al.* Pattern II and pattern III MS are entities distinct from pattern I MS: evidence from cerebrospinal fluid analysis. *Journal of Neuroinflammation 2017 14:1* **14**, 1–14 (2017).

21. Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurology* **17**, 162–73 (2018).

22. Chalmer, T. A. *et al.* Early versus later treatment start in multiple sclerosis: a register-based cohort study. *Eur J Neurol* **25**, 1262-e110 (2018).

23. Multiple Sclerosis International Federation. Atlas of MS 3rd Edition - Sweden Factsheet. https://www.atlasofms.org/fact-sheet/sweden (2020).

24. Dendrou, C. A., Fugger, L. & Friese, M. A. Immunopathology of multiple sclerosis. *Nature Publishing Group* (2015) doi:10.1038/nri3871.

25. Lorscheider, J. *et al.* Defining secondary progressive multiple sclerosis. *Brain* **139**, 2395–2405 (2016).

26. Cree, B. A. C. *et al.* Secondary Progressive Multiple Sclerosis. *Neurology* **97**, 378–388 (2021).

27. Magyari, M. & Koch-Henriksen, N. Quantitative effect of sex on disease activity and disability accumulation in multiple sclerosis. *J Neurol Neurosurg Psychiatry* **93**, 716–722 (2022).

28. Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis. *Neurology* **33**, 1444–1444 (1983).

29. Meyer-Moock, S., Feng, Y. S., Maeurer, M., Dippel, F. W. & Kohlmann, T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol* **14**, 1–10 (2014).

30. Roxburgh, R. H. S. R. *et al.* Multiple Sclerosis Severity Score. *Neurology* **64**, 1144–1151 (2005).

31. Manouchehrinia, A. *et al.* Age Related Multiple Sclerosis Severity Score: Disability ranked by age. *Mult Scler* **23**, 1938–1946 (2017).

32. Marrie, R. A. & Horwitz, R. I. Emerging effects of comorbidities on multiple sclerosis. *Lancet Neurol* **9**, 820–828 (2010).

33. Synnot, A. *et al.* Interventions for managing skeletal muscle spasticity following traumatic brain injury. *Cochrane Database of Systematic Reviews* vol. 2017 Preprint at https://doi.org/10.1002/14651858.CD008929.pub2 (2017).

34. Sommerfeld, D. K., Eek, E. U. B., Svensson, A. K., Holmqvist, L. W. & Von Arbin, M. H. Spasticity after Stroke: Its Occurrence and Association with Motor Impairments and Activity Limitations. *Stroke* **35**, 134–139 (2004).

35. Trompetto, C. *et al.* Pathophysiology of spasticity: Implications for neurorehabilitation. *BioMed Research International* vol. 2014 Preprint at https://doi.org/10.1155/2014/354906 (2014).

36. Bar-On, L. *et al.* Spasticity and Its Contribution to Hypertonia in Cerebral Palsy. (2015) doi:10.1155/2015/317047.

37. Verschueren, A., Grapperon, A. M., Delmont, E. & Attarian, S. Prevalence of spasticity and spasticity-related pain among patients with Amyotrophic Lateral Sclerosis. *Rev Neurol (Paris)* **177**, 694–698 (2021).

38. Rizzo, M. A., Hadjimichael, O. C., Preiningerova, J. & Vollmer, T. L. Prevalence and treatment of spasticity reported by multiple sclerosis patients. *Multiple Sclerosis* **10**, 589–595 (2004).

39. Malhotra, S. *et al.* An investigation into the agreement between clinical, biomechanical and neurophysiological measures of spasticity. *Clin Rehabil* **22**, 1105–1115 (2008).

40. Bethoux, F. & Marrie, R. A. A Cross-Sectional Study of the Impact of Spasticity on Daily Activities in Multiple Sclerosis. *Patient* **9**, 537–546 (2016).

41. Patejdl, R. & Zettl, U. K. Autoimmunity Reviews Spasticity in multiple sclerosis: Contribution of inflammation, autoimmune mediated neuronal damage and therapeutic interventions. *Autoimmun Rev* **16**, 925–936 (2017).

42. Gallien, P., Gich, J., Sánchez-Dalmau, B. F. & Feneberg, W. Multidisciplinary management of multiple sclerosis symptoms. *Eur Neurol* **72**, 20–25 (2014).

43. Haselkorn, J. K. & Loomis, S. Multiple Sclerosis and Spasticity. **16**, 467–481 (2005).

44. Milligan, J., Ryan, K. & Lee, J. Spasticity in Primary Care. *Canadian Family Physician* **65**, 697–703 (2019).

45. Aboud, T. & Schuster, N. M. Pain Management in Multiple Sclerosis: a Review of Available Treatment Options. *Curr Treat Options Neurol* **21**, (2019).

46. (Läkemedelsverket), P. R. A. S. Läkemedelsbehandling av multipel skleros (MS) – behandlingsrekommendation i Sverige [Pharmaceutical treatment of MS - Treatment guidelines in Sweden]. 13–25 (2015).

47. Otero-Romero, S. *et al.* Pharmacological management of spasticity in multiple sclerosis: Systematic review and consensus paper. *Multiple Sclerosis* **22**, 1386–1396 (2016).

48. Hugos, C. L. & Cameron, M. H. MS Spasticity: Take Control (STC) for ambulatory adults: Protocol for a randomized controlled trial. *BMC Neurol* **20**, (2020).

49. Shakespeare, D., Boggild, M. & Young, C. Anti-spasticity agents for multiple sclerosis (Cochrane review). *Cochrane Database of Systematic Reviews* (2010) doi:10.1002/14651858.CD001332.www.cochranelibrary.com.

50. Skierlo, S., Rommer, P. S. & Zettl, U. K. Symptomatic treatment in multiple sclerosis-interim analysis of a nationwide registry. *Acta Neurol Scand* **135**, 394–399 (2017).

51. Kutzelnigg, A. & Lassmann, H. Chapter 2 – Pathology of multiple sclerosis and related inflammatory demyelinating diseases. in *Handbook of Clinical Neurology* vol. 122 15–58 (2014).

52. Greenfield, A. L. & Hauser, S. L. B Cell Therapy for Multiple Sclerosis: Entering an Era. *Ann Neurol* **83**, 13 (2018).

53. Odoardi, F. *et al.* T cells become licensed in the lung to enter the central nervous system. *Nature* **488**, 675–679 (2012).

54. Tremlett, H. & Marrie, R. A. The multiple sclerosis prodrome: Emerging evidence, challenges, and opportunities. *Multiple Sclerosis Journal* 1–7 (2020) doi:10.1177/1352458520914844.

55. Wijnands, J. M. A. *et al.* Health-care use before a first demyelinating event suggestive of a multiple sclerosis prodrome: a matched cohort study. *Lancet Neurol* **16**, 445–451 (2017).

56. Wijnands, J. M. A. *et al.* Five years before multiple sclerosis onset: Phenotyping the prodrome. *Multiple Sclerosis Journal* 1–10 (2018) doi:10.1177/1352458518783662.

57. Yusuf, F. L. A. *et al.* A systematic review of morbidities suggestive of the multiple sclerosis prodrome. *https://doi-org.proxy.kib.ki.se/10.1080/14737175.2020.1746645* **20**, 799–819 (2020).

58. Makhani, N. & Tremlett, H. The multiple sclerosis prodrome. *Nature Reviews Neurology 2021 17:8* **17**, 515–521 (2021).

59. Montgomery, S. *et al.* Reply to " Concussion May Not Cause Multiple Sclerosis ". *Ann Neurol* **82**, 652–653 (2017).

60. Xu, Y. *et al.* Hospital-diagnosed infections before age 20 and risk of a subsequent multiple sclerosis diagnosis. *Brain* (2021) doi:10.1093/brain/awab100.

61. Montgomery, S. *et al.* Concussion in adolescence and risk of multiple sclerosis. *Ann Neurol* **82**, 554–561 (2017).

62. Bäärnhielm, M. *et al.* Sunlight is associated with decreased multiple sclerosis risk: no interaction with human leukocyte antigen-DRB1*15. *Eur J Neurol* **19**, 955–962 (2012).

63. Hedström, A. K., Hillert, J., Olsson, T. & Alfredsson, L. Smoking and multiple sclerosis susceptibility. *Eur J Epidemiol* **28**, 867–874 (2013).

64. Hedström, A. K., Bäärnhielm, M., Olsson, T. & Alfredsson, L. Exposure to environmental tobacco smoke is associated with increased risk for multiple sclerosis. *Multiple Sclerosis Journal* **17**, 788–793 (2011).

65. Thacker, E. L., Mirzaei, F. & Ascherio, A. Infectious mononucleosis and risk for multiple sclerosis: A meta-analysis. *Ann Neurol* **59**, 499–503 (2006).

66. Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J. P. A. & Tzoulaki, I. Environmental risk factors and multiple sclerosis: An umbrella review of systematic reviews and meta-analyses. *Lancet Neurol* **14**, 263–273 (2015).

67. Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (1979)* **375**, 296–301 (2022).

68. Correale, J., Balbuena Aguirre, M. E. & Farez, M. F. Sex-specific environmental influences affecting MS development. *Clinical Immunology* **149**, 176–181 (2013).

69. Patsopoulos, N. A. *et al.* Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science (1979)* **365**, (2019).

70. Xu, Y. *et al.* Higher body mass index at ages 16 to 20 years is associated with increased risk of a multiple sclerosis diagnosis in subsequent adulthood among men. *Multiple Sclerosis Journal* (2020) doi:10.1177/1352458520928061.

71. Montgomery, S. *et al.* Concussion in adolescence and risk of multiple sclerosis. *Ann Neurol* **82**, 554–561 (2017).

72. Hedström, A. K. *et al.* Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. *Neurology* **82**, 865–872 (2014).

73. Jacobs, B. M. *et al.* Gene-Environment Interactions in Multiple Sclerosis: A UK Biobank Study. *Neurology(R) neuroimmunology & neuroinflammation* **8**, (2021).

74. Rotstein, D. L. *et al.* MS risk in immigrants in the McDonald era. *Neurology* **93**, e2203–e2215 (2019).

75. Montgomery, S., Hillert, J. & Bahmanyar, S. Hospital admission due to infections in multiple sclerosis patients. *Eur J Neurol* **20**, 1153–1160 (2013).

76. Brand, J. S., Smith, K. A., Piehl, F., Olsson, T. & Montgomery, S. Risk of serious infections in multiple sclerosis patients by disease course and disability status: Results from a Swedish register-based study. *Brain Behav Immun Health* **22**, (2022).

77. Buljevac, D. *et al.* Prospective study on the relationship between infections and multiple sclerosis exacerbations. **125**, 952–960 (2002).

78. Sawcer, S., Franklin, R. J. M. & Ban, M. Multiple sclerosis genetics. *Lancet Neurol* **13**, 700–709 (2014).

79. Westerlind, H. *et al.* Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. *Brain* **137**, 770 (2014).

80. Fagnani, C. *et al.* Twin studies in multiple sclerosis: A meta-estimation of heritability and environmentality. *Mult Scler* **21**, 1404–1413 (2015).

81. Olerup, O. & Hillert, J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* **38**, 1–15 (1991).

82. Hillert, J. Human leukocyte antigen studies in multiple sclerosis. *Ann Neurol* **36**, S15–S17 (1994).

83. Brynedal, B. *et al.* HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS One* **2**, 1–5 (2007).

84. Moutsianas, L. *et al.* Class II HLA interactions modulate genetic risk for multiple sclerosis. **47**, 1107–1113 (2015).

85. Sawcer, S. The genetic aspects of multiple sclerosis. *Ann Indian Acad Neurol* **12**, 206–14 (2009).

86. Baranzini, S. & Sawcer, S. Genetic analysis of multiple sclerosis severity identifies a novel locus and implicates CNS resilience as a major determinant of outcome. (2022) doi:10.21203/RS.3.RS-1723574/V1.

87. Hogancamp, W. E., Rodriguez, M. & Weinshenker, B. G. The epidemiology of multiple sclerosis. *Mayo Clin.Proc.* **72**, 871–878 (1997).

88. World Health Organization. *Neurological disorders: public health challenges*. *Neurological disorders: public health challenges*. (2006). doi:10.1001/archneurol.2007.19.

89. Ahlgren, C., Odén, A. & Lycke, J. High nationwide prevalence of multiple sclerosis in Sweden. *http://dx.doi.org/10.1177/1352458511403794* **17**, 901–908 (2011).

90. Jacobs, B. M., Noyce, A. J., Giovannoni, G. & Dobson, R. BMI and low vitamin D are causal factors for multiple sclerosis: A Mendelian Randomization study. *Neurology(R) neuroimmunology & neuroinflammation* **7**, (2020).

91. Becklund, B. R., Severson, K. S., Vang, S. V. & DeLuca, H. F. UV radiation suppresses experimental autoimmune encephalomyelitis independent of vitamin D production. *Proceedings of the National Academy of Sciences* **107**, 6418–6423 (2010).

92. Spach, K. M. & Hayes, C. E. Vitamin D3 confers protection from autoimmune encephalomyelitis only in female mice. *J Immunol* **175**, 4119–4126 (2005).

93. Hochmeister, S. *et al.* Effect of Vitamin D on Experimental Autoimmune Neuroinflammation Is Dependent on Haplotypes Comprising Naturally Occurring Allelic Variants of CIITA (Mhc2ta). *Front Neurol* **11**, (2020).

94. Rhead, B. *et al.* Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. *Neurol. Genet.* **2**, e97 (2016).

95. Beaumont, K. A., Liu, Y. Y. & Sturm, R. A. *The Melanocortin-1 Receptor Gene Polymorphism and Association with Human Skin Cancer*. *Progress in Molecular Biology and Translational Science* vol. 88 (Elsevier Inc., 2009).

96. Cassidy, P. B., Abdel-Malek, Z. A. & Leachman, S. A. Beyond red hair and sunburns: Uncovering the molecular mechanisms of MC1R signaling and repair of UV-induced DNA damage. *Journal of Investigative Dermatology* **135**, 2918–2921 (2015).

97. Rouzaud, F., Kadekaro, A. L., Abdel-Malek, Z. A. & Hearing, V. J. MC1R and the response of melanocytes to ultraviolet radiation. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **571**, 133–152 (2005).

98. Catania, A. *Melanocortins: Multiple actions and therapeutic potential*. (Landers Bioscience and Springet Science+Business Media, LLC, 2010).

99. Mykicki, N. *et al*. Melanocortin-1 receptor activation is neuroprotective in mouse models of neuroinflammatory disease. *Sci Transl Med* **8**, 146–362 (2016).

100. Benjamins, J. A., Nedelkoska, L. & Lisak, R. P. Melanocortin receptor subtypes are expressed on cells in the oligodendroglial lineage and signal ACTH protection. *J Neurosci Res* 1–9 (2017) doi:10.1002/jnr.24141.

101. Dwyer, T. *et al*. Melanocortin 1 receptor genotype, past environmental sun exposure, and risk of multiple sclerosis. *Neurology* **71**, 583–589 (2008).

102. Ostkamp, P. *et al*. Sunlight exposure exerts immunomodulatory effects to reduce multiple sclerosis severity. *Proc Natl Acad Sci U S A* **118**, (2021).

103. Erratum: Correction for Ostkamp et al., Sunlight exposure exerts immunomodulatory effects to reduce multiple sclerosis severity (Proceedings of the National Academy of Sciences of the United States of America (2021) 118 1 PII: e2110306118). *Proceedings of the National Academy of Sciences of the United States of America* vol. 118 Preprint at https://doi.org/10.1073/pnas.2110306118 (2021).

104. Strange, R. C. *et al*. The Multiple Sclerosis Severity Score: associations with MC1R single nucleotide polymorphisms and host response to ultraviolet radiation. *Multiple Sclerosis Journal* **16**, 1109–1116 (2010).

105. Ramachandran, S. *et al*. Progression of disability in multiple sclerosis: A study of factors influencing median time to reach an EDSS value. *Mult Scler Relat Disord* **2**, 109–116 (2013).

106. Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte–stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet* **11**, 328–330 (1995).

107. Partridge, J. M. *et al*. Susceptibility and outcome in MS: associations with polymorphisms in pigmentation-related genes. *Neurology* **62**, 2323–2325 (2004).

108. Hepp, D., Gonçalves, G. L. & Freitas, T. R. O. de. Prediction of the Damage-Associated Non-Synonymous Single Nucleotide Polymorphisms in the Human MC1R Gene. *PLoS One* **10**, e0121812 (2015).

109. Wolf Horrell, E. M., Boulanger, M. C. & D'Orazio, J. A. Melanocortin 1 receptor: Structure, function, and regulation. *Front Genet* **7**, 1–16 (2016).

110. Herraiz, C. *et al.* Functional Characterization of MC1R-TUBB3 Intergenic Splice Variants of the Human Melanocortin 1 Receptor. *PLoS One* **10**, (2015).

111. Latremoliere, A. *et al.* Neuronal-Specific TUBB3 Is Not Required for Normal Neuronal Function but Is Essential for Timely Axon Regeneration. *Cell Rep* **24**, (2018).

112. Radwitz, J. *et al.* Tubb3 expression levels are sensitive to neuronal activity changes and determine microtubule growth and kinesin-mediated transport. *Cell Mol Life Sci* **79**, (2022).

113. Hillert, J. & Stawiarz, L. The Swedish MS registry - clinical support tool and scientific resource. *Acta Neurol Scand* **132**, 11–19 (2015).

114. Andersen, O. From the Gothenburg cohort to the Swedish multiple sclerosis registry. *Acta Neurol Scand* **126**, 13–19 (2012).

115. Alping, P., Piehl, F., Langer-Gould, A. & Frisell, T. Validation of the Swedish Multiple Sclerosis Register: Further Improving a Resource for Pharmacoepidemiologic Evaluations. *Epidemiology* **30**, 230–233 (2019).

116. Ludvigsson, J. F. *et al.* External review and validation of the Swedish national inpatient register. *BMC Public Health* **11**, 450 (2011).

117. Wettermark, B. *et al.* The new Swedish Prescribed Drug Register - Opportunities for pharmcoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* **16**, 726–735 (2007).

118. Ludvigsson, J. F. *et al.* Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* **31**, 125–136 (2016).

119. Ho, J. *et al.* The Swedish cause of death register. 765–773 (2017) doi:10.1007/s10654-017-0316-1.

120. Ludvigsson, J. F., Svedberg, P., Olén, O., Bruze, G. & Neovius, M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *Eur J Epidemiol* **34**, 423–437 (2019).

121. Hedström, A., Olsson, T. & Alfredsson, L. Smoking is a major preventable risk factor for multiple sclerosis. *Multiple Sclerosis Journal* **22**, 1021–1026 (2016).

122. Holmén, C. *et al.* A Swedish national post-marketing surveillance study of natalizumab treatment in multiple sclerosis. *Mult Scler* **17**, 708–719 (2011).

123. Olafsson, S. *et al.* Fourteen sequence variants that associate with multiple sclerosis discovered by meta-analysis informed by genetic correlations. *NPJ Genom Med* **2**, (2017).

124. Verlouw, J. A. M. *et al.* A comparison of genotyping arrays. *European Journal of Human Genetics* **29**, 1611–1624 (2021).

125. Kockum, I., Huang, J. & Stridh, P. Overview of Genotyping Technologies and Methods. *Curr Protoc* **3**, e727 (2023).

126. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* **34**, 591 (2010).

127. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387–406 (2009).

128. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics 2016 48:10* **48**, 1279–1283 (2016).

129. Freedman, M. L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nature Genetics 2004 36:4* **36**, 388–393 (2004).

130. Reich, D., Price, A. L. & Patterson, N. Principal component analysis of genetic data. *Nature Genetics 2008 40:5* **40**, 491–492 (2008).

131. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**, 100–106 (2014).

132. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82 (2011).

133. Johansson, Å., Ingman, M., Mack, S. J., Erlich, H. & Gyllensten, U. Genetic origin of the Swedish Sami inferred from HLA class I and class II allele frequencies. *European Journal of Human Genetics 2008 16:11* **16**, 1341–1349 (2008).

134. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med* **12**, 1–11 (2020).

135. Jiang, X. *et al.* Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nature Communications 2018 9:1* **9**, 1–12 (2018).

136. Knol, M. J., Vandenbroucke, J. P., Scott, P. & Egger, M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am J Epidemiol* **168**, 1073–1081 (2008).

137. Rose, S. & Van Der Laan, M. J. *Why Match? Investigating Matched Case-Control Study Designs with Causal Effect Estimation Why Match? Investigating Matched Case-Control Study Designs with Causal Effect Estimation *. The International Journal of Biostatistics* vol. 5 (2009).

138. Hennessy, S., Bilker, W. B., Berlin, J. A. & Strom, B. L. Factors influencing the optimal control-to-case ratio in matched case-control studies. *Am J Epidemiol* **149**, 195–197 (1999).

139. Smith, K. A. *et al.* Hospital diagnosed pneumonia before age 20 years and multiple sclerosis risk. *BMJ Neurol Open* **2**, e000044 (2020).

140. Wijnands, J. M. A. *et al.* Infection-related health care utilization among people with and without multiple sclerosis. *Multiple Sclerosis* **23**, 1506–1516 (2017).

141. Pearce, N. Analysis of matched case-control studies. *BMJ (Online)* **352**, (2016).

142. Breslow, N. E., Day, N. E., Halvorsen, K. T., Prentice, R. L. & Sabai, C. Estimation of multiple relative risk functions in matched case-control studies. *Am J Epidemiol* **108**, 299–307 (1978).

143. Murley, C., Friberg, E., Hillert, J., Alexanderson, K. & Yang, F. Validation of multiple sclerosis diagnoses in the Swedish National Patient Register. *Eur J Epidemiol* **34**, 1161–1169 (2019).

144. Bahmanyar, S., Montgomery, S. M., Hillert, J., Ekbom, A. & Olsson, T. *Cancer risk among patients with multiple sclerosis and their parents*. (2009).

145. Smith, K. A. *et al.* Spasticity treatment patterns among people with multiple sclerosis: a Swedish cohort study. *J Neurol Neurosurg Psychiatry* jnnp-2022-329886 (2022) doi:10.1136/JNNP-2022-329886.

146. Cox, D. Regression Models and Life-Tables. *Journal of the Royal Statistical Society*. **34**, 187–220 (1972).

147. Zhang, Z., Reinikainen, J., Adeleke, K. A., Pieterse, M. E. & Groothuis-Oudshoorn, C. G. M. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med* **6**, 121–121 (2018).

148. Bellera, C. A. *et al. Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer*. http://www.biomedcentral.com/1471-2288/10/20 (2010).

149. Schoenfeld, D. Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239–241 (1982).

150. Kleinbaum, D. G. & Klein, M. Survival Analysis. (2012) doi:10.1007/978-1-4419-6646-9.

151. Hernán, M. A. The hazards of hazard ratios. *Epidemiology* vol. 21 13–15 Preprint at https://doi.org/10.1097/EDE.0b013e3181c1ea43 (2010).

152. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294–305 (2011).

153. Revez, J. A. *et al.* Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nature Communications 2020 11:1* **11**, 1–12 (2020).

154. Stewardson, A., Allegranzi, B., Sax, H., Kilpatrick, C. & Pittet, D. Back to the future: rising to the Semmelweis challenge in hand hygiene. *Future Microbiol* **6**, 855–876 (2011).

155. Dean, G. & Kurtzke, J. F. On the risk of multiple sclerosis according to age at immigration to South Africa. *Br Med J* **3**, 725–729 (1971).

156. Alter, M., Leibowitz, U. & Speer, J. Risk of multiple sclerosis related to age at immigration to Israel. *Arch Neurol* **15**, 234–237 (1966).

157. Xu, Y. *et al.* Association of Infectious Mononucleosis in Childhood and Adolescence With Risk for a Subsequent Multiple Sclerosis Diagnosis Among Siblings. *JAMA Netw Open* **4**, e2124932–e2124932 (2021).

158. Kuri, A. *et al.* Epidemiology of Epstein-Barr virus infection and infectious mononucleosis in the United Kingdom. *BMC Public Health* **20**, 1–9 (2020).

159. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nature Reviews Immunology 2016 16:10* **16**, 626–638 (2016).

160. Yang, Y. & Kozloski, M. Sex differences in age trajectories of physiological dysregulation: inflammation, metabolic syndrome, and allostatic load. *J Gerontol A Biol Sci Med Sci* **66**, 493–500 (2011).

161. Milinis, K., Tennant, A. & Young, C. A. Spasticity in multiple sclerosis: Associations with impairments and overall quality of life. *Mult Scler Relat Disord* **5**, 34–39 (2016).

162. Thompson, A. E. *et al.* The influence of gender and other patient characteristics on health care-seeking behaviour: A QUALICOPC study. *BMC Fam Pract* **17**, 1–7 (2016).

163. Dobson, R. *et al.* Social determinants of health in multiple sclerosis. *Nature Reviews Neurology 2022 18:12* **18**, 723–734 (2022).

164. Sebsibe Teni, F. *et al.* Self-reported restrictions in different life domains and associated factors among people with multiple sclerosis in Sweden. *Eur J Neurol* **00**, 1–11 (2023).

165. Hernando, B. *et al.* Sex and MC1R variants in human pigmentation. Differences in tanning ability and sensitivity to sunlight between sexes. *J Dermatol Sci* **84**, 346–360 (2016).

166. Cawley, N. X., Li, Z. & Loh, Y. P. Biosynthesis, Trafficking and Secretion of Pro-opiomelanocortin-derived peptides. *J Mol Endocrinol* **56**, T77 (2016).

167. García-Borrón, J. C., Abdel-Malek, Z. & Jiménez-Cervantes, C. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. *Pigment Cell Melanoma Res* **27**, 699–720 (2014).

168. Chen, W., Li, J. & Zhou, Q. The melanocortin 1 receptor (MC1R) inhibits the inflammatory response in Raw 264 . 7 cells and atopic dermatitis (AD) mouse model. **6**, 1987–1996 (2013).

169. Wu, X. *et al.* NDP-MSH binding melanocortin-1 receptor ameliorates neuroinflammation and BBB disruption through CREB/Nr4a1/NF-κB pathway after intracerebral hemorrhage in mice. *J Neuroinflammation* **16**, 1–13 (2019).

170. Gatti, S. *et al.* Protective action of NDP-MSH in experimental subarachnoid hemorrhage. *Exp Neurol* **234**, 230–238 (2012).

171. Switonski, M., Mankowska, M. & Salamon, S. Family of melanocortin receptor (MCR) genes in mammals—mutations, polymorphisms and phenotypic effects. *J Appl Genet* **54**, 461 (2013).

172. Yang, Y. Structure, function and regulation of the melanocortin receptors. *Eur J Pharmacol* **660**, 125 (2011).

173. Wikberg, J. E. S. & Mutulis, F. Targeting melanocortin receptors: an approach to treat weight disorders and sexual dysfunction. *Nature Reviews Drug Discovery 2008 7:4* **7**, 307–323 (2008).

174. Herraiz, C., Garcia-Borron, J. C., Jiménez-Cervantes, C. & Olivares, C. MC1R signaling. Intracellular partners and pathophysiological implications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1863**, 2448–2461 (2017).

175. Xu, X. *et al.* Tubulin β-III modulates seizure activity in epilepsy. *J Pathol* **242**, 297–308 (2017).

176. Carlund, T. Upgrade of SMHI's meteorological radiation network 2006-2007 Effects on direct and global solar radiation. **148**, (2011).

177. McKenzie, C., Nahm, W. J., Kearney, C. A. & Zampella, J. G. Sun-protective behaviors and sunburn among US adults. *Arch Dermatol Res* 1 (2023) doi:10.1007/S00403-023-02547-Z.

178. Scragg, R., Sandin, S., Löf, M., Adami, H. O. & Weiderpass, E. Associations between sun exposure and other lifestyle variables in Swedish women. *Cancer Causes and Control* **28**, 985–996 (2017).

179. Lucas, R. M., Byrne, S. N., Correale, J., Ilschner, S. & Hart, P. H. Ultraviolet radiation, vitamin D and multiple sclerosis. *Neurodegenerative disease management* vol. 5 413–424 Preprint at https://doi.org/10.2217/nmt.15.33 (2015).

180. Ortiz, G. G. *et al.* Role of the blood-brain barrier in multiple sclerosis. *Arch Med Res* **45**, 687–697 (2014).

181. Hum, S., Lapierre, Y., Scott, S. C., Duquette, P. & Mayo, N. E. Trajectory of MS disease course for men and women over three eras. *Mult Scler* **23**, 534–545 (2017).

182. Ribbons, K. A. *et al.* Male Sex Is Independently Associated with Faster Disability Accumulation in Relapse-Onset MS but Not in Primary Progressive MS. *PLoS One* **10**, (2015).