Chapter 2

Genetic epidemiology of Multiple Sclerosis: a review of the literature

Maartje Boon
Jacques De Keyser
Charles H.C.M. Buys
Gerard J. te Meerman

Submitted for publication
Multiple Sclerosis (MS) is a disease characterized by inflammation and demyelination of the central nervous system (CNS). The severity and localization of the symptoms are very variable, from little or no residual symptoms over the years to severe handicap or even death within a few weeks to months. MS is a major cause of disability among young adults. In the pathogenesis, autoimmunity is likely to play a major role. Proteins that form part of the myelin sheath are suspected to be the target of the immunological reaction. However, some authors examine an alternative hypothesis of pathogenesis, suggesting that MS is a neurodegenerative and metabolic disorder with polygenic influence.

The etiology of MS is unknown, but there is evidence for both genetic and environmental factors playing a role in susceptibility to MS. Indications for environmental factors include the correlation between prevalence and geographic localization, migration studies, possible clustering and reports of specific antibodies in body fluids. The available epidemiologic data were recently reviewed by Marrie.

For almost all diseases, the genetic make-up of an individual will to some extent determine susceptibility to and expression of a given disease. Some diseases are entirely genetic in origin (e.g. myotonic dystrophy), although environmental factors may influence the course of the disease. In other diseases the balance weighs to the other side, e.g. in infectious diseases where an individual can be more or less susceptible to the external pathogen that is the primary cause of the disease. The influence of genes in susceptibility to a certain disease can be investigated by observational epidemiological studies (e.g. twin studies), association and linkage studies and alternative methods like haplotype sharing analysis.
2.1 Observational studies

2.1.1 Studies of familial aggregation

Over a century ago Eichhorst called MS an inherited, transmissible disease. The main indication for this hypothesis was the observation that MS tended to show familial aggregation. Studies of familial aggregation examine whether the disease prevalence in genetically related family members of affected individuals is increased compared with the prevalence of the disease in the general population. If so, this is an indication for the influence of genetic factors on susceptibility to the disease.

A chi-squared test of association is used to determine whether there is an increased disease frequency in the relatives of affected individuals as compared with controls. A 2x2 table is analyzed for a difference in proportions. A significant result indicates an increased (or decreased) incidence of disease in relatives of affected individuals as compared with controls.

Another statistic in this context is the familial relative risk (RR), which has the advantage that it quantifies the degree of risk to relatives. The RR is also denoted as $\lambda$. It is calculated as the ratio of the disease rate in relatives of affected individuals to the disease rate in relatives of controls. A RR significantly larger than 1 indicates that the disease is more prevalent in the relatives of affected individuals than in the general population (inferred from the controls).

The RR can be underestimated because of incomplete or reduced penetrance of the disease or high frequency of the disease susceptibility gene. Interesting in this respect are the studies that show biochemical and radiological manifestations of MS in clinically healthy sibs of MS patients (see paragraph “Subclinical MS?”). Overestimation can occur because of ascertainment bias: patients with a disease are more likely to know of individuals in their family with the same disease than unaffected controls are. An alternative for the familial RR is the population RR, where the probability of a relative of an affected individual being affected is compared with the probability of a random member of the population being affected.

In many of the classical family studies, the observations on recurrence of MS in families are reported as the frequency of familial cases, i.e. the proportion of MS index cases that have a close relative with MS. This parameter, however, also depends on the family sizes in the population and the mode of inheritance. Therefore, a better parameter to use is the prevalence of MS among specific types of relatives in proportion to the total number of relatives with the same family relation. In order to examine whether the
prevalence among relatives is increased, it needs to be compared to the prevalence in the population concerned and ideally corrected for age and sex.

Table 1: Comparison of age-adjusted lifetime risks by relationship to the proband - data for a Northern European population living in a temperate climate.

<table>
<thead>
<tr>
<th>Relationship to proband</th>
<th>Approximate relative risk (%)</th>
<th>Relative risk to general population</th>
<th>% Genetic sharing with the proband</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population</td>
<td>0.2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>First-degree relative</td>
<td>3.5</td>
<td>15-25</td>
<td>50</td>
</tr>
<tr>
<td>Dizygotic twin</td>
<td>3.5</td>
<td>15-25</td>
<td>50</td>
</tr>
<tr>
<td>Monozygotic twin</td>
<td>28</td>
<td>190</td>
<td>100</td>
</tr>
<tr>
<td>'Adopted' first-degree relative</td>
<td>0.2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Half-sib</td>
<td>1.3</td>
<td>6.5</td>
<td>25</td>
</tr>
<tr>
<td>Offspring of conjugal MS</td>
<td>29.5</td>
<td>147.5</td>
<td>50(^{b})</td>
</tr>
</tbody>
</table>

\(^a\) It may be more appropriate to compare crude rates for this group

\(^b\) The child shares 50% of the genetic material with the affected mother and 50% of the genetic material with the affected father (From Sadovnick et al.)

The recurrence risk for relatives of MS patients is increased proportional to the degree of kinship, probably as a manifestation of the amount of DNA shared. The risk for relatives is increased, depending on the sex of the index patient and the relative, the type of kinship and the age of the relative. The age-adjusted recurrence rate for first-degree relatives appears to be 1.3-5%. Since the population risk is 0.1-0.2%, this results in an RR (\(\lambda_r\), recurrence risk of relative/population risk) up to 40 or 50. The relative risk for second degree relatives is around 3 and in third degree relatives the risk is with 0.3 – 1.5% still considerably higher than the population risk.\(^{2,3}\)
2.1.2 Shared environment vs genes in familial clustering

Although familial clustering may also be the consequence of shared environment, various types of epidemiological studies argue against the importance of this factor in familial clustering in MS, unlike in regional clustering. 

Twin studies

Twin studies are a classical method to discriminate between the contributions of genes and environment in the etiology of diseases. Monozygotic twins are genetically almost identical, whereas dizygotic twins share on average 50% of their DNA, like sibs that are not twins. The 50% of DNA that dizygotic twins share on average is due to the fact that for 25% of the genome they share both chromosomes (100%), for 50% of the genome they share one chromosome (50%) and for 25% of the genome they share no chromosomes (0%).

Discordance for a trait between monozygotic twins is thus mainly attributed to environmental and not to genetic factors. In dizygotic twins, both genetic and environmental factors may be responsible for discordance. A possible source of bias in twin studies is excess inclusion of concordant pairs compared to discordant pairs. This type of bias, however, should not be different in dizygotic and monozygotic twins.

For MS, several twin studies have shown a consistent difference in concordance between monozygotic and dizygotic twins. Concordance in monozygotic twins is around 30%, whereas in dizygotic twins it is 3-5%. However, interpreting these numbers one has to realize that sometimes the healthy twin of a discordant pair of identical twins does have neurological symptoms, oligoclonal bands in the CSF or MRI abnormalities without ever meeting the criteria of MS. Another limitation of twin studies is that the sample will necessarily be relatively small.

An interesting observation from twin studies is, that even if two individuals are more or less genetically identical, still in 70-75% of the cases they are discordant for MS. Postzygotic events during formation of the embryo and placenta and epigenetic effects later in life are likely to play a role, also influenced by environmental factors. Differences in mitochondrial DNA between members of a twin pair are another possible influence.
Studies of half-sibs, adoptees and conjugal pairs

Ebers et al. found an age-adjusted recurrence risk for full-siblings of a patient of 3.11%, whereas for half-siblings in the same families it was 1.89%, indicating the effect of the amount of sharing of DNA\(^4\). There was a difference between paternal (1.31%) and maternal (2.35%) half-sibs, suggestive of a maternal effect. This may be due to various factors, e.g. environmental (pre-and postnatal), genetic (mitochondrial) or epigenetic (parental imprinting). There was no difference between the risk of half-sibs who did and did not live with the MS patient. This observation supports the hypothesis that the familial aggregation of MS is mainly caused by genetic, rather than environmental factors.

Non-biological first degree relatives of adopted patients have a risk comparable to the population risk\(^3\). Biological relatives that do not live with the patient, however, have the same risk as relatives from intact nuclear families, despite the ascertainment bias. This again indicates that genetic factors are more important than shared environment in familial aggregation of MS.

Children of conjugal pairs both with MS have a recurrence risk of MS of 1:17, age-adjusted 1:5, compared to a risk of 1:200, age-adjusted 1:50, for children of single affected parents. To assess the influence of environmental factors, conjugal pairs were compared. There was no evidence for clinical concordance, clustering at year of onset or distortion of expected pattern of age at onset in the second affected spouse\(^4\). The Canadian data also showed that the rate of MS among spouses of patients was not increased compared with the population risk, arguing against transmission\(^4\).

2.1.3 Effect of consanguinity

In several of the clusters described, consanguinity appears to play a role\(^4\). If a patients unaffected parents are first cousins, the recurrence risk for his/her siblings is nearly four times increased compared with the situation when the parents were unrelated\(^4\). This effect has been related to a recessive inheritance mode of susceptibility alleles.

2.1.4 Subclinical MS?

In epidemiological studies, most often only patients with definite or probable MS are considered affected. However, if clinically unaffected relatives of patients are examined by means of laboratory techniques, the number of
relatives showing abnormalities considered characteristic for MS is higher. MRI studies in relatives of MS patients show lesions, compatible with MS, in part of the clinically unaffected relatives. Likewise, examining healthy sibs in multiplex families Duquette et al. reported oligoclonal bands in 18% of sibs. Xu et al. found oligoclonal bands in the CSF of 13 of 15 clinically unaffected twins of MS patients, 3 of whom developed clinical manifestations of MS in the 6 year follow-up period. Haghighi et al. found two or more oligoclonal bands in 9/47 healthy sibs of MS-patients as opposed to 2/50 controls. Nuwer et al. have shown abnormal evoked potentials in up to 35% of healthy sibs of patients from multiplex families. MS lesions have also been found at autopsy in asymptomatic individuals.

There may be various explanations for these findings. The sibs that showed abnormal test results may have a subclinical type of MS that may or may not become clinically manifest at follow up (at this point, data are limited). It is possible that they carry the same "genetic burden" their affected sibs are carrying, but there is reduced penetrance. Differences in environmental factors or age at which sibs are exposed to these factors may play a role. If one assumes polygenic inheritance and/or interaction between genes contributing to MS, a sib could be more or less strongly predisposed to MS if he or she carries more or fewer disease-related alleles. Another explanation for the presence of subclinically affected sibs is, that the disease has a very mild course and the lesions are localized in clinically 'silent' areas of the brain. For the spinal cord, however, this might be harder to argue.

Differences in frequencies of MS in various ethnic or racial groups have been described. In epidemiology, these differences are often considered preliminary evidence of genetic contribution to the disease, since it is difficult to establish that the environments of the groups are the same.
2.2 Segregation/ Model of Disease

For many Mendelian diseases, underlying mutations in specific genes have been found. Over all, mutational diversity at each of these loci is high, each mutation is rare, having occurred in recent human history (no older than 2000 years) and each mutation is sufficient and necessary to cause the phenotype of interest. However, most human phenotypes and diseases are complex and Mendelian patterns do not apply.

The segregation of MS in multiplex families is not consistent with a fully penetrant single gene disorder. The prevalence in relatives of MS patients decreases with increasing genetic distance in a non-linear fashion, also pleading against single locus Mendelian inheritance. Epidemiologic characteristics suggest that MS is a complex disease. Complex diseases show familial aggregation but a high proportion of sporadic cases, no Mendelian pattern of inheritance, reduced penetrance, phenocopies, etiologic and genetic (locus and allelic) heterogeneity.

It is suspected that mutations leading to a complex phenotype occur at multiple genes. The genetic model is often represented as a curve where affected individuals are those that cross a biological threshold of risk. The contributing alleles at multiple independent loci may either be rare or common, dependent upon the model applied.

![Diagram: Competing theories for complex disease inheritance](From Chakravarti 60)

Figure 1

Competing theories for complex disease inheritance (a, b, c and so on are susceptibility/protective alleles at multiple, independent loci). For a fixed disease incidence, individuals who are clinically affected can either have mutations at only one of many possible disease loci (in which case the mutant alleles are rare in the population) or harbour mutations at multiple loci simultaneously (in which case the mutant alleles are common in the population). These hypotheses are the extremes of many other possible intermediate scenarios. (From Chakravarti 60)
For MS, many models have been proposed and tested\textsuperscript{61}. Parametric methods incorporate the parameters known or suspected to influence susceptibility and its segregation. Calculations are made, varying values and interactions of these parameters. The results are compared with the observed data and the best fitting model is proposed. Sometimes, in order to make calculations feasible, assumptions may be made that are not realistic\textsuperscript{62}.

Parametric methods specify a model for the disease in question, but the validity of this model is not always established. Therefore, nonparametric methods are often favoured\textsuperscript{63,64}. Nevertheless, (parametric) likelihood methods could in principle make better use of the data than non-parametric methods could, by interpreting it against the backdrop of evolutionary history that links the observed haplotypes\textsuperscript{65}. Non-parametric methods sometimes try to summarize information of all haplotypes in one value. Since similar values may result from different underlying evolutionary models in cases compared with controls, this might result in less power\textsuperscript{65}. Another limitation of non-parametric methods, especially when studying complex disorders, is that including covariates like other genetic or environmental factors is difficult\textsuperscript{66}.

The best fitting model is likely to be a multifactorial model with a number of genes of which one major gene, situated in the HLA-DRDQ region and one or more environmental factors. The MHC is estimated to account for 10-60\% of the genetic component of MS susceptibility in Caucasian populations of northern European descent\textsuperscript{67,68}. We have found indications that the inheritance mode of the HLA-related gene (or genes) is stronger than additive suggesting a recessive component\textsuperscript{69}.
2.3 Methods of genetic analysis

In complex diseases, frequent variants of a number of genes, each with a modest effect may play a role. Phenocopies, genetic (locus and allelic) heterogeneity, reduced penetrance and contribution of environmental factors complicate the finding of susceptibility genes even more. With the sequencing of the human genome, new possibilities to learn about the genetic background of complex diseases have emerged. Millions of single nucleotide polymorphisms (SNPs) provide potential markers. However, it is as important to develop strategies to distinguish between polymorphisms with relevance for susceptibility to or influencing the course of a disease and those that are only in linkage disequilibrium with them. For example, associations between HLA-types and a number of diseases are known since decades, but identification of the causal variants has proven very difficult.

2.3.1 Linkage analysis

Linkage analysis assesses the segregation of a genetic variant in families with multiple affected members. Classical linkage analysis within families will typically resolve the position of a novel gene to 10-20 cM, with further precise location obtained by using linkage disequilibrium mapping within this region. In complex diseases, linkage analysis has only had limited success. The majority of positive linkages for the same disease could not be replicated. No single study design consistently produced more significant results. The only variables independently associated with increased study success were increase in the number of individuals studied and sampling from only one ethnic group. Kruglyak and Lander have shown that for loci with a modest RR the sample size needed for high resolution linkage mapping is very large. The power of a linkage genome screen is dependent upon the frequency of susceptibility alleles in the population studied and may thus vary between populations.

In multiple sclerosis, linkage screens have so far had limited results. Although every genome screen identified regions of interest, none has demonstrated linkage with genomewide significance. Meta-analysis of the available raw data of 11 whole genome screens for linkage, using microsatellite markers with a density of about one marker every 10cM, included over 700 multiplex families. However, only in the MHC region linkage with genomewide significance was found. The apparent technical limitations were likely to have reduced the power of the analyses. This stimulated the
International Multiple Sclerosis Genetics Consortium to conduct a linkage screen in 730 multiplex families (comprising 2692 individuals) using a set of almost 6000 markers.

2.3.2 Association analysis

Association studies test whether a genetic marker (polymorphism or haplotype) occurs more frequently in cases than in controls. If significant association emerges and the possible bias of population stratification is excluded, the polymorphism itself is either in the susceptibility locus or in linkage disequilibrium with the susceptibility locus. Careful application of the method is warranted in order to avoid several confounding factors such as variable definitions of the phenotypes, the aforementioned population stratification or methodological flaws such as insufficient numbers. In order to avoid false positive results caused by population stratification, family-based controls can be used, for example trios consisting of a patient and both parents. The non-transmitted haplotypes of the parents can be used as (pseudo-)controls.

In monogenic diseases, single-locus association methods have proven to be powerful tools in mapping disease loci and identification of genes. However, in complex diseases single-locus methods have so far had limited success. In 1997, genome-wide association screening using pooled DNA and 6000 microsatellite markers was thought to be an efficient strategy to find predisposing genes outside the HLA-region. For MS, a number of studies did not show significant association (see Table 2). DNA was pooled to reduce the number of assays. This pooling of DNA appeared to be complicated by specific artifacts. In pooled DNA, there is a tendency to overestimate the frequency of short alleles, partly as a consequence of stutter bands that may be hard to distinguish from actual peaks. The use of pooled DNA is also hampered by length-dependent amplification, distorting allele frequencies.

If indeed a complex disease is caused by a number of genes, and for any of these genes by frequent alleles with each a small effect, very large sample sizes and much denser marker sets will be necessary to find these genes by means of association analysis. However, with the current availability of many SNPs and more efficient genotyping methods, mapping and identification of disease alleles contributing to complex diseases using association analysis may come within reach.
Linkage disequilibrium (LD)-based methods use multilocus association analysis. These methods are based on the assumption that affected individuals in the present generation have inherited their susceptibility alleles from common ancestors. A disease-related mutation has been introduced on a haplotype in a previous generation. With each generation, recombination events during meiosis may have led to reduction of the length of the original haplotype. Therefore, only if the distance between a marker locus and a disease locus is small, they are expected to be in LD. The strength of LD is influenced by a number of other factors such as number and age of disease mutations and population admixture.

Association-based studies using LD mapping are more sensitive to find also minor genes. However, also with multilocus analysis, a sufficient density of markers is necessary. The marker density needed depends on the average extent of LD and the number of haplotype blocks. Kruglyak estimated on the basis of simulation studies that a useful level of LD is unlikely to extend beyond an average distance of about 3 kb in the general population. This implies that approximately 500,000 to 1,500,000 SNPs will be required for whole genome studies. The use of microsatellite markers (that are multiallelic and thus potentially more informative) will probably be too inefficient and expensive in association studies. According to Kruglyak, the extent of LD is similar in isolated populations, unless the founding bottleneck is very narrow or the frequency of the variant is low. Other authors indicate that LD may be spanning megabases in special populations. In that case, much lower marker densities could generate substantial power. Issues concerning LD and haplotype blocks are reviewed by Wall and Pritchard.

Recently a special type of association analysis, admixture mapping, has been used to identify genes contributing to MS. Admixture mapping tries to identify genomic regions where individuals with MS from admixed populations tend to have an unusually high proportion of ancestry from either one or the other population of origin. This may be an indication for the presence of a multiple sclerosis risk variant that differs in frequency between the ancestral populations. Admixture mapping has statistical power to detect factors that differ markedly in frequency across populations. An advantage is, that for instance in the African American population there has been contact between the ancestral populations for on average six generations. Thus, few recombinations have taken place and long conserved segments of ancestral chromosomes allow less dense marker sets. Reich et al. reported strong association between the extent of European ancestry and MS around the centromere of chromosome 1. However, they did not find association with HLA. This is the consequence of a limitation of admixture...
mapping: it cannot detect disease loci at which the total risk summed over all alleles in each population is similar in both populations.

In order to confirm that a true disease-related SNP has been identified, replication of the association in another, independent sample is required. This is more likely to be successful using multi-locus (haplotypic) association than single-locus association methods. The probability of a unique haplotype being associated with a disease will be high, since frequencies of haplotypes are much lower than those of alleles.

Another issue concerns correction for multiple testing using correction methods such as Bonferroni’s or Holm’s. These methods are often conservative with as a consequence a reduction in power. Different statistic methods have been developed reducing the number of comparisons. The choice of a lodscore of 3 implies some correction for multiple testing in monogenic disease where a single gene causing the disease is likely. However, the appropriate multiple testing correction should depend on the number of independent tests. When the tests of adjacent markers are strongly correlated, a lower correction factor is required than when there is less dependency between tests, as in single locus association analysis. The degree of LD is an indication for the correlation between tests at consecutive loci in multilocus methods. Apart from multiple testing correction, the a priori probability of a gene with an effect that can be found with the current sample size is unknown, which leads to a situation where only replicated results can be trusted.

2.3.3 Non-parametric methods for gene mapping

Affected sibpair and pedigree member studies compare genotypes of two or more affected individuals from a pedigree in order to find genomic areas which are shared in excess. The problem of non-penetrance is avoided by this approach, but the problems of phenocopies and heterogeneity are not. In complex diseases, also the relatively low number of multiplex families often reduces the power of these methods. This is also the case in a number of studies in MS were these methods were applied.

The Transmission/disequilibrium test evaluates deviation from linkage equilibrium from a heterozygous parent to his/her affected child. Linkage equilibrium would imply that both alleles have equal probability to be transmitted (both 50%). If there is deviation from linkage equilibrium and one allele is transmitted preferentially to affected offspring, it is likely to be
associated with the disease in question. However, the TDT only shows significant association when the marker is linked to the disease locus. When applied to more complex diseases this method failed to identify relevant gene loci. Apart from the fact that most studies into complex genetic disease have too limited sample size, this can be attributed to genetic effects for which the TDT has reduced power, e.g. effects of frequent alleles of specific genes. This may be the case in MS as well (see also Table 2), but studies with larger numbers of patients show interesting results concerning the HLA region.

The Haplotype Sharing Statistic

Analysis of haplotype overlap can be used to map a disease gene. Once it is certain that haplotypes contain a disease gene, the smallest fragment shared by patients and not by controls is likely to contain this gene. However, when it is not certain that haplotypes contain a disease gene, the overlap of haplotypes needs to be evaluated statistically. Te Meerman and Van der Meulen developed the Haplotype Sharing Statistic (HSS), a non-parametric multilocus association method for finemapping disease genes. They described the perspectives of identity by descent (IBD) mapping in founder populations and showed the method to be efficient due to the large number of meioses implicitly observed and reduced heterogeneity in these populations. However, the applicability of the method is not limited to founder populations in the narrow sense, as with the current ultra-high marker density many populations show LD in sets of many nearby markers. The method can handle the situation of multiple introductions of a particular disease mutant and thus different surrounding haplotypes. Nolte et al. developed the method further.

HSS compares the length of shared haplotypes among patients with the length of shared haplotypes among controls. The length of haplotype sharing is taken to be an approximate measure for the probability of two haplotypes being identical by descent (IBD). The probability of IBD increases with increasing length of haplotype sharing, although not linearly. The hypothesis is, that haplotype sharing among patients is larger than among controls at loci involved in the disease, because (i) haplotypes containing risk alleles are likely to be similar more often especially close to the risk allele and (ii) haplotypes containing the risk alleles may be shared over longer stretches. The first factor is understandable with the concepts of association and LD. The second factor is the consequence of the patient haplotypes.
containing mutations that are genetically younger than the wild type allele. Thus, fewer meioses and consequently fewer recombinations have taken place on the patient haplotypes containing a disease mutation than on the wild type haplotype.

Figure 2: The sharing of two haplotypes

<table>
<thead>
<tr>
<th>loci</th>
<th>haplo 1</th>
<th>haplo 2</th>
<th>IBS</th>
<th>haplotype sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>loci 1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>loci 2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>loci 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>loci 4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>loci 5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>loci 6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>loci 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>loci 8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The sharing of two haplotypes at a locus is defined as the number of intervals between successive marker loci at which alleles of these haplotypes are identical (Figure 2). In order to know the phase of the haplotypes trios are used, preferably consisting of an affected individual and his or her parents. The non-transmitted haplotypes of the parents are used as controls. If parents are not available, a spouse and child can be used. In that case, the haplotypes of the spouse serve as controls. However, there is a risk of population stratification that is not present when non-transmitted haplotypes are used. The sharing of all pairs of haplotypes is calculated for each marker locus for the patient and control datasets separately and the results of the datasets are compared.

HSS is more powerful and more accurate in fine-mapping than association analysis and the TDT for high frequency risk alleles that are considerably younger than the wildtype alleles. If, however, the risk alleles and the
wildtype alleles are of the same age, HSS has no power but association analysis and TDT have. Nolte et al. demonstrated that HSS extracts different information from the data than do association analysis and TDT \(^{118}\). As Jorde \(^{120}\) concluded, it is unlikely that a single method would provide optimal power under all circumstances to detect susceptibility genes for complex diseases. Moreover, most often the underlying genetic model is not known. Therefore, it is recommendable to apply different methods and compare the results or even combine p-values after correction for correlation between statistics.

Haplotype sharing analysis was applied in MS on a genomewide \(^{121}\) and candidate region \(^{109,122-124}\) basis.

### 2.3.5 Candidate genes

Association studies into candidate genes that show positive results will only rarely be confirmed by a second investigation. This is caused by a number of factors, including too small sample sizes and publication bias \(^{125}\). Ioannidis even demonstrated that most published research findings are false \(^{95}\). This is the consequence of bias and lack of study power, but also the publication selection effect of many other studies on the same question and the ratio of “true relationships” to “no relationships” among those tested in the field, i.e. the prior probability of the research finding being true.

A large number of genes have been investigated for a possible role in susceptibility to MS (ref. www.ucsf.edu/msdb/r_ms_candidate_genes.html). The only association that has repeatedly been confirmed is the association with alleles of the human leukocyte antigen (HLA) system. In 1972, an association between alleles of the HLA system and MS was reported by Jersild et al.\(^{126}\). They demonstrated association with HLA class I antigens A3 and B7. This association provided the first direct evidence for a genetic contribution in the susceptibility to MS. Later, the association with HLA class II antigens appeared to be stronger than that with class I antigens. It has been confirmed by many investigators\(^{127-129}\). In Northern European populations, the associated HLA-type is DR2, Dqw1, DNA-subtyping DRB1*1501, DQA1*0102, DQB1*0602\(^{130-133}\). Differences in HLA-associations have been reported in different populations: DR15 and DR4 in Canary Islands\(^{134}\), DR4 and DR15 in Turkey \(^{135}\), DR4 and DR3 in Sardinia \(^{136}\). The association may also be related to the type of MS (Western or Asian in Japan \(^{137}\)). It has been suggested that some haplotypes may be protective \(^{114}\).

Conserved haplotypes with recombination hotspots in between characterize the HLA region. This results in regions with strong linkage dis-
equilibrium and reduced resolution to discriminate between genes on an ancestral haplotype. In most populations DRB1*1501 and DQB1*0602 are almost invariably linked, so it is still not clear whether DR or DQ or both are implied in susceptibility to MS. In our haplotype sharing analysis, the interval showing strongest haplotype sharing contained DQB1. The importance of DQB1-alleles was supported by studies in populations were different combinations of DR2 and DQ6 were present, whereas others on the contrary found that DRB1 was more important than DQB1.

The contribution of the extended haplotype HLA DRB1*1501, DQA1*0102, DQB1*0602 to the susceptibility to MS was estimated 14% by Ligers et al. There is evidence for the involvement of different non-HLA genes in HLA-DR15 positive compared with HLA-DR15 negative patients. An important question is, whether the association between HLA and MS is due to a causative role of HLA-molecules in the susceptibility to MS, or whether a separate gene, not functionally involved in the HLA-system, plays a role. The HLA-system plays an important role in the immune system. Among other functions, it is involved in antigen-presentation by macrophages to lymphocytes, after which the specific immunological reaction to this antigen can start. A preference of HLA-genes to present specific auto-antigens could make them indeed functionally involved in MS. Specific HLA-types may provide a selective evolutionary advantage from the point of view of infectious disease despite the fact that they make the individual more susceptible to an auto-immune disease. Another possibility is that a gene that is genetically close to the HLA-genes could "hitch hike" with such a haplotype, even though on itself, it might give a selective evolutionary disadvantage. This way, also genes that contain a mutation that makes the individual susceptible to a disease could reach higher population frequencies than expected on the basis of their selective disadvantage.

2.3.6 Whole genome screens

In 1996, three groups performed genome screens in multiplex families collected in Canada, the UK and the USA. They used (partly different) sets of polymorphic markers that spanned the whole genome. A multiple stage screen was performed in all three samples; the areas identified in a first screen were analyzed in a secondary screen by the Canadian and UK groups. An important and maybe surprising finding was, that there were no regions of interest that appeared strongly in all three initial screens. Overall, the contribution of all regions of interest in the initial screen appeared to be small.
in the combined datasets. Around 90% of the genome was excluded for a locus with rare alleles that contributes a sibling recurrence risk $\lambda_s > 3$.

However, two regions stood out in these studies. The HLA region on chromosome 6p21 appears as the region with strongest linkage in the USA study (peak MLS 3.57) and the UK study (peak MLS 2.8), though in the latter study becoming apparent only in the secondary screen. In the Canadian study, the MLS is 0.65; evidence for linkage disequilibrium was found for one of three markers on chromosome 6, but not for the other two markers which were closer to the HLA DR and DQ loci. The second region of interest was located on the short arm of chromosome 5.

A genomewide two-stage scan was done in multiplex families from an isolated Finnish region with high prevalence of MS. The population originates from a limited number of founders. This population structure was considered very useful in the study of complex disease, because of the likely restricted number of founder mutations. The first stage comprised multi-point, non-parametric linkage analysis of a low-resolution screen of 328 polymorphic markers. This did not reveal statistically significant linkage. There were however 10 regions of slight interest ($p=0.1-0.15$), of which 8 were analyzed in the second stage. In this stage, a denser marker screen was applied with average spacing of 4.5 cm (0.1-10cM). 5 multiplex families were added in this study. The results were analyzed with two-point parametric linkage analysis. Finally, the highest two-point LOD scores were found on chromosome 17 (Zmax=2.8, theta=0.04 under a dominant model). The UK genome screen showed evidence for linkage on 17q22-q24 with the same markers. Allelic association with markers on 17q22-q24 was found in neither of the two studies. Since these first whole genome screens several genome screens have been conducted with increasing insight into the requirements of genetic studies in complex diseases (Table 2). Loci with LOD scores < 2 are not shown.
Conclusions on genome screens

So far, only the HLA region has repeatedly shown significant association and linkage with MS. This is probably the consequence of sample sizes and marker densities having been still insufficient with regard to the complexity of the underlying genetic model. Genetic (locus and allelic) heterogeneity, reduced penetrance and contribution of environmental factors complicate the finding of susceptibility genes. It is to be expected that in the near future possibilities to find these genes will be much better with the availability of more of SNP markers and high throughput genotyping. However, funding will have to be sufficient as well, since large numbers of DNA samples will have to be examined. In order to obtain these large numbers of DNA samples, national and international cooperation will be required.
Table 2: Whole genome screens in Multiple Sclerosis

<table>
<thead>
<tr>
<th>Study, population</th>
<th>Patients</th>
<th>Markers</th>
<th>Average separation</th>
<th>Statistical methods</th>
<th>Loci (MLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sawcer</em> 1996 UK</td>
<td>Stage 1: 129 multiplex families, stage 2: 98 multiplex families, total 251 affected sibpairs</td>
<td>Stage 1: 311 microsatellite-markers; stage 2: 44 markers</td>
<td>12 cM</td>
<td>Linkage analysis (MAPMAKER/SIBS) maximum likelihood sharing probabilities for each sibpair, MLS</td>
<td>Stage 1: 19 loci with MLS &gt; 0.7 (nominal significance 5%); 6 loci with MLS &gt; 1.8; stage 2: 17q22 (2.7), 6p21 (2.8)</td>
</tr>
<tr>
<td><em>Haines</em> 1996 US</td>
<td>Stage 1: 52 multiplex families, stage 2: 23 multiplex families, total 126 affected sibpairs and 88 other affected relative pairs</td>
<td></td>
<td>9.6 cM</td>
<td>Linkage (FASTLINK), sibpair (SAGE/SIBPAL), affected relative pair (SimIBD)</td>
<td>(significant on three tests) 7q21-22 (2.86)</td>
</tr>
<tr>
<td><em>Ebers</em> 1996 Canada</td>
<td>Stage 1: 61 multiplex families, 100 affected sibpairs, stage 2: 444+78 sibpairs</td>
<td>Stage 1: 257 microsatellite markers, stage 2: 5 microsatellite markers on 5p and 3 on 6p21 in all 3 datasets</td>
<td>15.2 cM (stage 1)</td>
<td>Multipoint linkage analysis and multipoint sibpair analysis, TDT</td>
<td>Stage 1: 5 loci with MLS &gt; 1, of which D5S406 (4.24), Stage 2: 5p, 6p21 non-significant, D6S461 on TDT $\chi^2$ 10.8 (p&lt;0.01) and 10.9 (p&lt;0.0005)</td>
</tr>
<tr>
<td><em>Kuokkanen</em> 1997 Finland</td>
<td>Stage 1: 16 multiplex families, stage 2: 21 multiplex families</td>
<td>Stage 1: 328 microsatellite markers</td>
<td>4.5 cM</td>
<td>Stage 1: Multipoint non-parametric linkage analysis, stage 2: two-point parametric linkage analysis</td>
<td>Stage 1: no significant linkage, stage 2: 17q22-24 (2.8)</td>
</tr>
<tr>
<td>Study, population</td>
<td>Patients</td>
<td>Markers</td>
<td>Average separation</td>
<td>Statistical methods</td>
<td>Loci (MLS)</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Broadley 2001, Italy</td>
<td>40 multiplex families with 37 sibpairs and three other relative pairs</td>
<td>321 microsatellite markers and HLA DRB1</td>
<td>Multipoint non-parametric linkage analysis (GENEHUNTER-PLUS), TDT</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Coraddu 2001, Sardinia</td>
<td>49 multiplex families, 46 sibpairs and 3 sibtrios</td>
<td>324 microsatellite markers, HLA DRB1, DQA1 and DQB1</td>
<td>Nonparametric linkage analysis (MAPMAKER/SIBS) and TDT</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Dyment 2001, Canada</td>
<td>219 sibpairs</td>
<td>105 markers previously showing increased sharing</td>
<td>Multipoint linkage analysis and transmission disequilibrium (Aspex Statistical Package)</td>
<td>5p14 (2.27), transmission disequilibrium D17S789 (p=0.0015)</td>
<td></td>
</tr>
<tr>
<td>The Transatlantic Multiple Sclerosis Genetics Cooperative 2001, US, UK, Canada</td>
<td>US: 52 multiplex families, 133 affecteds, UK: 128 families, 264 affecteds, Canada: 61 families, 139 affecteds</td>
<td>US: 442, UK: 314, Canada: 257 microsatellite markers with different overlap between screens</td>
<td>Meta-analysis, non-parametric multipoint linkage (GENEHUNTER)</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Akesson 2002, Scandinavia</td>
<td>136 sibpairs</td>
<td>399 microsatellite markers</td>
<td>Multipoint nonparametric linkage analysis (MAPMAKER/SIBS)</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Ban 2002, Australia</td>
<td>54 sibpairs</td>
<td>397 microsatellite markers</td>
<td>Multipoint nonparametric linkage analysis (MAPMAKER/SIBS)</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Study, population</td>
<td>Patients</td>
<td>Markers</td>
<td>Average separation</td>
<td>Statistical methods</td>
<td>Loci (MLS)</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Haines 2002 US</td>
<td>1: 52 multiplex families with 135 affecteds, 2: 46 multiplex families with 131 affecteds</td>
<td>80 microsatellite markers in 19 previously determined regions</td>
<td>Parametric and nonparametric linkage analysis (FASTLINK, HOMOG, ASPEX, ALLEGRO), stratification for HLA DR2</td>
<td>5 regions with LOD score &gt;2.0; after stratification for DR2 in 2 additional regions</td>
<td></td>
</tr>
<tr>
<td>He 2002 Sweden</td>
<td>Stage 1: 5 MS patients from 4 families, stage 2: 15 MS patients and healthy relatives (genetically isolated population)</td>
<td>Stage 1: 390 (380) microsatellite markers</td>
<td>TDT on haplotypes</td>
<td>Stage 1: 7 fragments shared by &gt;=4/8 patient chromosomes; stage 2: conserved haplotype of 10cM on 17p11 in 12/15 patients, 4 marker haplotype 6/15, p&lt;0.01</td>
<td></td>
</tr>
<tr>
<td>Sawcer 2002 UK</td>
<td>Pooled DNA of 1: 216 patients vs 219 controls and 2: 745 patients vs 1490 parents</td>
<td>6000 microsatellite markers</td>
<td>Association analysis, Chi-square test</td>
<td>10 most promising markers p-value &lt;5%</td>
<td></td>
</tr>
<tr>
<td>Coraddu 2002 Sardinia</td>
<td>Pooled DNA of 1: 229 MS patients vs 264 controls and 2: 235 trios of patient vs parents; all patients had RR MS</td>
<td>2764 microsatellite markers</td>
<td>Association analysis, Chi-square test</td>
<td>5 markers with p-value &lt;5% in both screens</td>
<td></td>
</tr>
</tbody>
</table>

**Genetic epidemiology of MS**
<table>
<thead>
<tr>
<th>Study, population</th>
<th>Patients</th>
<th>Markers</th>
<th>Average separation</th>
<th>Statistical methods</th>
<th>Loci (MLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goedde 156 2002</td>
<td>Pooled DNA of 198 HLA-DR15+ MS patients and 198 controls</td>
<td>6000 (4666) microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>87 markers with nominal p-value &lt; 0.05 (not corrected for multiple testing)</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ban 157 2003</td>
<td>Pooled DNA of 217 HLA DR15 positive MS patients and 187 controls</td>
<td>6000 (4346) microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>7 markers with p-value &lt; 1%</td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bielecki 158 2003</td>
<td>Pooled DNA of 200 MS patients vs 200 controls and 129 trios</td>
<td>6000 (4696) microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>5 markers with p-value &lt; 5%</td>
</tr>
<tr>
<td>Poland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eraksoy 159 2003</td>
<td>43 multiplex families: 16 with &gt;=2 sibs, 27 other relative pairs (consanguinity in 13 families). Stage 1: 92 affecteds, stage 2: 78 unaffecteds (45 markers)</td>
<td>392 microsatellite markers</td>
<td>10 cM</td>
<td>Linkage analysis (GENEHUNTER-PLUS),</td>
<td>No regions with genomewide significance</td>
</tr>
<tr>
<td>Turkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eraksoy 160 2003</td>
<td>Pooled DNA of 197 patients with RR/SP MS vs 199 controls</td>
<td>6000 (4359) microsatellite markers</td>
<td></td>
<td>Association analysis, Chi-square test</td>
<td>12 regions with empirical p-value &lt; 5%; 5p15 confirmed by linkage</td>
</tr>
<tr>
<td>Turkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, population</td>
<td>Patients</td>
<td>Markers</td>
<td>Average separation</td>
<td>Statistical methods</td>
<td>Loci (MLS)</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Giedraitis 2003</td>
<td>54 MS patients and 114 healthy family members (genetically isolated population)</td>
<td>1040 (834) microsatellite markers</td>
<td>4.4 cM</td>
<td>Marker- and haplotype-based (TRANSMIT) TDT, nonparametric linkage analysis (Genehunter)</td>
<td>7 regions in transmission disequilibrium with MS, 1 strongly (14q24-32); lodscores around 2 for several regions; both on 17q12-24</td>
</tr>
<tr>
<td>Goris 2003</td>
<td>Pooled DNA of 204 MS patients vs 198 controls and 131 trios</td>
<td>6000 (4875) microsatellite markers</td>
<td>Association analysis, Chi-square test</td>
<td>27 markers p-value&lt;5%</td>
<td></td>
</tr>
<tr>
<td>Heggarty 2003</td>
<td>Pooled DNA of 200 MS patients vs 200 controls</td>
<td>6000 (2537) microsatellite markers</td>
<td>Association analysis, Chi-square test</td>
<td>22 markers p-value&lt;5%, 5 markers p-value&lt;1% (of which 2 in HLA region)</td>
<td></td>
</tr>
<tr>
<td>Hensiek 2003</td>
<td>Previous 129 multiplex families + 97 multiplex families, analyzed together and stratified on HLA-DRB1*1501</td>
<td>242 microsatellite markers (+111 previous)</td>
<td>Multipoint nonparametric linkage analysis (MAPMAKER/SIBS)</td>
<td>No regions with genomewide significance</td>
<td></td>
</tr>
<tr>
<td>Laaksonen 2003</td>
<td>Pooled DNA of 195 patients vs 205 controls</td>
<td>5532 (5522) microsatellite markers</td>
<td>Association analysis, Chi-square test</td>
<td>108 markers with hypothetical p-value&lt;5%, 5 genomic regions with markers &lt;1 cM apart</td>
<td></td>
</tr>
<tr>
<td>Study, population</td>
<td>Patients</td>
<td>Markers</td>
<td>Average separation</td>
<td>Statistical methods</td>
<td>Loci (MLS)</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Liguori\textsuperscript{165} (2003) Italy</td>
<td>Pooled DNA of 224 MS patients vs 231 controls and 185 trios</td>
<td>6000 (4789) microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>D2S367 (p-value 0.015)</td>
</tr>
<tr>
<td>Santos\textsuperscript{166} (2003) Portugal</td>
<td>Pooled DNA 188 MS patients vs 188 controls</td>
<td>6000 (4661) microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>10 markers empirical p-value &lt;0.01</td>
</tr>
<tr>
<td>Weber\textsuperscript{167} (2003) Germany</td>
<td>Pooled DNA of 234 MS patients vs 209 controls and 68 trios</td>
<td>6000 microsatellite markers</td>
<td>0.75 Mb</td>
<td>Association analysis, Chi-square test</td>
<td>11 markers with empirical p-value &lt;5%</td>
</tr>
<tr>
<td>GAMES and Transatlantic Multiple Sclerosis Genetics Cooperative\textsuperscript{16} (2003)</td>
<td>&gt;700 multiplex families</td>
<td>Around 6000 microsatellite markers</td>
<td>0.75 Mb</td>
<td>Meta-analysis of raw genotypes, linkage analysis</td>
<td>Genomewide significance in the MHC region, not outside it. Limited accuracy of data</td>
</tr>
<tr>
<td>Kenealy\textsuperscript{168} (2004) US and France</td>
<td>456 affected relative pairs from 245 families</td>
<td>390 microsatellite markers</td>
<td>&lt;10cM</td>
<td>Parametric and non-parametric analysis of linkage (FASTLINK and HOMOG), “score pairs” and exponential model (Allegro)</td>
<td>No regions with genomewide significance</td>
</tr>
<tr>
<td>Dyment\textsuperscript{169} (2004) Canadian</td>
<td>522 sibpairs from 442 families (stage 1: 219, stage 2: 333 sibpairs)</td>
<td>Stage 1: 498 and stage 2: 35 microsatellite markers</td>
<td>7 cM</td>
<td>Multipoint linkage analysis and transmission distortion (Aspex Statistical Package)</td>
<td>Only linkage on 6p (4.40) Suggestive 2q27 (2.27) and 5p15 (2.09)</td>
</tr>
<tr>
<td>Study, population</td>
<td>Patients</td>
<td>Markers</td>
<td>Average separation</td>
<td>Statistical methods</td>
<td>Loci (MLS)</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Goedde et al. 2005 Germany</td>
<td>Pop1 and pop2: 100 MS patients (85% HLA DR2+) vs 100 unstratified controls; pop3 180 MS patients vs 180 controls, both unstratified</td>
<td>11,555 SNPs on DNA chips</td>
<td>105 kb (mean genetic distance 0.32 cM, gaps on chr. 16, 17 and 19)</td>
<td>Association analysis, exact Fisher test for two-by-three tables; Bonferroni correction for multiple testing</td>
<td>SNPs in the vicinity of HLA-DRA show strong association (p-value $4 \times 10^{-10}$) (extended haplotype?); no other significant association after Bonferroni correction</td>
</tr>
<tr>
<td>International Multiple Sclerosis Genetics Consortium 2005 Northern European</td>
<td>2692 individuals from 730 multiplex families: 1595 affecteds of whom 830 sib pairs and 172 other relative pairs</td>
<td>5858 (4506) SNPs</td>
<td>&lt;500 kb</td>
<td>Multipoint non-parametric linkage analysis</td>
<td>6p21 (11.66), 17q23 (2.45), 5q33 (2.18)</td>
</tr>
<tr>
<td>Reich et al. 2005 African Americans</td>
<td>605 patients and 1043 controls</td>
<td>1555</td>
<td></td>
<td>Association analysis/Admixture mapping</td>
<td>Centromere 1 (5.2)</td>
</tr>
</tbody>
</table>
References


117. van der Meulen MA, te Meerman GJ. Haplotype sharing analysis in affected individuals from nuclear families with at least one affected offspring. Genet Epidemiol 1997; 14(6):915-920.


