Utilizing Twins Concordance Rates to Infer the Predisposition to Myasthenia Gravis

Ryan Ramanujam,1 Ritva Pirskanen,2 S. Ramanujam3 and Lennart Hammarström1

1 Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet at Karolinska University Hospital Huddinge, Stockholm, Sweden
2 Department of Neurology, Karolinska University Hospital Solna, Stockholm, Sweden
3 The Peter J. Tobin College of Business, St. John’s University, New York, United States of America

Myasthenia gravis (MG) is an autoimmune disorder in which patients experience muscular fatigability due to the presence of anti-acetylcholine receptor (AChR) antibodies which inhibit signal transduction across the neuromuscular junction. Like all complex disorders, disease is caused by an interaction between genetic and environmental factors. Although several genes have been identified which appear to be associated with MG, both classic twin studies and current multi-gene models are insufficient to explain either disease pathogenesis or inheritance. We examined the literature on MG to determine both mono- and dizygotic twin concordance rates, and used this data to (1) estimate the proportion of the population with underlying genetic predisposition to MG and the frequency of the environmental component and (2) derive the number of inherited genetic regions that are required to confer predisposition to MG. Using a MZ twin concordance rate of 35.5%, and a dizygotic rate of approximately 4–5% (based on family data), the probability of encountering environmental components necessary to develop MG in an individual with genetic predisposition is approximately 52.4%, making the frequency of predisposition (1:5240) roughly twice the rate of incidence. Furthermore, the number of genetic regions co-inherited between affected individuals is between two and four, which may be large haplotypes with interacting activity. Determining these haplotypes, by fully sequencing associated regions in cases and controls to identify mutations present, may therefore be a practically step toward the understanding of complex disease.

Keywords: autoimmunity, genetics, mathematical modeling, complex disease

The concordance rate of twins is frequently used as a standard criterion for explanation of a genetic background for predisposition to a disorder. In the classical twin design (CTD), the pairwise concordance is given by

\[
\frac{C}{C + D}
\]

where \(C\) indicates a concordant twin pair and \(D\) indicates a discordant twin pair. The probandwise concordance rate is given by

\[
\frac{2C}{2C + D}
\]
and has been suggested to be used between studies to give similar results over different ascertainment probabilities (McGue, 1992). The monozygotic twin (MZ) concordance rate is defined as the additive genetic effect \( V_A \) plus the environmental effect \( V_C \). The dizygotic twin (DZ) concordance rate is given as

\[
\frac{1}{2} V_A + V_C
\]

due half of all genes shared, so that solving for \( V_A \) gives the heritability as equal to twice the difference in concordance between MZ and DZ twins. This is often used as a proxy for the maximum genetic component expected to be observed in case-control studies (Boomsma et al., 2002).

However, if a complex disease requires certain genetic factors to be present, which are necessary but not sufficient for pathogenesis, the disease can be considered to be 100% inherited due to inclusion of said factors in all cases (Rothman & Greenland, 2005). Similarly, if all cases have encountered certain necessary environmental factor(s), these too can be considered to be 100% required (Rothman & Greenland, 2005), making estimation of genetic/environmental contribution to a disorder difficult. Due to this situation, CTD is not useful to determine further information about either a disease’s underlying predisposition or the environmental component. For this reason, we have created a method to utilize the MZ and DZ twins rate in a constructive manner. We have therefore built a statistical model to determine the frequency of the underlying disease predisposition as well as the nature/number of targets for multi-factorial disease genetics studies. Using data from a comprehensive review of literature on MG twins, we use the model to estimate these for the disease, and further apply the technique to similar complex disorders.

**Methods and Modeling**

Monozygotic and dizygotic twins rates for both concordance and discordance was estimated by compiling literature from all known published twins studies on myasthenia gravis. Searches on PUBMED were conducted with the terms ‘myasthenia gravis’ and ‘twins’, with all returning results examined for accuracy. Those publications found to report twins with at least one diagnosed with myasthenia gravis were selected. Additionally, previous publications referenced in publications, which were not immediately returned by a PUBMED search, were considered. Cases in twins of neonatal myasthenia gravis, a transitory disorder caused by maternal anti-AChR antibodies in newborns, were omitted, as were others in which autoimmune myasthenia gravis was not indicated.

A summary of these studies was made with respect to the number and type of twin pairs, and whether the pair was concordant or discordant. The data are tabulated in Table 1.

### Table 1. The Number and Type of Twin Pairs, and Whether the Pair

<table>
<thead>
<tr>
<th>Twin 1</th>
<th>Twin 2</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>((1 - P_T) \times (1 - P_T))</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>((1 - P_T) \times P_T)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>(P_T \times (1 - P_T))</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>(P_T \times P_T)</td>
</tr>
</tbody>
</table>

The probabilities then simplify to:

\[
p(0,0) = \text{No twins with MG} \quad (1 - P_T)^2 \\
p(1,0),(0,1) = \text{Exactly one twin with MG} \quad 2 P_T \times (1 - P_T) \\
p(1,1) = \text{Both twins with MG} \quad P_T^2
\]

The rate of MZ twins concordance is equal to the proportion of both twins with MG \( p(1,1) \) to the total MZ twin pairs observed \( p(1,1) + p(1,0), (0,1) \), which is equal to

\[
MZ \text{ concordance} = \frac{P_T^2}{P_T^2 + 2P_T(1-P_T)}
\]

\( P_T \) can be solved and is given as follows:

\[
P_T = \frac{2 \times MZ}{1 + MZ}
\]

**Estimating Predisposing Genetic Regions With Twin Data**

MG is known to be a relatively rare disorder, with a worldwide frequency of approximately 1 in 10,000 individuals. Due to the necessity for both a genetic predisposition and an environmental trigger, the population background with disease predisposition, who have not contracted the disease, must be greater than or equal to the disease frequency. Since the concordance rate is reported in the range of 30–40% in monozygotic twins, it appears that the triggering event is not rare among individuals with predisposition for MG. An estimate may be obtained by assuming that monozygotic twins with predisposition for MG take the form \{(0,0),(1,0),(0,1),(1,1)\} where 0 denotes a twin who does not develop the disease and 1 denotes a twin who does develop MG. MZ twins concordant for MG are denoted as (1,1), while MZ twins discordant for MG consist of both subsets (1,0) and (0,1). Although nothing is known about the frequency of the twin pair (0,0), we assume that the environmental factor is an independent event in households with twins predisposed to MG.

We denote the probability of the first MZ twin developing MG as \( p(T_1) \), and the probability of the second twin developing MG as \( p(T_2) \). The probabilities are assumed to be independent, and equal, so that \( p(T_1) = p(T_2) \), hereafter referred to as \( P_T \), the probability of an MZ twin with genetic predisposition developing MG. Thus, the probability of an MZ twin not developing MG is 1 - \( P_T \). The probabilities of the set \{(0,0),(1,0),(0,1),(1,1)\} are then given by:

- Twin 1 Twin 2 Probability
  - 0 0 \((1 - P_T)^2\)
  - 0 1 \((1 - P_T) \times P_T\)
  - 1 0 \(P_T \times (1 - P_T)\)
  - 1 1 \(P_T^2\)

The probabilities then simplify to:

\[
p(0,0) = \text{No twins with MG} \quad (1 - P_T)^2 \\
p(1,0),(0,1) = \text{Exactly one twin with MG} \quad 2 P_T \times (1 - P_T) \\
p(1,1) = \text{Both twins with MG} \quad P_T^2
\]

The rate of MZ twins concordance is equal to the proportion of both twins with MG \( p(1,1) \) to the total MZ twin pairs observed \( p(1,1) + p(1,0), (0,1) \), which is equal to

\[
MZ \text{ concordance} = \frac{P_T^2}{P_T^2 + 2P_T(1-P_T)}
\]

\( P_T \) can be solved and is given as follows:

\[
P_T = \frac{2 \times MZ}{1 + MZ}
\]

**Determination of Genetic Regions Inherited**

The ratio of monozygotic and dizygotic twins also provides a basis to conduct an approximation of the number
### TABLE 1

Published Articles on Myasthenia Gravis in Twins

<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Reference</th>
<th>Concordant MZ</th>
<th>Discordant MZ</th>
<th>Concordant DZ</th>
<th>Discordant DZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1944</td>
<td>(Wilson &amp; Stoner, 1944)</td>
<td>Myasthenia gravis: A consideration of it causation in a study of 14 cases</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>(Osserman &amp; Teng, 1956)</td>
<td>Studies in myasthenia gravis: Neonatal and juvenile types</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>(Alter &amp; Talbert, 1960)</td>
<td>Myasthenia gravis in one monozygotic twin</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>(Simpson, 1960)</td>
<td>Myasthenia gravis. A new hypothesis</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>(Adler, 1966)</td>
<td>Myasthenia gravis bei einigen Zwillingen</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>(Haralanov &amp; Kutchonkov, 1966)</td>
<td>Myasthenie familiale (pere et fille) avec hyperplasie du thymus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>(Osborne &amp; Simcock, 1966)</td>
<td>Myasthenia gravis in identical twins</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>(Motoki et al., 1966)</td>
<td>A case of myasthenia gravis in identical twin brothers</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1967</td>
<td>(DeGroot et al., 1967)</td>
<td>Remission of myasthenia gravis and thyrotoxicosis after thymectomy</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1969</td>
<td>(Herman, 1969)</td>
<td>Familial myasthenia gravis</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1969</td>
<td>(Hokkanen, 1969)</td>
<td>Myasthenia gravis. A clinical analysis of the total material from Finland and special reference to endocrinological and neurological disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td>(Herrmann, 1971)</td>
<td>The familial occurrence of myasthenia gravis</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td>(Namba et al., 1971b)</td>
<td>Myasthenia gravis occurring in twins</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td>(Bundey, 1972)</td>
<td>A genetic study of infantile and juvenile myasthenia gravis</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td>(Keeseey et al., 1982)</td>
<td>Acetylcholine receptor antibody titer and HLA-B8 antigen in myasthenia gravis</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td>(Pirskanen, 1977)</td>
<td>Genetic aspects in myasthenia gravis. A familial study of 164 Finnish patients</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td>(Michalski et al., 1978)</td>
<td>Monozygotic twins with Klinefelter’s syndrome discordant for systemic lupus erythematosus and symptomatic myasthenia gravis</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>(Seybold &amp; Lindstrom, 1981)</td>
<td>Anti-acetylcholine receptor antibody and its relationship to HLA type in asymptomatic siblings of a patient with myasthenia gravis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>(Allen et al., 1984)</td>
<td>Myasthenia gravis in monozygotic twins. Clinical follow-up nine years after thymectomy</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>(Murphy &amp; Murphy, 1986)</td>
<td>Myasthenia gravis in identical twins</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>(Lefvert et al., 1989)</td>
<td>B cell and autoantibody repertoire in a pair of monozygotic twins discordant for myasthenia gravis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>(Dias-Tosta et al., 1989)</td>
<td>Familial myasthenia gravis: a case report in identical twins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>(Grinlinton et al., 1991)</td>
<td>A pair of monozygotic twins who are discordant for myasthenia gravis but became discordant for systemic lupus erythematosus post-thymectomy</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>(Agafonov et al., 1997)</td>
<td>Twin studies of myasthenia</td>
<td>4</td>
<td>5</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>2004</td>
<td>(Kakoulidou et al., 2004)</td>
<td>The autoimmune T and B cell repertoires in monozygotic twins discordant for myasthenia gravis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>(Punga et al., 2009)</td>
<td>Monozygous twins with neuromuscular transmission defects at opposite sides of the motor endplate</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note: Studies on twins with autoimmune MG, dating back to 1944, have been compiled. In each case, the number of MZ and DZ twins both concordant, and discordant for the disease have been tabulated.
of inherited regions required for predisposition to the disease. Since monzygotic twins share 100% of their genetic material (with the possible exception of copy number variation of genes) and dizygotic twins share 50%, the lower rate of disease concordance in dizygotic twins is related to reduced inheritance of the disease causing regions. In both situations, we assume that the risk of a triggering event is equal; that is, in the households/regions in question, the twins have had an equal risk of exposure to environmental factors irrespective of whether they are mono- or dizygotic. The equal environments assumption (EEA) has been shown to be valid (Derks et al., 2006; Kendler et al., 1993) despite some social differences between MZ/DZ twins (Horwitz et al., 2003), and we therefore omit this term. We further assume that in each of the cases of discordant twins, the same disease predisposing regions have been inherited. We then estimate the number of regions by assigning the probability of dizygotic twins (Di) inheriting the disease to be equal to that of monzygotic twins (Mo) multiplied by an inheritance ratio (β) and a recombination probability (RF) for each of n regions.

\[ p(Di) = p(Mo) \times \beta \times (1 - RF)^n \]

where the term RF represents the chance of recombination within the given region disturbing required genetic components and is therefore equal to the Morgan distance of the region.

**Inclusion of Genetic Modifiers**

Genes that are associated to a disorder may be neither a necessary, nor sufficient, component of disease predisposition. These genes and SNPs may be those that are frequently found to be associated with complex diseases with modest odds ratios (OR), which consistently replicate in several populations. Due to their role in immune system regulation, they might confer an additional risk of contracting the disease, but only for those groups who carry the required underlying predisposition. In this sense, these genes can be considered as modifiers, whose effects are similar to environmentally contributing factors, and should therefore be included in the varying component. Epigenetic factors, which may vary even between twins, may also be considered as modifiers since the underlying disease genes are unaltered.

It is likely that many genetic associations that have been found for complex disorders are of the ‘genetic modifier’ variety, which occur with relatively similar allele frequencies between cases and controls, but are slightly elevated in cases wherein the effect is observed. An exacerbating circumstance is that many subgroups (or syndromes of disease) may be present for complex disorders, and therefore each modifier gene may act in only one of these groups. A recent study investigating the effect of modifier genes on subgroups determined that using 2000 samples, modifier genes can be detected at 80% power only for ORs greater than 1.8 (Bergen et al., 2010). In MG, the strength of association with non-MHC genes rarely exceeds this limit, nor is association frequently observed across replicate cohorts. For example, the PTPN22 R620W has consistent association with MG, yet it has reported ORs in the 1.52 (Lefvert et al., 2008) to 1.97 (without thymoma, anti-titin negative) (Vandiedonck et al., 2006) range.

Genetic modifiers may also confound the previous calculation with respect to inherited regions required, if they are infrequent with large effects. We can modify the previous equation by first including separate values of β and RF per inherited disease haplotype, since these may have different inheritance patterns:

\[ p(Di) = p(Mo) \times \prod_{j=1}^{n} \left[ \beta_j \times (1 - RF^j) \right] \]

and then include genetic modifiers, denoted as ε with effect sizes (0<ε<1), occurring m times:

\[ p(Di) = p(Mo) \times \prod_{j=1}^{n} \left[ \beta_j \times (1 - RF^j) \right] \times \prod_{m=1}^{\epsilon} \left[ \beta_m \times (1 - \xi_m) \right] \]

With the inclusion of genetic modifier(s) occurring with effect >0, the number of necessary inherited regions would reduce for all β, < 1, and therefore we accept the number of regions (with a recessive model, to be discussed) to be the upper limit.

**Results**

Based on these studies, the monzygotic twin concordance rate in myasthenia gravis is 11 of 31 pairs, or 35.5% pairwise (52.4% probandwise). Of 16 dizygotic twin pairs, none were concordant for MG. However, this number is too low to draw an accurate conclusion, especially if the rate is less than 5%, and the true DZ twins rate may be similar to the familial rate, which measured to be between 3.8 and 7.2% (Namba et al., 1971a; Pirskanen, 1977).

The result of the predisposition calculation in MG using an MZ pairwise concordance rate of 35.5% is that \( P_{\tau} \), or the rate of the environmental trigger, is estimated to be 52.4% in predisposed twins. Probandwise concordance rates are not used since MZ concordance may already be overrepresented due to selection bias and with high ascertainment probability, the expected probandwise concordance is inflated relative to pairwise concordance (McGuire, 1992). As a consequence of this calculation is that this rate can be used to approximate the frequency of genetically predisposed individuals in the general population, by dividing the rate of incidence by \( P_{\tau} \). In the case of MG, a 52.4% triggering event would make the overall frequency of genetic predisposition approximately 1 in 5240. Therefore, the inclusion of genetically susceptible individuals in the control population is unlikely, as it requires over 3600 controls to expect one to be included (1- \( p(x) \) > 0.50).
This method can be applied to other complex diseases, such as RA, which has a pairwise MZ concordance of approximately 13.5% (12–15%) (Gregersen, 1999). This gives a \( P_T \) value of 0.238, suggesting that the environmental factor(s) are more rarely encountered. It also implies that the number of people with predisposition to RA is approximately 4 times higher than the rate of incidence (1–2%), and that individuals with genetic predisposition to the disease are likely to be included in control groups at a rate of 3–6%. In celiac disease (CD), pairwise MZ concordance of 75% (Greco et al., 2002) and 71.4% (Nisticò et al., 2006) give a \( P_T \) value of approximately 0.83, suggesting that the environmental component is encountered by most predisposed individuals. In Type I diabetes (T1D), pairwise concordance has been shown to be 27.3% (Hyttinen et al., 2003) in 22,650 Finnish MZ twin pairs, yielding a \( P_T \) value of 0.44. The best estimate for systemic lupus erythematosus (SLE) concordance, 24% in MZ twins (Deapen et al., 1992), gives a \( P_T \) value of 0.39, very similar to that in T1D.

When estimating the inherited regions in MG, the term \( p(Mo) \) is given by the MZ concordance rate, while \( p(Di) \) is given by the DZ concordance rate, estimated from the familial rate. For mathematical simplicity, approximations for \( p(Mo) \) of 40% and \( p(Di) \) of 2.5% are used. We can thereafter determine the expected minimum and maximum values under the assumption that all regions must be inherited from both parents (\( \beta = 0.25 \)) under a recessive model or that a single copy is required (\( \beta = 0.50 \)) using a dominant model. Solving the equation for \( n \) using these \( \beta \)s (and for the moment assuming RF = 0) gives a solution that between two and four regions are required, which are inherited in their entirety. If these regions are long and RF is greater than 0, the decreased chance of joint inheritance will result in a reduced number of required regions. However, since meiotic recombination occurs relatively infrequently over vast stretches of DNA, it is likely that this term will be very close to zero.

For RA, a large-scale study found MZ and DZ concordance rates to be 15.4% and 3.6% (Silman et al., 1993), respectively, which corresponds to between one and two inherited regions. In T1D, MZ (27.3%) and DZ (3.8%) concordance rates correspond to between one and three regions of inheritance. In CD, the average twins concordance rates (73.2% in MZ twins and 10.1% in DZ twins; Greco et al., 2002; Nisticò et al., 2006) indicate that between one and three predisposing regions are present. In SLE (24% in MZ twins and 2% in DZ twins; Deapen et al., 1992), the number of regions can be estimated to be at least two and up to six. Although these regions may contain many SNPs, this appears to contradict the current assumption that large numbers of genes acting with small effects cause complex disorders, and this premise will therefore be examined.

**Discussion**

The proposed model utilizes twins data in two novel ways, namely to estimate the frequencies of both disease predisposition and the environmental triggering event and to estimate the number of jointly inherited regions which cause disease. Both of these applications greatly enhance the use of twin data, which previously were used only to estimate the genetic/environmental contributions via 'heritability', which in the case of MG would be \( 2 \times (35.5\% - 3-7\%) \), or approximately 60%. Determining the rate of predisposition (approximately 1 in 5420) for MG and complex disorders gives important clues as to the underlying inheritance mechanisms, as does the number of regions contributing to disease.

The finding that a small number of regions may be necessary to be jointly inherited to give predisposition to many complex disorders is surprising as it has been widely assumed that all genes associated with a disorder have a role in disease pathogenesis, which is unlikely given this inheritance pattern. There are several possible explanations for this result, given the high number of associated genes to certain complex disorders, which can number 40 or more (Manolio et al., 2009).

First, there may exist subclasses of complex diseases, which are responsible for the large genetic heterogeneity observed. We have thus far considered only an individual family’s inheritance of disease predisposing elements, such that the inherited regions are representative of one subclass of MG causing disease in a particular twin pair. However, the Rothman pie model and ‘syndrome of diseases’ approach to complex disorders point to a more complicated situation, in which genes causes which predispose to disease are not similar for all patients (Figure 1). This is highly likely for the vast majority of complex disorders exhibiting genetic heterogeneity; otherwise,

**FIGURE 1**

Rothman pie model of sufficient causes underlying a complex disorder.

Note: The Rothman pie model of complex disease illustrates how multiple factors act in tandem in complex disease. Each pie represents one subclass of disease, or an individual disorder in a ‘syndrome of diseases’, and each letter represents a genetically necessary, but not sufficient contributor to disease predisposition. Taken together, all required elements of one pie confer disease predisposition.
genome wide scans would certainly have detected SNPs with extremely high odd ratios consistently in all patients. Myasthenia gravis is known to have different associations in subgroups of patients, such as the previously mentioned HLA association of EOMG and LOMG. These genetic subclasses may overlap with clinical subgroups that show such strong association.

Second, the SNPs known to be associated with a complex disease may not be the actual predisposing genetic elements, but instead could be markers for the causative elements in linkage disequilibrium (LD) with them. This would reduce the observable proportion of the marker SNP’s risk allele in affected individuals, and thereby mask the allele frequency of the disease predisposing SNP(s). Given that a SNP B is in LD with a the true predisposing SNP A, the measure of correlation (r) gives the normalized value of the linkage disequilibrium (D) and is 0 when LD is not observed and 1 when SNP B is a perfect proxy for SNP A. A comprehensive study on the issue determined that maximum alleleic variation between two SNPs with a squared correlation (r^2) of 0.8 is +/-0.06, which is sufficient to reduce the power of a study from 80% to detect Type I error of 5x10^-5 to 60% (Wray, 2005). For r^2 = 0.5, the maximum allele variation increases to 0.16, strongly reducing the ability to detect the SNP in case/control studies.

Lastly, although the number of regions contributing to complex disease may be low, the number of SNPs per region is unknown and may be high. A SNP known to be associated with a complex disease may interact with other SNPs nearby, thereby diluting the measured association of the SNP individually. This constitutes a special case of a SNP in LD with the predisposing elements, with the additional constraint of a haplotype of two or more predisposing SNPs. Here the effect of SNP B in LD with predisposing SNPs A and C cannot be estimated by examining the r^2 values between B and each of the predisposing SNPs, since the presence of conserved regions and recombination hotspots alters the linkage patterns. Therefore, a shift in focus in complex disease genetics to large haplotypes may be useful.

Another interpretation of these findings may be to clarify the frequency of disease causing mutations/regions in the population. The common disease common variant (CDCV) hypothesis holds that SNPs with allele frequencies > 1% in the population without deleterious effects alone may contribute to complex disease, in tandem with other genetic factors as assumed in the subclass model. However, if MG predisposition occurs in roughly 1:5240 individuals in the population, with the same 2–4 disease causing regions in all, an equal distribution of these regions in the population would dictate their frequency to be between 1:8.5 and 1:70 (\(\sqrt[4]{5240}\)). The presence of subclasses would reduce these frequencies, by altering the frequency of predisposition in the population, that is 1:5240 x subclasses; therefore, 4 equally sized subclasses with normal distribution would exist in the population with frequency 1:12 to 1:144. As mentioned, haplotypes that occur at the lower end of these frequencies are unlikely to be represented by SNP markers adequately, and therefore regions that are associated with a disorder should be fully sequenced in patients and controls to determine the relevant genetic elements present.

**Acknowledgments**

The study was supported by the Swedish Research Council, the Palle Ferb Foundation, and by a grant (U19AI067152) from the US National Institute of Allergy and Infectious Diseases.

**References**


