Editorial

Issues in Investigating the Genetic Epidemiology of Disease Expression

Despite the extensive statistical modeling that frequently accompanies genetic studies, the disciplines of genetics and epidemiology have historically functioned separately. However, over the last 2 decades, epidemiological methods have been slowly integrated into genetic concepts. Often the initial recognition of a genetic contribution to a complex disease has been based on convincing epidemiological data from family and twin studies. Further, the magnitude of genetic burden, and to some extent the genetic modeling, can be determined by the recurrence risk in relatives, as proposed by Risch[1]. Thus epidemiologic studies have had important implications for genetic counseling, genetic modeling, and gene identification.

To date, much of the effort in studying the genetics of complex disease has been directed toward elucidation of genetic factors related to disease susceptibility. An area of increasing interest is the identification of genetic determinants of disease expression. This interest has led to a recent surge of genetic epidemiological studies from cohorts that were assembled primarily for linkage studies, as illustrated by recent studies in lupus, rheumatoid arthritis, and juvenile rheumatoid arthritis[2–4]. The question now is whether cohorts assembled for identification of disease susceptibility genes can also be used to ascertain relevant information regarding the epidemiology of disease expression[5].

In this issue of *The Journal*, the concordance of disease severity among family members with ankylosing spondylitis (AS) is addressed with respect to disease activity, function, and radiologic change, as well as the prevalence of iritis and juvenile onset arthritis[6]. The immediate rationale for exploring the epidemiology of familial disease expression in AS is to further delineate the relative genetic versus environmental influence on various disease manifestations, as well as to assess the clinical relevance of concordance data regarding its utility for genetic counseling of individual families. The longterm goal of such an initiative is to aid in the identification of genetic factors related to disease expression so that our ability to prognosticate outcome and rationalize therapeutic interventions is enhanced.

Identification of susceptibility genes for complex diseases such as AS has proven difficult. Thus, it is not unreasonable to postulate that identification of genetic determinants for disease expression may be an even more daunting task given the additional heterogeneity that is likely to exist with expression of various manifestations[7]. Addressed in this editorial are additional issues that need to be considered when exploring the epidemiology of disease expression as opposed to disease susceptibility.

First, in order to study disease expression, we need to expand our present practice regarding data collection in genetic studies of complex disease. Currently there has been a welcoming trend towards strict adherence to diagnostic criteria, with minimal inclusion of patients with atypical features, in an attempt to reduce phenocopies and locus heterogeneity. While this degree of scrutiny to diagnose a given disorder is also essential for studies in disease expression, it is not sufficient. In addition to confidently diagnosing the disease, a concerted effort must be made to adequately phenotype relevant disease manifestations using a standardized and systematic approach. Such an undertaking necessitates a better validation of the parameters being collected. This is because disease manifestations that are not essential to the diagnostic criteria are often solely ascertained by self-reporting from patients. In the present study by Brophy, *et al*[6], AS was diagnosed by stringent criteria (New York disease criteria[8]) that required clinical and radiological evidence. A fair degree of scrutiny also went into diagnosing selected extraarticular manifestations such as iritis, inflammatory bowel disease, and psoriasis as these manifestations were confirmed by physicians. Unlike certain genetic databases, the Bath Ankylosing Spondylitis database is a multipurpose database with a strong foundation in epidemiological as well as genetic studies.

See Concordance of disease severity among family members with AS? *page 1775*

Second, prior to venturing into epidemiological studies of disease expression, one must assemble an adequate sample size. From a traditional epidemiological standpoint, for a given sample size the power to detect various disease manifestations (or subsets of a disease) is less than the power to detect an association with the primary disease, and the variance of the disease manifestations will also likely be greater than the disease. Thus larger cohorts need to be assembled to answer questions related to disease expression. Brophy, *et al* have gained access to valuable resource to address their questions: they are working with the Bath Ankylosing Spondylitis database, which includes data on 5507 AS cases[6]. This resource includes 406 families with 325 sibling pairs and 213 affected parent-child pairs. But even in this large sample cohort, the concordance of rare manifestations of AS, such as apical lung fibrosis, cannot be adequately studied due to insufficient sample size.

Third, selection of the disease manifestations (phenotypes) being evaluated is critical. Unlike the genotype, which can now be confidently measured whether by polymerase chain reaction or DNA sequencing, the phenotype is a complex product of the genotype, environment, and epigenetic factors. To facilitate identification of potentially heritable disease manifestations, epidemiological studies have been incorporated into traditional genetic designs, such as affected sibling pairs. These studies have demonstrated clustering of several manifestations in families with lupus (thrombocytopenia, discoid rash, neurologic disorders, and hemolytic anemia)[2]; rheumatoid arthritis (rheumatoid factor, RF, nodules, and age of onset)[3]; and juvenile rheumatoid arthritis (tenosynovitis, leukocytosis, RF, and anemia)[4]. In the study by Brophy, *et al*, children of AS parents with iritis were more likely to develop iritis than children of AS parents without iritis, and parents with juvenile AS were more likely to have children with juvenile AS compared to non-juvenile AS parents[6]. From a genetic perspective this would imply that there is an underlying set of genes, independent of susceptibility to AS, that encodes for these manifestations. These attempts are justified by the assumption that finding the gene(s) for these simpler phenotypes will enhance our knowledge of the genetic mechanisms in complex disease.

A potentially more meaningful phenotype is that of an endophenotype. As noted in a recent editorial, an endophenotype-based approach has the potential to enhance the genetic dissection of complex diseases[9]. The criteria set forth by Gottesman and Gould for psychiatric genetics can also be applied to rheumatology[9]. They suggested that the endophenotype should be heritable, be primarily state independent (manifest in an individual whether or not illness is active), cosegregate with illness within families, and be found in non-affected family members at a higher rate than in the general population. Unfortunately, the identification of such phenotypes has often been elusive. Another helpful feature that may facilitate gene identification is the use of quantitative rather than categorical definitions of phenotypes.

A divergent approach was also undertaken by Brophy, *et al* as they assessed the concordance of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and Bath Ankylosing Spondylitis Functional Index (BASFI), which are complex composite phenotypes. In a previous study from this group, the BASDAI and BASFI were noted to have a heritability of 0.51 and 0.63, respectively, which refers to the proportion of the total variance that is genetic[10]. In the present study, Brophy, *et al* concluded that sibling pairs showed significantly more concordance for disease activity and function than did parent-child pairings. These efforts, however, are not devoid of difficulties, and should be interpreted with caution until they are independently validated. For instance, disease activity is often a transitory state reflecting a single point in time and fluctuates greatly based on numerous parameters including physician advocated treatments. Further, the BASDAI and BASFI are ascertained via questionnaire, and this self-reporting is subject to variability among different individuals or even within the same individual at different points in time. Additional concerns are raised when assessing the concordance of parent-child pairings, as marked differences in ages among such pairings can result in significant clinical discordance, even when seemingly appropriate adjustments are made for disease duration. Finally, the clinical relevance of reporting such a correlation should be carefully evaluated. Despite a concordance of 0.27 and 0.36 in sibling pairs for the BASDAI and BASFI, respectively, the authors dutifully note that the concordance data were not clinically relevant for predicting disease outcome, as the proportion of variation of disease activity that could be accounted for by the disease activity of the sibling was only 7%.

The Bath Ankylosing Spondylitis Radiologic Index (BASRI), on the other hand, may be a more robust measure of disease concordance. Unlike the BASDAI and BASFI, the BASRI is a cumulative index and likely exhibits greater objectivity, as this is not a self-reported measure. Thus the BASRI may be a better phenotype to assess the concordance of disease severity in AS. Brophy, *et al*, note a heritability of 0.62 for the BASRI[6]. Multiple testing and the generation of numerous false positive associations that predictably follows is also a concern, as many more phenotypes are now being simultaneously tested with countless numbers of genotypes. In order to analyze these studies, departures from conventional statistical methods, such as principal component analysis, may be required.

Finally, under most circumstances epidemiological designs are suitable for assessing the impact of genes as a potential risk factor. Thus, collection of DNA in most epidemiologically designed studies should allow molecular characterization of the phenotype being studied. On the other hand, as linkage studies are not part of a typical epidemiological design, such cohorts should be used cautiously when

studying the epidemiology of disease expression. For instance, there may be an ascertainment bias when sibling pairs gathered for linkage studies are used to study the epidemiology of disease expression, as linkage studies often oversample families with multiple affected siblings. That being said, as noted by these recent studies, potentially informative data can be elicited from such cohorts.

PROTON RAHMAN, MD, MSc, FRCPC,
Associate Professor of Medicine,
Division of Rheumatology,
Memorial University of Newfoundland,
St. John's, Newfoundland, Canada

Address reprint requests to Dr. Rahman, St. Clare's Mercy Hospital, 1 South, 154 LeMarchant Road, St. John's, Newfoundland A1C 5B8, Canada

REFERENCES

1. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 1990;46:222-8.
2. Tsao BP, Grossman JM, Riemekasten G, et al. Familiality and co-occurrence of clinical features of systemic lupus erythematosus. Arthritis Rheum 2002;46:2678-85.
3. Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. Clustering of disease features within 512 multicase rheumatoid arthritis families. Arthritis Rheum 2004;50:736-41.
4. Moroldo MB, Chaudhari M, Shear E, Thompson SD, Glass DN, Giannini EH. Juvenile rheumatoid arthritis affected sibpairs: extent of clinical phenotype concordance. Arthritis Rheum 2004;50: 1928-34.
5. Risch N. Evolving methods in genetic epidemiology. II. Genetic linkage from an epidemiologic perspective. Epidemiol Rev 1997;19:24-32.
6. Brophy S, Hickey S, Menon A, et al. Concordance of disease severity among family members with ankylosing spondylitis? J Rheumatol 2004;31:1775-8.
7. Criswell LA. Familial clustering of disease features: implications for the etiology and investigation of systemic autoimmune disease. Arthritis Rheum 2004;50:1707-8.
8. Moll JMH, Wright V. New York clinical criteria for ankylosing spondylitis. Ann Rheum Dis 1973;32:354-63.
9. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 2003;160:636-45.
10. Hamersma J, Cardon LR, Bradbury L, et al. Is disease severity in ankylosing spondylitis genetically determined? Arthritis Rheum 2001;44:1396-400.