

Collaborative epidemiology: From concept to execution

A Dessa Sadovnick

Department of Medical Genetics & Faculty of Medicine, Division of Neurology, University of British Columbia, Vancouver, Canada

Abstract

Genes, environment and the interaction thereof are widely indicated in multiple sclerosis (MS) susceptibility and pathobiology. MS is now being increasingly recognized in immigrants from “low risk” to “high risk” regions. Thus there is great interest in characterizing the MS phenotype and genotype for various Pan-Asian ethnic groups and in comparing MS immigrants who develop MS after migration to those of the same ethnicity who remain in the low risk regions. Much of what we have learned about the genetics, genetic epidemiology, environmental risk factors, roles of genes in disease susceptibility and resistance, clinical course, and progress, all important for better understanding the disease susceptibility, pathogenesis and natural history has come from longitudinal, multicentred research projects including the Canadian Collaborative Project on Genetic Susceptibility to MS (CCPGSMS). Now that the Pan-Asian Committee for Treatment and Research in Multiple Sclerosis (PACTRIMS) has had its inaugural meeting and the Registry of Asian Pacific Idiopathic Demyelination (RAPID) study is in progress, it is critical that collaborative endeavors must be organized and standardized to allow comparison of findings not only within Pan-Asian regions but also with data from other countries to which persons from these regions have migrated. This article focus on lessons learned in conducting the Canadian Collaborative Project on Genetic Susceptibility to MS (CCPGSMS) which will hopefully facilitate the collaborative endeavours planned for the Pan-Asian region.

INTRODUCTION

My objective with the title “Concept to execution” is to discuss the practical aspects of collaborative research to ensure data are comparable and results are reproducible.

Many lessons have been learned from the longitudinal, population-based Canadian Collaborative Project on Genetic Susceptibility to MS (CCPGSMS). Unique strengths of CCPGSMS include:

1. *Living database:* This means that it is not static in time, thus allowing updates and changes in all field. The database must be adaptable to allow linkage to other databases as necessary and the ability to access subsets of the database for specific analyses.
2. *Blank spaces:* All data forms and fields in the database have specific codes. Blank spaces are queried and not accepted. Blanks can be misleading. For a specific question, e.g. “was CSF examined”, a blank space could mean “the question was erroneously not asked or not replied to”, “the test was done but results are not known”, “the centre never does this test”, the test was “not applicable” for this person because of other medical issues, etc.
3. *Avoidance of duplicate entry:* In the CCPGSMS, various methods are used to ensure that duplicate entry of cases is avoided to the best of our ability.
4. *Longitudinal nature:* This was deliberate when the CCPGSMS was first envisioned. This allows for follow-up of families and individuals as well for enlargement of sample size using the same inclusion and exclusion criteria as well as replication of findings. Without the longitudinal component, several key findings of the CCPGSMS may have been missed such as maternal effects¹⁻³ and temporal prevalence changes.⁴ The longitudinal nature of the CCPGSMS also allows adjustments and reassessments as individuals enter and exit the most common age of onset decades for multiple sclerosis (MS), for example verifying “age adjustment” for recurrence risk data.^{5,6} This also allows for answering questions raised by results of earlier CCPGSMS studies.

5. *Ongoing contact with families and ability to update both clinical data and biological samples:* This is possible for the CCPGSMS as the families are contacted through the Canadian MS Clinic that they attend. In other words, the high level of cooperation from MS patients and their family members appears to come from the fact that the participants deal with personnel whom they can associate with the source of their ongoing care. While data basing is centralized, this regional collection of data is important to optimize cooperation.
6. Data (demographics, medical, occupational, ethnicity, gender, age, environmental exposure measures, etc.) and biological materials (DNA, RNA, serum) are collected from affected & unaffected family members, including various degrees of affected individuals and intervening unaffected relatives, as deemed appropriate for the study endeavours and with appropriate consents.
7. A variety of controls are available and this is critical. Suitable controls vary with the questions being studied. The CCPGSMS study group represents a wide range of ethnic diversity and socio-economic status.

CONTROLS

The selection of controls at the beginning of any multi-centred study must be carefully evaluated. For long-term, multifaceted studies, different controls are needed for different comparisons. The case-control model is useful in many instances but is not the only control to be considered. Controls may need to be ascertained from the general population, from "other disease" groups, from unaffected family members, from friends, from spouses, etc. Appropriate selection of controls is important. However, one consideration often overlooked in the selection of controls is the influence of gender. This is particularly important in diseases including MS, which have a gender bias among patients.⁴

Recent work from the CCPGSMS has been important in this area. The sex-specificity of recall and reporting bias and the greater female awareness of medical history have long been qualitatively known to experienced practitioners. Recent publications on risk factors and the occurrence of other autoimmune diseases are believed to be, when taken together, the first to both validate and quantify this in MS.⁷⁻¹⁰

DIAGNOSTIC RELIABILITY: INDEX CASES AND AFFECTED FAMILY MEMBERS

Accurate diagnosis is critical not only for the patient identified through the study but also when investigating the familial nature of the disease, i.e. reported cases in other family members. The methodology used for the diagnostic reliability has been outlined in numerous CCPGSMS publications.^{1,2,6} This issue is more complex for studies in Pan-Asian communities where diagnoses include "classical MS", opticospinal MS, neuromyelitis optica, and as has been discussed at PACTRIMS, some diagnostic criteria may be neither clear-cut nor universally accepted.

ASKING THE QUESTIONS

Questions, whether asked in person or through questionnaires must be asked with the same cues. For example, a pilot study was done in the MS Clinic at the University of British Columbia several years ago (AD Sadovnick, unpublished data) on the accuracy of reporting familial MS. The clinic neurologist (DW Paty) would ask a patient "does anyone in your family have MS". Later, during the same clinic visit, the geneticist (ADS) would ask about specific categories of relatives (siblings, parents, aunt/uncles, nieces/nephews, first cousins) and for each category, would specifically question about anyone with neurological disorder, multiple sclerosis, sudden loss of vision especially in one eye, numbness, and/or gait disturbance. With appropriate consents, medical records were obtained for these family members. For first degree relatives (parents, siblings), the information obtained by the neurologist was accurate. However, for the more distant relatives, it was rare that persons with MS were disclosed to the neurologist. A twin study¹² found that monozygotic twins were twice as likely to be recorded as twins in medical records than were dizygotic twins. In contrast, when specifically asked by a geneticist, all live born twins were identified.

DEFINE EXACTLY WHAT IS BEING ASKED

Data collection, whether by interviews or self-reporting on forms or online, must be structured and validated. However, even an apparently simple query must be well thought out in multi-centre studies to ensure consistence of meaning. Terms must be clearly defined. For this purpose, the

CCPGSMS has a detailed study manual for each centre participating in the collaboration as well as a toll-free “help line” allowing all study personnel direct access to the study principle investigators and study-wide coordinators.

Simple information such as onset of disease and occupation can be fraught with potential pitfalls with respect to consistency of data collected. In preparing for the CCPGSMS, we were surprised to find that interviewers were confused between the terms “onset” and “diagnosis”. Both were necessary for various analyses but it was critical that interviewers were clear on the definitions as used for the CCPGSMS.

Other unexpectedly complex questions include deceptively ones about “marital status” and “occupation”. Various points in a person’s life may be important for analyses. Therefore, multi-centre collaborations should be clear if questions refer to specific points in a participant’s life, e.g. MS onset, pre-MS onset, after MS diagnosis, etc.

TERMINOLOGY CAN DIFFER EVEN WHEN YOU THINK YOU ARE USING THE SAME LANGUAGE!

Multicentre studies can cross many languages and cultures. Common questionnaires (whether completed by an interviewer or self-completed) need translation and then back translation. This is complicated even when centres appear to use the same language. To illustrate, English is used in North America and the United Kingdom (UK). Hence, asking about the type of school attended would, on the surface, appear to be clear cut, but this is not the case. In North America, a “public” school is government funded and accessible to all children whereas a “private” school has specific entrance requirements that can include special exams, grades, parentage, gender, etc. and private payment. In the UK, these terms mean exactly the reverse – a public school is “exclusive” whereas a “private” school is available to everyone.

Back translation is helpful and important. In the CCPGSMS, we use English and French forms. One question asks about early life events for MS patients when they were in kindergarten. In English Canada, this refers to a half-day program at the same school where the person will then attend grade 1. Upon back translation from the French form, it became clear that the interviewers in Quebec interpreted this term to mean “day care in general” rather than a school level.

Obviously, if such caution must be taken in Canada which has a relatively homogeneous population, this will be carefully addressed in any multicentre Pan-Asian endeavour.

PARTICIPANTS VERSUS NON-PARTICIPANTS AND DROP-OUTS

These must be recorded in a standardized manner with respect not only to reason, but also to include profile on whatever demographics are available (e.g. gender, urban vs rural, age, degree of disability). As learned through drug trials, drop-out may not necessarily reflect compliance and this can confound results if not adjusted for.

COLLABORATORS MUST FEEL THAT THEY ARE BENEFITTING FROM THE COLLABORATION

Multicentre collaborations can fail for a variety of reasons but common reasons include lack of proper direction and collaborators feeling that their own interests are overshadowed by the goals of the multicentre collaborative effort. Effort should be made in any Pan-Asian collaboration to respect and enable the objectives of the various members of the research group so that their special interests are encouraged.

LEADERSHIP

Large collaborative multicentre efforts need leadership if progress is to occur. In addition to ensuring that participating sites are productive and adhering to protocol, leaders must address various issues including:

1. Priorities (ranking) for collaborative research proposals after objective, informed reviews;
2. Responsibilities (e.g. grant submission, paper submission, etc.);
3. Allocation of funds;
4. Assessment of productivity;
5. Authorship issues.

CONFIDENTIALITY MUST ALWAYS BE MAINTAINED

This is of course a major concern in multicentre studies and a thorough discussion is well beyond the scope of this elongated abstract. However, specific for Pan-Asian collaborations, it must be remembered that although the requirements of local ethics committees may be different from those for multicentre studies in North America, attention should be given to this matter if the long-

term intent is to compare data from Pan-Asian studies with those for similar ethnic groups from North America.

SO WHY DO PAN-ASIAN MULTICENTRE STUDIES?

There are many reasons why the Pan-Asian region, including Iran, provides an excellent opportunity for multi-centre collaboration studies in MS. These include:

1. Genetic and environmental heterogeneity of the population in this region.
2. Local regional differences towards illness (disclosure, patient care including pharmaceutical versus alternate/therapies management, community services, acceptance, etc.).
3. Asians may respond differently to therapies, especially disease modifying therapies than do Caucasians.

The Pan-Asian region can also add much necessary information about effects of genetic and environment on disease susceptibility and course because this region is noted for having a large migratory population to high risk areas such as North America and Europe. Canada (and British Columbia in particular) is well populated by new migrants, first-generation Canadian born individuals, and offspring of multicultural matings to allow many comparisons between the Asians who have move and those who have remained in their region of origin including:

1. The clinical course of MS (classical versus opticospinal) can be compared among migrants, non-migrants and first generation in terms of frequency, natural history and response to treatment.
2. Large migrant population also allows comparison including potential differences (genetic, environmental, clinical) among those who migrate and those who do not. One cannot assume that migrants resemble non-migrants in all ways.
3. Comparisons of new migrants who develop MS to first generation and those from multicultural matings, especially focusing on whether the parent from a "low risk" group is the mother or the father.

In conclusion, although the effort to standardize data collection over a very large geographic region

will be intense, the insights to be gained into the disease pathobiology should readily overcome the initial work necessary to get such projects underway. The Registry of Asian Pacific Idiopathic Demyelination (RAPID) study is an excellent first step but research in this region of the world needs to be increased and coordinated.

ACKNOWLEDGEMENTS

The author thanks Bayer Canada for providing travel funds to PACTRIMS. The CCPGMSMS is funded by the MS Society of Canada Scientific Research Foundation.

REFERENCES

1. Ebers GC, Sadovnick AD, Dyment DA, Yee IML, Willer CJ, Risch N. Parent of origin effect in multiple sclerosis: observations in half siblings. *Lancet* 2004; 363:1773-4.
2. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC and The Canadian Collaborative Study. Twin concordance and sibling recurrence rates in multiple sclerosis. *PNAS* 2003; 100(22):12877-82.
3. Herrera M, Ramagopalan SV, Orton S, Chao MJ, Yee IM, Sadovnick AD, Ebers GC. Parental transmission of MS in a population-based Canadian cohort. *Neurology* 2007; 69(12):1208-12.
4. Orton SM, Herrera BM, Yee IM, Valdar W, Ramagopalan SV, Sadovnick AD, Ebers GC; Canadian Collaborative Study Group. Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol* 2006; 5(11):932-6.
5. Ebers GC, Koopman WJ, Hader W, Sadovnick AD, Kremenchutzky M, Mandalfino P, Wingerchuk DM, Baskerville J, Rice GP. The natural history of multiple sclerosis: a geographically based study: 8: familial multiple sclerosis. *Brain* 2000; 123 (part 3):641-9.
6. Ebers GC, Sadovnick AD, Risch NJ, and the Canadian Collaborative Study Group. A genetic basis for familial aggregation in multiple sclerosis. *Nature* 1995; 377(6545):150-1.
7. Ramagopalan SV, Dyment DA, Valdar W, Herrera BM, Criscuoli M, Yee IM, Sadovnick AD, Ebers GC. Autoimmune disease in families with multiple sclerosis: a population-based study. *Lancet Neurol* 2007; 6(7):604-10.
8. Ramagopalan SV, Herrera BM, Valdar W, Dyment DA, Orton SM, Yee IM, Criscuoli M, Atkins K, Ebers GC, Sadovnick AD. No effect of birth weight on the risk of multiple sclerosis: a population based study. *Neuroepidemiology* 2008; 31(3): 181-4. Epub 2008 Sept 11.
9. Rapagopalan SV, Valdar W, Dyment DA, DeLuca GC, Orton SM, Yee IM, Criscuoli M, Ebers GC, Sadovnick AD for the CCPGMSMS. No effect of preterm birth on the risk of multiple sclerosis: a population based study. *Biomed Central Medical Genetics* 2008; 8(30): pg. Epub 2008 Aug 1.

10. Ramagopalan SV, Valdar W, Dyment DA, DeLuca GC, Yee IM, Giovannoni G, Ebers GC, Sadovnick AD for the Canadian Collaborative Study Group. Association of Infectious Mononucleosis with multiple sclerosis: A population-based study. *Neuroepidemiology* (in press).