

Supplementary Online Content

Kuo C-F, Grainge MJ, Valdes AM, et al. Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. *JAMA Intern Med*. Published online July 20, 2015.
doi:10.1001/jamainternmed.2015.3528.

eMethods. Methods for Genealogy Reconstruction, Threshold Liability, Model, and Definitions of Socioeconomic Data

eTable. Code List for Sjögren Syndrome and Other Autoimmune Diseases

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Methods for Genealogy Reconstruction, Threshold Liability, Model, and Definitions of Socioeconomic Data

Genealogy reconstruction using the National Health Insurance Research

Database

Genetic epidemiology is a powerful discipline in genetic studies to investigate the gene-environment interaction in causing diseases and their mode of inheritance within families.¹ Accurate and comprehensive genealogical and medical data are prerequisites for such research to generate evidence of familial aggregation of diseases and to test the mode of inheritance. These advances may aid clinicians to identify predispositions of individuals based on their family histories.

Population-based genealogy databases with linked medical data greatly facilitate genetic research and clinical application.² However, good resources are rarely available for research and often have restrictions on access or linkages due to technical difficulties or privacy considerations. Comprehensive genealogies of the Utah³ and Icelandic populations⁴ are well-known resources for research, with linkages to medical data in some of the participants. The Scotland⁵ and the Scandinavian countries have established nationwide genealogical registries which contain demographic data and linkage to specific disease registries.⁶ The resource is lacking in population of non-European descents.

The National Health Insurance Research Database (NHIRD) contains comprehensive information on demographics, medical diagnoses, medical expenditure, details of prescriptions, examinations, and procedures of the entire population of Taiwan. In particular, relationships between beneficiaries are also available, which allows for genealogy reconstruction in a national level.

The registry of beneficiaries, one of the registration files, contains details of demographics, residence, family relationships, occupation categories, insurance status and insurance amount of all beneficiaries of National Health Insurance. By the NHI Act and its enforcement rules, parents and grandparents who are unemployed or offspring or grandchildren who are either under 20 years of age and unemployed (including those who are in school without employment) or over 20 years of age but incapable of making a living can serve as dependents of the insured person. Unemployed spouse can also serve as dependent. The procedure to ascertain dependent status is strict and is part of the civilian registration; in general, a birth certificate issued by the medical facility who delivered the child or a DNA parentage testing for those who were not born in medical facilities is required.

The relationships recorded in the NHIRD include spouse, parent, offspring, and grandparents, grandchildren, great grand-parent and great grand-children on both paternal and maternal sides. Indirect identification of parent-offspring relationship is also possible. A parent-offspring relationship exists between grand-parents and parents of an individual. Offspring can also be recognised as children of the spouses.

The identification of siblings was based on sharing one or more common parents. In this study, we did not identify half-siblings although identification of half-siblings is possible. Only spouses in their first marriage were used to construct a family. If a couple did not have offspring, they were not assigned to the same family because they were not related based on a common blood relative.

The Registry of beneficiaries is updated biannually. Any change in residence, employment, insurance status (insurant or dependent) and relationships between insurant and dependents could incur an individual record in the registry. We exploit this dynamic nature to maximise the identification of possible family links by incorporating the entire registry records from 1995 to 2010.

Next we used the previous links to assemble pedigrees. We defined a family as a cluster of individuals who were related to each other by blood or by at least one common blood relative. Those without any parent identified were founders. We assembled a pedigree for each founder and then pedigrees were linked if they have common descendants. It should be noted that spouses without a child were not classified into one pedigree. Among 28,402,865 individuals contained in the registry of beneficiaries, 8,186,069 registered with themselves over a span of 15 years. The rest 20,216,796 were classified into 4,229,301 families. The mean family size was 4.8 persons and the largest families contained 144 members. There were 2 to 5 generations in these families.

Estimation of heritability and familial transmission using threshold liability model

We used the standard ACE model to examine the influences of additive genetic (A), common environmental factors shared by family members (C) and non-shared environmental factors (E) accounting for variance in a phenotype (P). This model can be expressed as:

$$\sigma_p^2 = \sigma_A^2 + \sigma_C^2 + \sigma_E^2$$

where $\sigma_p^2 = \text{total phenotypic variance}$; $\sigma_A^2 = \text{additive genetic variance}$; $\sigma_C = \text{shared environmental variance}$; and $\sigma_E^2 = \text{non-shared environmental variance}$.

Heritability was defined as the proportion of phenotypic variance that is attributable to genetic factors and can be expressed as $\frac{\sigma_A^2}{\sigma_p^2}$, and the familial transmission was expressed as $\frac{\sigma_A^2 + \sigma_C^2}{\sigma_p^2}$,

which is the sum of heritability and shared environmental variances. Familial transmission and heritability can be calculated using the polygenic liability model to calculate both measures.⁷⁻⁹ This model assumes a normally distributed liability of disease resulting from a large number of unspecified genes and environmental factors, each with small and additive influences. The liability of the affected individuals is greater than a critical threshold, the value of which can be determined with the information of the disease prevalence in the affected and the general population.

The familial transmission is the function of the difference of normal deviation of the threshold from the mean liability between individuals with affected relatives and the normal population. The original model assumes zero common environmental variance and therefore familial transmission equals heritability. To account for contributions of shared environmental factors to phenotypic variance, we used the spouse as a control, assuming that spouses share the family environment but have no close genetic similarity with blood family members. We restricted family history to first degree relatives and assumed an average of two siblings in a family. For full discussion of the model and the exact methods of familial transmission and heritability estimation please refer to a previous study relating to gout in Taiwan¹⁰ and the publication by Yang et al.¹¹ The familial transmission and heritability were calculated as:

$$\text{Familial transmission} = \frac{T_0 - T_1 \times \sqrt{1 - (T_0^2 - T_1^2) \times \left(1 - \frac{T_0}{i}\right)}}{a_R \times [i + T_1^2 \times (i - T_0)]}$$

$$\text{Heritability} = \frac{T_s - T_1 \times \sqrt{1 - (T_s^2 - T_1^2) \times \left(1 - \frac{T_s}{i}\right)}}{a_R \times [i + T_1^2 \times (i - T_s)]}$$

where $T_0 = \Phi^{-1}(1 - p)$; $T_s = \Phi^{-1}(1 - \text{spouse RR} \times p)$; $T_1 = \Phi^{-1}(1 - \text{sibling RR} \times p)$; $p =$ prevalence of SLE in the normal population); a_R : the additive genetic relationship between the relatives, for full sibling, $a_R = 0.5$; $i = z/p$; z , the height of the standard normal curve pertaining to SLE prevalence, and Φ , standard normal cumulative distribution function.

Socioeconomic data

Individual demographic and socioeconomic information were obtained from the registry of beneficiary. A place of residence for each individual was assigned as one of 369 towns or districts in Taiwan. The level of urbanisation for these 369 towns or districts was designated as either urban, suburban, or rural, based on 5 indices (population density, percentage of residents with college or higher education, percentage of residents >65 years of age, percentage of residents who were agriculture workers, and the number of physicians per

100,000 people).²³ Occupations were classified into 5 categories: (1) civil servants, teachers, military personnel and veterans; (2) non-manual workers and professionals; (3) manual workers; (4) other and (5) the dependents. Income levels were approximated based on the payroll of the employees and civil servants and the business income of employers. We categorised income levels into sex-specific income quintiles.

eTable. Code List for Sjögren Syndrome and Other Autoimmune Diseases

Disease	ICD-9 codes
Sjögren syndrome	710.2
Systemic lupus erythematosus	710.0
Rheumatoid arthritis	714.0
Systemic sclerosis	710.1
Idiopathic inflammatory myositis	710.3 or 710.4
Type 1 diabetes mellitus	250.01, 250.03, 250.11, 250.13, 250.21, 250.23, 250.31, 250.33, 250.41, 250.43, 250.51, 250.53, 250.61, 250.63, 250.71, 250.73, 250.81, 250.83, 250.91, 250.93
Multiple sclerosis	340
Myasthenia gravis	358
Inflammatory bowel diseases	555, 556.0 to 556.6, 556.8, 556.9
Vasculitis	443.1, 446.0, 446.1, 446.2, 446.4, 446.5, 446.7

References

1. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet*. 2005;366(9489):941-951.
2. Stefansdottir V, Johannsson OT, Skirton H, Tryggvadottir L, Tulinius H, Jonsson JJ. The use of genealogy databases for risk assessment in genetic health service: a systematic review. *J Community Genet*. 2012.
3. Cannon Albright LA. Utah family-based analysis: past, present and future. *Hum Hered*. 2008;65(4):209-220.
4. Thorsson B, Sigurdsson G, Gudnason V. Systematic family screening for familial hypercholesterolemia in Iceland. *Arterioscler Thromb Vasc Biol*. 2003;23(2):335-338.
5. Brewster DH, Fordyce A, Black RJ, Scottish Clinical G. Impact of a cancer registry-based genealogy service to support clinical genetics services. *Fam Cancer*. 2004;3(2):139-141.
6. Stefansdottir V, Johannsson OT, Skirton H, Tryggvadottir L, Tulinius H, Jonsson JJ. The use of genealogy databases for risk assessment in genetic health service: a systematic review. *J Community Genet*. 2013;4(1):1-7.
7. Falconer DS. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet*. 1967;31(1):1-20.
8. Reich T, James JW, Morris CA. The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. *Ann Hum Genet*. 1972;36(2):163-184.

9. Reich T, Rice J, Cloninger CR, Wette R, James J. The use of multiple thresholds and segregation analysis in analyzing the phenotypic heterogeneity of multifactorial traits. *Ann Hum Genet.* 1979;42(3):371-390.
10. Kuo CF, Grainge MJ, See LC, et al. Familial aggregation of gout and relative genetic and environmental contributions: a nationwide population study in Taiwan. *Ann Rheum Dis.* 2013.
11. Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. *Eur J Hum Genet.* 2010;18(9):1039-1043.