

# Clinical methods in psychiatric genetics

## I. Robustness of genetic marker investigative strategies

E. S. Gershon and L. R. Goldin

Clinical Neurogenetics Branch, National Institute of Mental Health, Bethesda, U.S.A.

---

**ABSTRACT** – Population stratification, secondary effects of illness or treatment, biological heterogeneity of a clinical syndrome, or complex biology underlying a syndrome (where only one component is measured) are conditions which may obscure the association of a genetic risk factor with a clinical syndrome. We consider several investigative strategies under each of these conditions. Only segregation-based paradigms are robust to genetic heterogeneity and population stratification. But secondary effects on the risk factor produced by illness or treatment require other strategies for their detection.

*Received November 1, 1985; accepted for publication March 1, 1986*

---

A genetic vulnerability marker for an illness (genetic risk factor) may be defined as a heritable trait, associated with a causative pathophysiologic factor in an inherited disease. By “heritable” trait, we mean either a trait determined by a known single locus polymorphism (such as an HLA type or an ABO blood group type) or another measured trait (either quantitative or qualitative) where the exact mode of genetic transmission is unknown but evidence for some genetic transmission exists from family or twin studies. Several investigative strategies exist to identify the association of genetic risk factors to diseases. In psychiatry and medicine, the diseases studied are often those where the inherited pathophysiologic mechanism is unknown, the specific gene(s) involved are not identified, and the mode of genetic transmission is not specified. In such an illness, the biology of the disease may present conditions to which a particular investigative strategy is not robust. By robustness we mean the ability of an investigative strategy to correctly detect or reject a putative genetic marker under various conditions.

Numerous investigative strategies have been proposed in psychiatry to test the validity of a genetic marker, and investigators must ask them-

selves how they might judge the validity of a particular strategy. We present in this paper a systematic consideration of the robustness of several research strategies under certain conditions which could lead to false positive or false negative conclusions on the validity of a marker. These conditions include population stratification, secondary effects of illness or treatment on a marker, biological heterogeneity of the clinical syndrome, and complex biology of the syndrome where only one component is measured. The first condition can lead to false positive conclusions about validity of a marker, the second to false positive or false negative conclusions, and the last two conditions can lead to false negative conclusions.

Population stratification exists when a gene frequency for the trait studied is higher in one subgroup of a population than another. Sickle cell anemia, for example, is more frequent in Black people. If our knowledge on this disorder were comparable to our knowledge of schizophrenia, it is easy to imagine a serious hypothesis that inherited anemias are heterogeneous, with the inheritance of skin pigment related to the inheritance of anemia in some cases. It is not difficult to also imagine an investigative strategy

Table 1  
Robustness of genetic marker investigative strategies under various conditions

	Population stratification	Secondary effect of illness or treatment	Genetic heterogeneity of illness	Complex biology with unmeasured components
Vulnerability markers (inherited risk factors)				
Segregation of marker in pedigrees	robust	not robust	robust	robust
Comparison of patients by family history	not robust	robust	not robust	robust
High risk study: well relatives of patients vs. controls	not robust	robust	not robust	not robust
Discordant vs. concordant MZ twins	not robust	robust	not robust	robust
Chromosomal linkage markers				
Linkage of marker to illness in pedigrees	robust	robust	not robust	robust

to test this hypothesis even if hemoglobin S were unknown, such as by examining skin color and anemia within families of patients or within the Black African population subgroup. The problem becomes more difficult if the correct population subgroup is unknown (in this example, if there were no obvious way to distinguish persons of black African descent). The ethnic associations of schizophrenia and affective disorders in European populations (1) do imply that other, as yet undiscovered but fortuitous associations with psychiatric disorders could exist.

The same problem of stratification could exist for inherited characteristics which do not have single major gene locus transmission.

Secondary effects of illness or treatment present another condition where a marker could be mistakenly accepted as valid. The reported ventricular enlargement in schizophrenia may be such a finding (2, 3). Theoretically, treatment could reverse a defect, thus leading to false negative conclusions on the validity of the defect as a marker.

Biological heterogeneity within a clinical syndrome can obscure a marker valid for only one of the several forms of illness. Diabetes mellitus was long thought to be inherited polygenically, with no single locus component (4). It is now known that a single locus is associated with a subgroup of patients with an insulin-dependent early onset clinical presentation, and that this

locus is associated with or identical to certain HLA antigens. It is readily apparent that the rarer the form of illness for which a marker is valid, the more difficult it will be to detect.

An illness with a complex biology, with several genetically independent risk factors, might be studied when only one of the risk factors is measurable. That is, several risk factors must be simultaneously present for illness to occur. This is not the same as heterogeneity, where each of several risk factors, in the absence of other factors, can lead to illness. In the complex biology situation, many persons, including relatives of known patients with a valid marker, can have the marker without being ill and thus it will not be detected as a risk factor.

For the remainder of this paper, we will consider the robustness of various investigative strategies under these four conditions (Table 1). The conditions are not mutually exclusive, but we will consider each one separately and disregard possible combinations of conditions.

## Investigative strategies

### Vulnerability markers (genetic risk factors)

*Segregation of marker in pedigrees.* Rieder & Gershon (5) designed a research strategy for

identification of vulnerability markers, based on segregation of a marker and illness within pedigrees, which is designed to be robust to these causes of false positive or negative conclusions. The criteria are straightforward:

1. Marker is associated with illness, in the population.
2. Marker is heritable.
3. Marker is state-independent (manifest whether or not illness is active).
4. Within families, marker and illness co-segregate.

We use the term "co-segregate" in the sense that among relatives who manifest the marker the prevalence of illness is higher than among relatives who do not. This is not the precise meaning of segregation in genetics, which refers to distribution in pedigrees of single gene locus characteristics, but it is a general analogy to single locus segregation which applies to multifactorial and other complex modes of inheritance.

Association with illness is defined to exist when the marker is associated with increased probability of illness. A simple patient-control difference will demonstrate this for a common finding in an illness. For a rare cause of illness, an estimation (usually indirect) of the rate of illness among persons with the marker, compared to the rate among persons without the marker, is needed. By this procedure, even a rare cause of a common illness can be detected (5).

The reason for requiring state-independence is that relatives may be studied before the age of onset, or between episodes. Theoretically, a genetic defect could manifest itself only during illness, but this would be exceedingly difficult to demonstrate as a marker since well relatives could not be evaluated for absence of the defect. In a remitting illness, one would expect the marker to be positive even when the persons is well, but some additional demonstration is needed that the observed status of the marker is not secondary to illness. By itself, this strategy will fail to correctly reject a putative marker which is actually a secondary effect of illness or treatment.

The requirement of segregation makes the

strategy robust in the presence of a population association, since all members of a family are part of the same population. This requirement applies only to pedigrees in which both illness and marker are present, in at least one individual. With this stipulation, genetic heterogeneity is resolved, since the subgroup of families selected will be relatively homogeneous with respect to biology of illness, if the marker is valid. This approach remains robust when only one of several contributing factors is identified as a putative marker, because selection of families is through a proband (presenting patient) in whom the marker and illness co-exist. That is, many persons in the population may have the putative marker but not the illness, because of the absence of other (unknown) contributing factors, and many patients may have the illness without the putative marker because of heterogeneity. By selecting families where the proband has the marker, the unknown contributing factors should also be present since the illness is present, heterogeneity has been bypassed, and relatives should not be ill without the marker. This point, on the robustness of segregation strategy defined here, has been reviewed at length elsewhere (6). Another advantage of this strategy is that it does not depend on a specific mode of genetic transmission of the illness and marker, although the statistical power will be greatest in straightforward single locus transmission with complete penetrance, and less in complex modes of transmission.

A practical issue often arises with quantitative measures, when there is overlap between patients and controls. Even if the only patients selected for study are those who are clearly different from controls, random variation among relatives may be expected, so that there will be ill relatives whose classification is uncertain, or in the control range. In addition, in a heterogeneous illness, a small proportion of relatives may be ill from causes other than marker-related. To test a quantitative marker, quantitative comparisons as developed by Rieder & Gershon (5) are used. For example: if, on a quantitative measurement, patients have higher mean values than controls, for a valid marker the key prediction is that ill relatives should have similar values to patients,

and well relatives should have similar values to controls. In some cases, it may be possible to dichotomize the data and measure relatives of "deviant" patients. Here the test is the same as described above. However, even on a qualitative marker, there may be individual relatives who are in the "wrong" range of values. Among well relatives there may be some who are biologically at risk but who have not manifested illness, and among ill relatives there may be phenocopies (persons ill from other causes).

The segregation approach differs from the other risk factor strategies in that it requires examination of ill relatives, to whom a genetic characteristic may or may not have been transmitted along with illness. In the other strategies considered here, relatives may not be examined at all (comparison of patients by family history), they may be examined when it is not known if they are affected (high risk strategy), or, when discordant MZ twins are compared, there is no opportunity for a genetic characteristic *not* to be shared.

*Comparison of patients by family history.* In this method, the usual comparison is between the frequency of a marker in patients with any ill relatives versus patients with no ill relatives. Even if the families are studied carefully, with all relatives examined directly, there will be an imprecision in the separation. For many modes of genetic transmission, and the family sizes commonly encountered, a substantial proportion of cases of inherited disease will occur with no family history. *Only* if there is also a substantial proportion of non-familial non-genetic cases, will the negative-family-history patients include fewer genetic cases than the positive-family-history, and useful quantitative comparisons might be made between the two types of patient. The necessity for invoking such a "best-case" assumption is a weakness of this strategy.

This strategy is not robust to population stratification. Consider the sickle-cell anemia example. If this were the only form of inherited anemia, one would find more Black than non-Black patients with positive family history, but it would be a mistake to consider race a physiologic genetic vulnerability factor.

The strategy is robust to secondary effect of illness or treatment, since this effect would be independent of morbid risk. Patients with and without family histories would not be differentiated on such a putative marker, so a false positive conclusion would not be reached.

Heterogeneity would obscure the validity of a marker, in this investigative strategy because it would differentially reduce the proportion of marker positive patients with positive-family-history. In one special case however, a derivative of the family history method can resolve heterogeneity: when a specific mode of genetic transmission can be identified in a subgroup of patients, by clinical or pathological traits associated with a particular pedigree structure (7).

*High risk studies (well relatives of patients vs. controls).* This design can be executed in two ways, by ascertaining well relatives of known patients for comparison with controls on presence of a marker, or, as proposed by Buchsbaum et al. (8), by screening normals, dividing them into marker present or absent, and comparing family history of each group. For consideration of robustness, the same arguments apply to both methods.

Population stratification will lead to falsely positive conclusions about a proposed marker, because prevalence of illness is higher in the marked sub-population, as described above in the discussion of race as a marker for sickle cell anemia. Genetic heterogeneity and complex biology can each confound this strategy, as discussed elsewhere at length by Cloninger et al. (6). Briefly, in the presence of genetic heterogeneity the frequency of the marker in relatives of patients is reduced, and the morbid risk in controls without the marker is increased. Thus, either of the comparisons that can be made will show a smaller or vanishing difference between groups. Complex biology can lead to many more normals than patients having any single component of risk, which makes the Buchsbaum strategy non-robust to this condition.

The advantage of high risk strategies is that they are robust to secondary effects of illness or treatment, because the individuals examined on their marker status have never been ill or treated.

A powerful variant of the high risk method is to combine it with segregation. One can select patients who have a marker, and examine their offspring at a young age. The predicted outcome is that offspring with the marker are much more likely to develop illness than offspring without. Of course, this strategy can require a long time to be implemented. This time is reduced if offspring are studied just before they enter the age of risk, and are followed only long enough for a statistically usable proportion to develop illness.

*Discordant vs. concordant monozygotic twins.* Monozygotic (MZ) twins are genetically identical, and so do not afford any opportunity for parental genes to segregate independently between them. The assumption behind using the concordant-discordant comparison to detect a genetic marker, is that in the discordant pairs neither twin is genetically vulnerable but in concordant pairs both are, so a genetic marker should be found in concordant but not discordant pairs. The assumption does not allow for the possibilities that discordance can be produced by failure of a genetic tendency to express itself (variable penetrance in the case of a single locus trait), or that concordance will occur when non-genetic illness factors are shared by identical twins (such as viral infection *in utero*).

Even if these possibilities are disregarded and the assumption is accepted, genetic heterogeneity will diminish the observable differences between the two types of MZ twin pairs, by reducing the number of concordant pairs who will show a marker. For the same reason, however, the complex biology condition has no effect on MZ twins; they share all genes, including genes which are components of vulnerability unrelated to the marker studied. Population stratification leads to false positive conclusions in this strategy, because there is no opportunity for independent assortment of a disease and population marker.

If a putative marker is actually produced by illness or treatment, it will not be manifested in the discordant twin who has never been ill. For this condition, the MZ twin strategy is robust, and can usefully test whether a finding is secondary to illness.

## Chromosomal linkage markers

This is a classic genetic strategy, which is discussed here for comparison with genetic vulnerability marker [risk factor] strategies. An important difference between a linkage marker and a risk factor marker is that linkage is necessarily an event that occurs in relation to a single genetic locus, whereas a risk factor may be polygenetically determined or be the result of complex interactions between several loci.

A second difference between linkage and risk factor strategies is that the risk factor is necessarily associated with the physiologic process that causes the disease, whereas in linkage equilibrium the marker is not associated with the disease in the population. (Within families, however, one allele occupying the marker locus will be consistently associated with illness.) When there is a population association due to linkage disequilibrium, the linkage analysis can still be performed, with appropriate modifications to allow for association (9, 10).

Secondary effects of illness or treatment would not be expected to affect a marker locus.

Genetic heterogeneity reduces the power of linkage analysis, since the analysis generally considers a cumulative result over all pedigrees. In some instances, it is possible to detect linkage despite the presence of heterogeneity (11), but the greater the number of forms within the heterogeneity, the less likely that a particular linkage will be detectable.

## Discussion

No one investigative strategy for vulnerability markers is robust to all the conditions we have considered. Among strategies discussed here, the segregation paradigm of Rieder & Gershon (5) is the only one which is robust to genetic heterogeneity. It would not be prudent to reject a vulnerability marker until the possibility of heterogeneity is eliminated by such a paradigm.

Population stratification presents another major problem for validation of a vulnerability marker. Only strategies that examine the assortment of the marker in pedigrees with illness appear to be robust to this condition. For a

single locus marker, pedigree linkage analysis does this best. However, when the inheritance of the marker does not follow Mendelian inheritance, a strategy such as that of Rieder & Gershon (5) must be followed. It appears, then, that a genetic vulnerability marker can not be successfully established without investigating pedigrees of patients, but this has rarely been done in psychiatric disorders, except in the investigation of linkage markers. By itself, however, the study of pedigrees, with ill and well relatives compared, is not robust to false positive or negative conclusions based on secondary effects of illness or treatment. To rule this out, one of the strategies that examines individuals at risk who have not developed illness seems most appropriate, such as the study of offspring of known patients as they enter and pass through the age of risk.

## References

1. Torrey E F, McGuire M, O'Hare A, Walsh D, Spellman M P. Endemic psychosis in Western Ireland. *Am J Psychiatry* 1984;141:966-970.
2. Reveley A M, Reveley M A, Clifford C A, Murphy R M. Cerebral ventricular size in twins discordant for schizophrenia. *Lancet* 1984;1:540-541.
3. DeLisi L E, Goldin L R, Hamovit J R, Maxwell M E, Kurtz D, Gershon E S. A family study of the association of increased ventricular size with schizophrenia. *Arch Gen Psychiatry* 1986;43:148-156.
4. Childs B. Medical genetics. *Neurosci Res Prog Bull* 1976;14:13-20.
5. Rieder R O, Gershon E S. Genetic strategies in biological psychiatry *Arch Gen Psychiatry* 1978;35:866-873.
6. Cloninger C R, Lewis C, Rice J, Reich T. Strategies for resolution of biological and cultural inheritance. In: Gershon E S, Matthyse S, Breakefield X O, Ciaranello R D, eds. Genetic research strategies in psychobiology and psychiatry. Pacific Grove CA: Boxwood Press, 1981: 319-332.
7. Matthyse S, Kling M. Neurochemical and genetic bases of psychopathology: future directions. *Behav Genet* 1982;12:101-109.
8. Buchsbaum M S, Coursey R D, Murphy D L. The biochemical high-risk paradigm: behavioral and familial correlates of low platelet monoamine oxidase activity. *Science* 1976;194:339.
9. Gershon E S, Matthyse S. X-linkage: ascertainment through doubly ill probands. *J Psychiatr Res* 1977;13:161-168.
10. Morton N E, Lalouel J M. Resolution of linkage for irregular phenotype systems. *Hum Hered* 1981;31:3-7.
11. Morton, N E. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 1956;9:55-75.

## Address

Elliot S. Gershon, M.D.  
 Clinical Neurogenetics Branch  
 National Institute of Mental Health  
 Building 10, Room 3N218  
 9000 Rockville Pike  
 Bethesda, Maryland 20892  
 U.S.A.