Published in final edited form as:

Stat Med. 2002 July 15; 21(13): 1887–1897.

Concordance of disease form in kindreds ascertained through affected individuals

Melodie Winawer^{1,2}, Ruth Ottman^{1,3}, and Daniel Rabinowitz^{4,*†}

1 G.H. Sergievsky Center and Mailman School of Public Health (Epidemiology Division), Columbia University, New York, U.S.A.

2 Department of Neurology, Columbia University, New York, U.S.A.

3 Epidemiology of Brain Disorders Research Department, New York State Psychiatric Institute, New York, U.S.A.

4 Department of Statistics, Columbia University, New York, U.S.A.

SUMMARY

When designing or conducting genetic epidemiological studies of a disease with several distinct forms, it is useful to know whether susceptibilities to the different forms are conferred by different genes or whether there are genes that confer susceptibility to multiple forms. A natural approach to exploring these issues is to examine how the disease forms cluster in kindreds. When inclusion in the study is based on the affection status of multiple relatives, however, distorted patterns of familial clustering of disease form can be evident. The purpose here is to present statistical methods for adjusting for this distortion. In particular, approaches to testing two null hypotheses are presented: a null hypothesis that corresponds to all genes acting in the same way on the relative risk of the different disease forms, and a null hypothesis that corresponds to each gene conferring susceptibility to distinct disease forms. The approaches are illustrated through an application to the generalized and localization-related forms of epilepsy.

INTRODUCTION

Some diseases with a genetic influence have distinct forms. For example, among the different forms of epilepsy are the localization-related and generalized forms, and individuals usually suffer from only one form but occasionally suffer from both. For diseases such as epilepsy that have distinct forms, it may be that some genes confer susceptibility to only one form. Conversely, there may be genes that confer susceptibility to multiple forms.

If a disease has multiple forms, then when designing and conducting a genetic epidemiological study to locate genes involved in the aetiology of the disease or a study to examine the influence of candidate mutations on the disease, it is useful to know whether susceptibilities to the different forms are conferred by different genes or whether there are genes that confer a susceptibility to more than one form. If each gene contributes to a separate form, then the different forms of disease should be analysed separately, but if there are genes that contribute to more than one form, then it might be appropriate to pool data across disease forms.

*Correspondence to: Daniel Rabinowitz, Department of Statistics, Mail Code 4403, Columbia University, New York, NY 10027, U.S.A..

†E-mail: dan@stat.columbia.edu

Contract/grant sponsor: NIH; contract/grant number: GM5597, NS07153, NS20656

Contract/grant sponsor: Taub Institute for Research on Alzheimer's Disease and the Aging Brain

A natural approach to exploring whether there are genes that confer susceptibility to distinct forms and whether there are genes that confer shared susceptibility is to examine how disease forms cluster in kindreds. If kindreds tend to be concordant for disease form, that is, if members of the same kindred tend to manifest the same disease form, one might infer that each form of disease is influenced by a different gene. If kindreds tend to be discordant for disease form, one might infer that some genes confer susceptibility to several forms.

An issue encountered when examining clustering of disease form is that study sample inclusion criteria may involve the affection status of related individuals. For example, ascertainment might be initiated through affected individuals coming to the attention of researchers. Subsequently, the researchers might examine the available relatives of the individual and include the individual's kindred only if, for example, there were at least two affected first degree relatives.

Inclusion criteria that stipulate that a kindred must contain multiple affected individuals result in samples that are particularly suitable for linkage analyses. Such inclusion criteria, however, can result in distorted patterns of clustering of disease form. Examination of patterns of clustering of disease form in such samples requires, therefore, methods that account for the inclusion criteria.

Distortion induced by inclusion criteria results, for example, when there are many susceptibility genes, each with very low penetrance. In this case, it might be difficult for any one gene to give rise to several affected individuals in the same kindred, and so, in kindreds chosen for having many affected members, different genes would most likely have been responsible for the different manifestations of the disease. Even if every gene conferred susceptibility to only one form of disease, ascertainment schemes that required kindreds to contain several affected individuals would therefore result in many kindreds in the study sample in which more than one disease form were represented.

The purpose here is to present methods for examining clustering of disease form in study samples composed of kindreds ascertained through inclusion criteria that involve multiple affected individuals. Approaches to testing two null hypotheses are presented. The first null hypothesis corresponds to the situation where the effect of all susceptibility genes is on the probability of disease, but not on the relative probabilities of the disease forms. This null hypothesis is formalized as disease forms in affected kindred members following multinomial distributions with common relative frequency parameters. The second null hypothesis corresponds to the situation where no gene confers susceptibility to more than one disease form. This null hypothesis is formalized as independence of the occurrence of the different disease forms in kindred members. These two hypotheses represent different extremes; rejecting one is not incompatible with rejecting the other.

The methods developed here are predicated on some assumptions about the ascertainment process. Ascertainment is assumed to be initiated through affected individuals who come to the attention of investigators through their affection status. (Such individuals will here be termed probands if they and their relatives are ultimately included in the study sample. Although this usage differs from the convention of denoting as probands only those individuals who, independent of all other individuals, causes the kindred to be included in the sample, it helps in making the distinction between the individuals who initially come to the attention of researchers, and the several kindred members who are jointly responsible for the inclusion of the kindred.) It is further assumed that kindreds are screened according to inclusion criteria that specify that certain numbers of certain types of relatives (for example, at least two first-degree relatives) must be affected. No assumptions are made about the influence of an individual's disease form on the probability that the individual will come to the attention of

investigators. It is assumed, however, that, given the disease form of the individual, the probability of coming to the attention of researchers is conditionally independent of presence or absence of disease and of disease form, in the individual's relatives. These assumptions preclude, for example, individuals seeking inclusion in the study because they have affected relatives, or because their disease form is concordant with that of their relatives. The assumptions may be relaxed, when testing the first null hypothesis, to allow the probability that an individual comes to the attention of researchers to depend on the affection status of relatives, as long as the dependence is only through the presence or absence of disease, and not on the affected relatives's disease forms.

The methods proposed for the first null hypothesis rest on a permutation test approach. See, for example, Cox and Hinkley [1]. A test statistic that reflects clustering of disease form in kindreds, such as the number of kindreds concordant for disease form, may be compared to its permutation distribution. Here, by concordant it is meant that all subjects in the kindred manifest the same disease form. The methods proposed for the second null hypothesis involve defining several strata each corresponding to a different disease form. Under the null hypothesis, within each disease form's stratum, despite the ascertainment scheme used to obtain the original sample, in related individuals in a stratum, the other forms of the disease are distributed as in related individuals in the general population. To test the null hypothesis, therefore, stratum-specific occurrence of these other disease forms may be compared to historical data or known population rates.

The approaches are illustrated through an application to the localization-related and generalized forms of epilepsy. The co-occurrence in kindreds of these two forms of epilepsy has been investigated previously. See, for example, Ottman *et al.* [2,3], Jain *et al.* [4], Callenbach *et al.* [5] and the Italian League Against Epilepsy Genetic Collaborative Group [6]. The analysis presented here provides strong evidence against the first null hypothesis; it appears that some factors confer relatively greater susceptibility to one or the other of the disease forms than do other factors. Although there is a substantial number of kindreds in the data set in which both disease forms are represented, the approach developed here for the second type of null hypothesis indicates that the discordant kindreds are potentially an artefact of the ascertainment scheme; there appears to be no significant evidence in these data against the hypothesis that each gene confers susceptibility to a distinct disease form.

TESTING FOR CONCORDANCE

In this section, the concern is with testing the null hypothesis that the multinomial frequency parameters of the conditional distribution of disease form, given the number of affected members, are homogeneous across the kindreds. This null hypothesis does not preclude different genes conferring different susceptibilities to the disease as a whole, but it specifies that the conditional frequencies of disease form are invariant across susceptibility genes. Familial clustering of disease forms, as measured, for example, by concordance of disease form within kindreds, therefore, is evidence against this null hypothesis. In what follows, a permutation test approach is described and the mean and variance under the permutation distribution of the number of kindreds concordant for disease form are presented.

Under the assumptions about the ascertainment scheme described in the first section, the distribution of the number of affected individuals in kindreds is completely unspecified. The relative frequencies of the different forms of disease are also not specified, and, because the probability of becoming a proband is not assumed independent of disease form, the frequencies of the different forms of disease in the probands may differ from the frequencies in the other kindred members. However, the relative frequencies of the disease forms in affected subjects other than probands are, under the null hypothesis, invariant across kindreds. This suggests

that correct type I error rates could be obtained by computing conditional p -values under the permutation distribution that keeps fixed the disease form of the probands, keeps fixed the numbers of affected subjects in each kindred, keeps fixed the pooled numbers of subjects with each form, but which otherwise assigns disease form without regard to kindred membership. This permutation distribution may be generated by fixing the probands's disease forms at their observed values, forming the pooled sample of disease forms in the remaining individuals, and then re-assigning those disease forms by sampling randomly, without replacement, from the pooled sample. An SAS macro that implements this algorithm is available upon request from the corresponding author.

In moderate and large samples, it may be reasonable to compute p -values according to asymptotic approximations to the permutation distribution. This section concludes with an illustration of how asymptotic approximations might be computed. The test statistic used in the illustration is the count of the number of kindreds that are concordant for disease form.

For increasing numbers of kindreds in the data set, the distribution of the number of kindreds concordant for disease form, after normalization by the expectation and standard deviation, is approximately that of a standard normal variable. It follows that, in order to compute approximate p -values, it suffices to compute the expectation and variance of the number.

Let n_i denote the number of affected individuals, excluding the proband, in the i th kindred. Let t_i denote the disease form of the proband in the i th kindred. Let n_t denote the total number of individuals, excluding probands, with disease form t , so that n_{ti} denotes the number of individuals in the sample, excluding probands, who have the same disease form as does the proband in the i th kindred. Let n denote the total number of affected individuals in the sample, excluding the probands. Then, the expectation of the number of kindreds concordant for disease form is given by the sum over kindreds of μ_i , where

$$\mu_i = \sum_{j=0}^{n_i-1} \left(\frac{n_{t_i} - j}{n - j} \right)$$

The variance is given by the double sum over kindreds

$$\sum_i \sum_j \begin{cases} \mu_i - \mu_i^2 & \text{if } i = j \\ \prod_{k=0}^{n_i+n_j-1} \left(\frac{n_{t_i} - j}{n - j} \right) - \mu_i \mu_j & \text{if } i \neq j \text{ and } t_i = t_j \\ \prod_{k=0}^{n_i-1} \left(\frac{n_{t_i} - k}{n - k} \right) \times \prod_{k=0}^{n_j-1} \left(\frac{n_{t_j} - k}{n - n_i - k} \right) - \mu_i \mu_j & \text{if } i \neq j \text{ and } t_i \neq t_j \end{cases}$$

A program for the Windows operating system that implements these methods is available upon request from the corresponding author.

TESTING FOR DISCORDANCE

This section is concerned with testing the null hypothesis that any factor that confers susceptibility to any given disease form does so independently of any factors that confer susceptibility to the other disease forms. This null hypothesis is a proxy for the null hypothesis that no factor confers susceptibility to more than one disease form. The two hypotheses are equivalent, however, only when the different risk factors occur independently. Without observing the risk factors themselves, it is not possible to distinguish the co-occurrence of two

factors each of which raises the risk for a different disease form from a single risk factor that increases the risk for both disease forms.

The null hypothesis does not specify the distribution of the occurrence of any disease form. However, it does imply that in kindreds identified through the presence of a particular disease form, the occurrence of the other disease forms in that kindred corresponds to that of the general population. To test the null hypothesis in population-based samples, therefore, one could, for example, examine the incidence of one disease form in kindreds in which another disease form occurs; an excess relative to population rates would indicate aetiologies that conferred susceptibility simultaneously to both disease forms. This approach is not possible in samples ascertained through affected relatives, however, as in such samples the incidence of all disease forms could be artificially inflated. In what follows is a description of an approach to identifying subsets of the study sample kindreds in which the occurrence of particular disease forms may be compared to that of the general population.

The approach advocated here involves stratifying by proband disease form, and, in each stratum, examining the presence of the other forms of disease. To avoid bias induced by the ascertainment scheme, an additional requirement is placed on individuals before they are included in a stratum: individuals included in the strata must satisfy the inclusion criteria through the members affected with the proband's disease form. That is, individuals are only included in a stratum if they would have been included in a study sample ascertained purely through a single disease form. For example, if only first-degree relatives of probands were considered for inclusion, and the inclusion criterion specified the existence of at least two affected first-degree relatives of the proband, then inclusion of a first-degree relative of a proband in a stratum corresponding to a disease form $\mathcal{A}$ would require that the proband's disease form be $\mathcal{A}$ and that the proband had at least two first-degree relatives not simply affected with the disease, but affected with the form $\mathcal{A}$. Note that in some ascertainment schemes it may be that some but not all members of a kindred are included in a stratum. For example, if first-degree relatives of previously included affected individuals are systematically sequentially examined, only subjects whose inclusion could be traced back through a sequence of relatives affected with disease form $\mathcal{A}$ would be included in the stratum corresponding to the disease form $\mathcal{A}$.

The validity of this approach to constructing strata follows from the independence, under the null hypothesis, of the different disease forms. Although kindreds are originally ascertained through the presence of any affected individuals, the related individuals in the strata may be viewed as kindreds ascertained solely through the presence of a particular disease form. For this reason, individuals in kindreds whose probands have multiple forms of the disease are not included in any stratum. The probands might have come to the attention of researchers through carrying both forms of the disease, and their relatives might therefore be at greater risk, even under the null hypothesis, of carrying both forms of the disease. Under the null hypothesis, therefore, in the stratum corresponding to a disease form $\mathcal{A}$, there should be no excess of the disease forms other than $\mathcal{A}$, and a relatively high incidence of these other disease forms is evidence against the null hypothesis.

How the occurrence of disease forms other than $\mathcal{A}$ ought to be compared to the distribution of disease forms in the population must depend on the available information. In one extreme, there could be available information from population-based samples that could be used to determine age-specific population hazard rates in strata defined by demographic variables. In this case, age-at-onset information in the members of the kindreds included in the strata could be used to compute hazard ratio estimates. In these computations, the minimum of age at ascertainment, age at death, and age at loss to follow-up would be treated as a censoring variable. (In subjects experiencing onset of more than one form of disease, if only the age at

onset of the first occurring disease form is recorded, and if the first occurring form is the proband's disease form, then the definition of the censoring time should be augmented to include the age at the first occurring onset.) In the other extreme, there might be no population data and only prevalence information in the study subjects. In this case, one could only compare the prevalence data for the disease forms other than $\mathcal{A}$ to generally accepted estimates of the population prevalence. In any case, when making formal statistical comparisons, the clustering by kindreds must be taken into account when computing standard errors.

This section concludes with a pseudo-likelihood approach that may be used for inference about the ratio of the hazard for onset of disease forms other than $\mathcal{A}$ to their hazard in the general population. Let Z_{ij} denote demographic data on the j th member of the i th kindred in the stratum defined in terms of $\mathcal{A}$, and let T_{ij} denote the age at onset of any disease form other than $\mathcal{A}$. Let C_{ij} denote a corresponding censoring time, the minimum of the subject's age at ascertainment, death or loss to follow-up. Let Δ_{ij} denote the indicator that onset of a form other than $\mathcal{A}$ is observed in the subject; that is, Δ_{ij} is the censoring indicator $1_{\{T_{ij} \leq C_{ij}\}}$. Let $\lambda(t|z)$ denote the conditional hazard for onset of disease forms other than $\mathcal{A}$ at age t given demographic variables z , and let $\Lambda(t|z)$ denote the corresponding cumulative hazard function. Let θ denote the log hazard ratio associated with stratum membership. That is, the hazard for onset of disease forms other than $\mathcal{A}$ in the stratum defined in terms of $\mathcal{A}$, given demographic variables z , is $\exp(\theta)\lambda(t|z)$.

The likelihood for θ depends on the joint distribution of onset in related individuals. This joint distribution is generally unknown, and might be difficult to model. However, a generalized estimating equation may be developed as the score equation from a likelihood that treats all subjects as independent. This approach side-steps modelling of the joint distribution. See, for example, Zeger and Liang [7]. This likelihood is the product over the individuals, excluding the probands, in the stratum

$$\prod_{ij} \{\lambda(T_{ij} | Z_{ij})^{\Delta_{ij}} \exp(-\theta)^{\Delta_{ij}} \Lambda(\min(T_{ij}, C_{ij}) | Z_{ij})^{\exp(\theta)}\}$$

The corresponding estimating equation is

$$\sum_{ij} \Delta_{ij} - \exp(\theta) \log[\Lambda(\min(T_{ij}, C_{ij}) | Z_{ij})] = 0$$

Let $\hat{\theta}$ denote the solution to the estimating equation. Then, an estimator of the variance of $\hat{\theta}$ that reflects within-kindred correlations is

$$\frac{\sum_i (\sum_j \Delta_{ij} - \exp(\hat{\theta}) \log[\Lambda(\min(T_{ij}, C_{ij}) | Z_{ij})])^2}{(\sum_i \sum_j \exp(\hat{\theta}) \log[\Lambda(\min(T_{ij}, C_{ij}) | Z_{ij})])^2}$$

See Wei *et al.* [8] for a development of this approach to computing standard errors.

This standard error computation may be used in hypothesis tests when the number of events $\sum_{ij} \Delta_{ij}$ is not small. However, when the disease frequency is small and the sample size is only moderate, there might be only a few events in any of the strata. In this case, an alternative approach is required. A conservative approach to computing p -values is to count at most one event per kindred, and to approximate the distribution of the count as if the study subjects were not related. Such an approximation is given by treating the count as a Poisson variable with a parameter that may be estimated by the sum of the cumulative hazards, $\sum_{ij} \Lambda(C_{ij})$.

EXAMPLE

In this section, the results of the application of the methods developed here to data on idiopathic/cryptogenic generalized and idiopathic/cryptogenic localization-related epilepsy are presented. (Idiopathic/cryptogenic refers to disease whose cause cannot be attributed to known external factors such as head injury.) Generalized seizures are those in which the first clinical changes indicate initial involvement of both cerebral hemispheres, and partial seizures are those in which the first clinical and electroencephalographic changes indicate activation of a system of neurons limited to part of one cerebral hemisphere, see the Commission of Classification and Terminology of the International League Against Epilepsy [9]. The purpose here is not to present substantive results, but simply to provide an illustration of the methodology proposed here. A more thorough analysis of an augmented data set may be found in Winawer *et al.* (in preparation).

The kindreds used in this example were obtained through a two-stage sampling design. In the first stage, kindreds came to the attention of researchers through affected volunteers, and the affection status of these potential probands' first-degree relatives was ascertained. In the second stage, the kindreds in which the volunteers had at least two affected first-degree relatives were identified, and, sequentially in each kindred, affection status was ascertained in the first-degree relatives of previously identified affected individuals. Finally, when possible, disease form (generalized or localization-related) was determined in the affected subjects obtained from both stages.

There were 25 kindreds included in the analysis; eleven had a proband with the generalized form, thirteen had a proband with the localization form and one had a proband with both. Among the 25 kindreds, there were three, thirteen, six, one, one and one, who had, respectively, two, three, four, five, eight and ten affected members with identified disease forms. Excluding the probands, there were 30 subjects with the generalized form, 34 with the localization-related form and two with both. Despite the inclusion criterion of three or more affected first-degree relatives (the proband and two affected first-degree relatives), there are kindreds with but two affected individuals. This occurred because in some of the relatives epilepsy could not be classified as generalized or location-related, see Table I.

For testing the first kind of null hypothesis, if having both forms of disease is treated as a separate form, under the permutation distribution, the expected number of concordant kindreds is 5.66 and the standard deviation is 2.51. The number of concordant kindreds was fourteen, so the Z-statistic was 3.32; the observed number of concordant kindreds is significantly larger than what would be expected under the first null hypothesis. The results are essentially the same if subjects with both forms of the disease are removed from the sample, and if having both forms of the disease is treated as concordant with having any form.

For testing the second null hypothesis, individuals were included in the stratum corresponding to one or the other of the disease forms if the associated proband had the disease form (but not both forms) and if there were a total two or more additional first-degree relatives who had either both forms of the disease or the same disease form as the proband. There were five kindreds with a total of 18 affected subjects (excluding marry-ins) with no disease or an identified disease type included in the stratum corresponding to the generalized form of the disease, and six kindreds with a total of 17 such subjects in the stratum corresponding to the localization-related form.

None of the kindreds included in the stratum defined in terms of the generalized form of epilepsy had any members with the localization-related form of the disease, nor did any of the kindreds in the stratum defined in terms of localization-related form have any subject with the generalized form. Despite the presence of eleven kindreds in the original sample of 25 that

were not concordant for disease type, after applying the inclusion criteria for stratum membership, there remains no evidence, in these data, against the hypothesis that all factors confer susceptibility to distinct disease forms. (Although the issue is moot for these data, in principle, for estimating the hazard ratio θ , the population hazard rate could be obtained from the report in Hauser *et al.* [10] of the results of a population-based sample.)

It should be noted that the power for detecting factors that confer susceptibility to more than one disease form may be quite weak in this sample. The number of relatives with the proband's disease form above what is needed for inclusion in a stratum provides a rough estimate of the order of magnitude of what the expected number with a different disease form would be under the null hypothesis. The number of such relatives from the two strata is only two. In this light, finding zero is not exceptional.

Finally, it is important to note the limitations that family-based sampling schemes appear to confer on the statistical analysis of concordance of disease form. The methods developed here are applicable only to two fairly extreme null hypotheses, and rejection of these null hypotheses cannot be unequivocally interpreted in terms of the genetic phenomena that are of primary interest.

DISCUSSION

The approaches presented here to examining clustering of disease form in kindreds have focused on testing two extremes. One extreme corresponds to all susceptibility genes acting in the same fashion with respect to disease form. The other corresponds to no susceptibility gene acting on more than one disease form. In general, one might expect a mixture of different kinds of genetic effects. Thus, while it might be of interest to test the two extremes, it would also be of interest to make more general inferences. However, beyond the two extremes considered here, there do not appear to be meaningful hypotheses about the nature of genetic effects that, without possibly untestable assumptions about penetrances and frequencies, lead to testable constraints on the distribution of disease forms. That is, the focus here is on testing for familial aggregation of disease form, rather than on a full segregation analysis. Nevertheless, the distributions estimated from the strata defined in the previous section could, in large samples, provide exploratory information about the range of possible genetic effects.

The methods developed here are useful for adjusting for effects of the ascertainment scheme that may obscure patterns of clustering of disease form. Use of the methods in kindreds ascertained through the disease status of individuals, however, does not guarantee that effects will be revealed; in some settings, the ascertainment scheme itself results in kindreds in which, after adjustment, there is no information about effects that may exist.

The assumptions that underlie the approach to adjusting for effects of the ascertainment scheme are predicated on the ascertainment process being carried out in two stages. In the first stage, individuals come to the attention of researchers through their affections status. In the second stage, inclusion criteria are applied to the individuals's family members. The assumptions that underlie the approach developed here specify that in the first stage, for the first null hypothesis, conditionally given an individual's disease form, the disease form of the individual's affected relatives should not influence the probability that the individual comes to the attention of researchers. For the second null hypothesis, it is additionally specified that the affection status of the individuals's relatives should also not influence the probability that the individual comes to the attention of researchers. The conditioning approach used for the first null hypothesis and the stratification approach used for the second null hypothesis are designed to adjust for the inclusion criteria applied in the second stage; the validity of the approaches rests on the assumptions on the first stage.

For example, if there is one gene that has high penetrance but leads to only one kind of disease, and there is another gene that has low penetrance, but confers susceptibility to all forms of the disease, in kindreds with many affected individuals, it is more likely that the cause of disease is the former gene. If the inclusion requirements call for many affected individuals, then most kindreds would be concordant for the disease form associated with the former gene and there would be little evidence of the shared susceptibility gene. While the methods developed here protect the researcher from erroneous conclusions resulting from not taking into account the ascertainment scheme, population based samples of kindreds may have better power to detect evidence for the different possible kinds of genetic mechanisms. In the simple case, a population based sample would result not just in kindreds with many affected individuals all concordant for disease form, but also discordant kindreds with few affected individuals.

It has been implicit throughout much of the previous sections that the causes of the disease are genetic. However, there may be both gene–environment interactions and solely environmental factors that confer susceptibility to the disease. The results of analyses should therefore be interpreted with care; inferences resulting from the analyses might pertain to the influence of non-genetic rather than genetic influences on the different forms of the disease.

Similarly, the formalizations of the null hypotheses in terms of independence of disease forms and in terms of homogeneous multinomial frequency parameters can correspond in certain anomalous circumstances to genetic effects that would be thought of as belonging with the alternative. For example, even if all susceptibility genes were specific for distinct disease forms, it is possible that through assortative mating of individuals carrying the different genes, the occurrence of the different disease forms in kindreds might be homogeneous. It is, of course, not possible to disentangle such circumstances from genetic effects without genotypic data.

Finally, the methods developed here detect the presence of aetiological factors that both aggregate in families, and that influence the relative frequency of the different disease forms in affected individuals. In some cases, some such factors might be available to the analyst. For example, for diseases where different forms have different onset distributions, clustering of age at ascertainment could result in evidence against the first null hypotheses (testing for concordance). When the ages of the subjects enrolled in the study are recorded, it might be of interest to examine whether clustering of age at ascertainment could explain apparent familial aggregation of disease form. A detailed discussion of this issue is beyond the scope of this manuscript, but it might be noted that adjusting for such factors could involve either modelling or stratification. In particular, in applying the permutation approach to testing for concordance, subjects might only be permuted within strata defined by age at ascertainment. (This strategy was implemented for the data of the illustrative example; the results were not qualitatively different from the results of the analysis with no stratification for strata defined both in terms of age in years or several coarser stratifications.)

Acknowledgements

This work was supported in part by NIH grants GM5597, NS07153 and NS20656 and by the Taub Institute for Research on Alzheimer's Disease and the Aging Brain. The authors appreciate Andrew Gelman's thoughtful comments on a preliminary draft.

References

1. Cox, DR.; Hinkley, DV. Theoretical Statistics. Halsted Press: New York; 1974.
2. Ottman R, Annegers JF, Hauser WA, Kurland LT. *Epilepsia* 1989;30:157–161. [PubMed: 2494041]
3. Ottman R, Lee JH, Hauser WA, Risch N. Are generalized and localization-related epilepsies genetically distinct? *Archives of Neurology* 1998;55:339–344. [PubMed: 9520007]

4. Jain S, Padma MV, Puri A, Jyoti, Maheshwari MC. Occurrence of epilepsies in family members of Indian probands with different epileptic syndromes. *Epilepsia* 1997;38:237–244. [PubMed: 9048678]
5. Callenbach PMC, Geerts AT, Arts WFM, van Donselaar CA, Peters ACB, Stroink H, Brouwer OF. Familial occurrence of epilepsy in children with newly diagnosed multiple seizures: Dutch study of epilepsy in childhood. *Epilepsia* 1998;39:331–336. [PubMed: 9578054]
6. Italian League Against Epilepsy Genetic Collaborative Group. Concordance of clinical forms of epilepsy in families with several affected members. *Epilepsia* 1993;34:819–826. [PubMed: 8404731]
7. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–130. [PubMed: 3719049]
8. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989;84:65–1073.
9. Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia* 1989;30:389–399. [PubMed: 2502382]
10. Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935–1984. *Epilepsia* 1993;34:468–543.

Table I

Probands's forms and numbers of subjects affected with the localized, generalized or both forms in the 25 kindreds included in the illustrative analysis.

Probands	Non-probands		
	Generalized	Localized	Both
Generalized	1	1	0
Localized	0	2	0
Generalized	0	2	0
Generalized	2	1	0
Generalized	2	0	0
Generalized	1	1	0
Localized	0	9	0
Generalized	2	0	0
Localized	3	1	0
Localized	0	2	0
Localized	1	1	1
Generalized	2	0	0
Localized	1	0	0
Generalized	3	0	0
Localized	0	1	0
Localized	0	2	0
Generalized	1	2	0
Localized	1	2	0
Both	5	0	2
Localized	2	0	0
Localized	0	2	0
Generalized	3	0	0
Generalized	0	1	0
Localized	0	2	0
Localized	0	2	0