Am. J. Hum. Genet. 62:1212–1227, 1998

Statistical Evaluation of Age-at-Onset Anticipation: A New Test and Evaluation of Its Behavior in Realistic Applications

Veronica J. Vieland[1] and Jian Huang[2]

[1]Departments of Preventive Medicine, Division of Biostatistics, and Psychiatry, College of Medicine, and [2]Department of Statistics and Actuarial Science, University of Iowa, Iowa City

Summary

The discovery that microsatellite repeat expansions can cause clinical disease has fostered renewed interest in testing for age-at-onset anticipation (AOA). A commonly used procedure is to sample affected parent-child pairs (APCPs) from available data sets and to test for a difference in mean age at onset between the parents and the children. However, standard statistical methods fail to take into account the right truncation of both the parent and child age-at-onset distributions under this design, with the result that type I error rates can be inflated substantially. Previously, we had introduced a new test, based on the correct, bivariate right-truncated, age-at-onset distribution. We showed that this test has the correct type I error rate for random APCPs, even for quite small samples. However, in that paper, we did not consider two key statistical complications that arise when the test is applied to realistic data. First, affected pairs usually are sampled from pedigrees preferentially selected for the presence of multiple affected individuals. In this paper, we show that this will tend to inflate the type I error rate of the test. Second, we consider the appropriate probability model under the alternative hypothesis of true AOA due to an expanding microsatellite mechanism, and we show that there is good reason to believe that the power to detect AOA may be quite small, even for substantial effect sizes. When the type I error rate of the test is high relative to the power, interpretation of test results becomes problematic. We conclude that, in many applications, AOA tests based on APCPs may not yield meaningful results.

Received September 30, 1997; accepted for publication January 22, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. Veronica J. Vieland, Preventive Medicine and Environmental Health, 2800 SB, University of Iowa College of Medicine, Iowa City, IA 52242. E-mail: veronica-vieland@uiowa.edu

0002-9297/98/6205-0027$02.00

1. Introduction

The discovery of expanding trinucleotide-repeat (ETNR) diseases has sparked renewed interest in statistical evaluation of age-at-onset anticipation (AOA), which is the tendency for children to develop clinical disease at an earlier age than their affected parents. Since ETNRs are, at present, the only known genetic explanation for true AOA, detection of AOA for a disorder of otherwise unknown genetic etiology may shed light on the mode of inheritance and may also influence strategies for genomic linkage studies. Furthermore, investigators involved in linkage studies will have large numbers of multiplex families on hand and readily available for secondary analyses, and AOA testing based on these samples is virtually cost free. These considerations have led to a recent proliferation of reports of AOA in a variety of disorders, including, among others, schizophrenia (Bassett and Honer 1994), bipolar disorder (McInnis et al. 1993), leukemia (Horwitz et al. 1996), Parkinson disease (Bonifati et al. 1995), breast cancer, colon cancer, Alzheimer disease, maturity-onset diabetes of the young, and insulin-dependent diabetes mellitus (Paterson et al. 1996). (For a recent review, see McInnis 1996.)

Heiman et al. (1996) pointed out a serious flaw in the statistical approach of many of these reports (also see Penrose 1948). A standard design is to test for a difference in mean age at onset between parents and children in a sample of affected parent-child pairs (APCPs), using a paired *t*-test or related nonparametric statistical procedure (e.g., see Myers et al. 1985; Zatz et al. 1995; Paterson et al. 1996). Clearly, since both the parent and the child in each pair are affected, whatever the current age of each may be, the age at onset must be prior to the current age. However, since the child is in general younger than the parent at the time of assessment, this right truncation of the age-at-onset distribution will be more pronounced in the children than in the parents. This produces, purely as an artifact of truncation, a tendency for the mean observed age at onset in the children to be lower than the mean in their parents. The result is statistical bias away from the null hypothesis of no AOA.

Using simulations, Heiman et al. (1996) showed that the propensity of the ordinary paired *t*-test to reject the hypothesis of no AOA when in fact it is true—that is, when in fact there is no AOA—can be extremely high, depending on the generating model. For some of the models they considered, false rejection rates could be virtually 100%. This raises obvious difficulties for interpretation of the positive AOA findings in the literature.

In an earlier paper (Huang and Vieland 1997), we proposed a new test for AOA, based on the correct, bivariate right-truncated, age-at-onset distribution. We showed that this new test has the correct (nominal) asymptotic type I error rate for a random sample of APCPs, in marked contrast with the paired *t*-test and other related testing procedures. However, in that paper we did not address two key statistical complications that are relevant to real applications of the test. First, we did not explicitly define what was meant by a *random sample of APCPs*. In the current paper, we show that, in this context, random sampling of APCPs corresponds to a form of single ascertainment (viz., generalized single ascertainment; Hodge and Vieland 1996) and that any other ascertainment scheme will affect the type I error rate of the test. In particular, we show that, for the types of data generally available for AOA testing—namely, linkage samples—the type I error rate of the test will tend to be inflated.

A second statistical consideration that, to our knowledge, has not received any systematic attention is the power of any statistical test to reject the null hypothesis when in fact it is false—that is, when there really is AOA. In this paper, we consider the nature of the sampling frame when the underlying cause of true AOA is an ETNR. We show that there is good reason to suspect that the power of any correctly formulated test may be extremely low, even for substantial effect sizes. We conclude that, under sampling conditions that are realistic, both in terms of ascertainment and sample size, even the best tests may be expected to have inflated type I error rates and quite low power. Under these circumstances, interpretation of test results becomes problematic (see section 5).

The paper is organized as follows: following (1) this Introduction, we (2) briefly review the formulation of a correct test statistic for random APCPs; we then (3) define what is meant by a "random" sample, place this definition in the context of the problem of ascertainment, and show that when APCPs are obtained from linkage samples, we should expect the type I error rate of our AOA test to be inflated; next (4), we consider the statistical model appropriate under the alternative hypothesis and show that the power of any AOA test may be expected to be low when the underlying biological mechanism is an ETNR; and, finally (5), we briefly consider the difficulties of interpreting statistically significant findings in view of the conclusions drawn in sections 3 and 4. Computational details for the new AOA test are given in Appendix A. Appendix B contains details of the simulation procedures (see sections 3 and 4).

2. An Appropriate Statistical Test

In this section, we briefly review the derivation of a new AOA test statistic following Huang and Vieland (1997). Consider a random sample of APCPs (see section 3 for a definition of "random" in this context). Let C_1 and C_2 be the parent's and the child's current ages (at interview), respectively; let X_1 and X_2 be the parent's and child's respective ages at onset. We assume that the pairs (C_1, C_2) and (X_1, X_2) are independent of each other in the population. Note that this assumption precludes an effect of the disease on fertility, which may be unrealistic for many genetic disorders. We assume that age at onset follows a bivariate normal distribution in APCPs. This assumption provides the theoretical framework for development of the test statistic, and it is likely to be quite innocuous in practice: the test statistic is going to be a function of a sum of random variables, so that, at least in large samples, the underlying distribution of (X_1, X_2) is immaterial.

Let $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ be the parameters of the age at onset distribution, so that the marginal mean age at onset for the parents is μ_1, the marginal mean for the children is μ_2, and so forth. Let f_θ and F_θ represent the probability density function (PDF) and cumulative distribution function (CDF), respectively, of the age-at-onset pairs (X_1, X_2); let g_γ and G_γ represent the PDF and CDF, respectively, of the current-age pairs (C_1, C_2). The subscript γ is used here to indicate the parameters of the current-age distribution. However, we leave the distributional form of g (and G) completely unconstrained in the likelihood. Accordingly, γ is completely unspecified as to both number and functional form of its constituent parameters.

For a single APCP, the joint likelihood of θ and γ is defined, up to an arbitrary multiplicative constant, by

$$L(\theta, \gamma) \propto \frac{f_\theta(x_1, x_2) g_\gamma(c_1, c_2)}{P_{\theta,\gamma}(X_1 \leqslant C_1, X_2 \leqslant C_2)} .$$

We follow the usual convention that uppercase letters represent random variables and lowercase letters represent their values. Thus, the expression in the denominator is independent of the observed current ages (c_1, c_2). Note again that the form of g (or G) is left completely unspecified, so that this is actually a semiparametric likelihood. We can write the PDF of the *observed* current-age pairs as

$$g^*(c_1, c_2) = \frac{g_\gamma(c_1, c_2)F_\theta(c_1, c_2)}{P_{\theta,\gamma}(X_1 \leqslant C_1, X_2 \leqslant C_2)} \ .$$

Thus, by multiplying the numerator and denominator of $L(\theta, \gamma)$ by a common factor $F_\theta(c_1, c_2)$, we can rewrite the likelihood as

$$L(\theta, \gamma) = \frac{f_\theta(x_1, x_2)}{F_\theta(c_1, c_2)} \times \frac{g_\gamma(c_1, c_2)F_\theta(c_1, c_2)}{P_{\theta,\gamma}(X_1 \leqslant C_1, X_2 \leqslant C_2)}$$

$$= \frac{f_\theta(x_1, x_2)}{F_\theta(c_1, c_2)} dG^*(c_1, c_2) \ ,$$

where G^* is the CDF corresponding to g^* and where $dG^*(c_1, c_2)$ is the mass that G^* puts at (c_1, c_2). Because G is completely nonparametric, so is G^*. It is well known that the (nonparametric) maximum-likelihood estimator of G^* is the empirical distribution function based on observed current-age pairs, which does not involve θ. Thus the joint likelihood of θ and γ (the completely unconstrained parameter set) is proportional, with respect to θ, to

$$L_c(\theta) = \frac{f_\theta(x_1, x_2)}{F_\theta(c_1, c_2)} \ .$$

This likelihood is simply (proportional to) the *conditional* probability $P(x_1, x_2 | X_1 \leqslant C_1, X_2 \leqslant C_2, c_1, c_2|)$—that is, the conditional probability of the observed ages at onset *given* the current ages (and given that both parent and child are affected), which does not involve the distribution of the current ages. Thus, when the form of g (or G) is left completely unconstrained in $L(\theta, \gamma)$, the joint likelihood is proportional to the conditional likelihood $L_c(\theta)$, which conditions on the observed current ages and does not involve the current-age parameter γ. In order to calculate the likelihood for a full data set, $L_c(\theta)$ is multiplied over all observed APCPs in the usual way.

[The proof of equivalence between $L(\theta, \gamma)$ and $L_c(\theta)$ given here is the direct analog, in the continuous case, of an argument given elsewhere by Hodge (1988), in the discrete case (also see Kalbfleisch and Sprott 1970; Ewens 1982; Hodge 1985; Vieland and Hodge 1995). In this application, we consider the relevant pedigree structure to be the current-age structure of the observed APCPs, and the relevant phenotypes are the ages at onset. Note that both this proof and the earlier one are restricted to what has been called proband-independent, or PI, sampling (Vieland and Hodge 1995).]

Given that $L_c(\theta)$ is equivalent to $L(\theta, \gamma)$, standard theory shows that the maximum-likelihood estimator $\hat{\theta}$ based on $L_c(\theta)$ is consistent for θ and has an asymptotic normal distribution, and its variance-covariance matrix can be estimated in the usual way, via the observed information. Let $\hat{\sigma}$ be the estimated standard error of $\hat{\mu}_1 - \hat{\mu}_2$. It can be shown that the test statistic $Z_{AO} = (\hat{\mu}_1 - \hat{\mu}_2)/\hat{\sigma}$ is approximately distributed as a standard normal random variable in large samples (Wald 1943). Therefore, a Z-test based on Z_{AO} yields the correct type I error rate, in marked contrast with the paired t-test examined and criticized by Heiman et al. (1996). We have confirmed through simulation studies, following the procedures of Heiman et al. (1996), that this holds even for quite small samples ($n \geqslant 25$) of randomly sampled APCPs (results available on request; also see Huang and Vieland 1997).

3. Ascertainment Bias and the Type I Error Rate of Our AOA Test

The previous section established the asymptotic distribution of the test statistic Z_{AO}, for *random* samples of APCPs. In this section, we show that sampling APCPs from families obtained for linkage studies, as is often done, will tend to inflate the type I error rate of our AOA test. Specifically, we (*i*) define explicitly what is meant by a *random sample* of APCPs; (*ii*) we show that *random* sampling of APCPs corresponds to an ascertainment scheme defined elsewhere (Hodge and Vieland 1996) as *generalized single ascertainment*; and (*iii*) we then show that preferential selection of pedigrees for the presence of multiple affected individuals, as is common in linkage studies, or what we shall call "multiplex ascertainment," does *not* correspond to generalized single ascertainment. Finally (*iv*), we show that multiplex ascertainment will tend to inflate the type I error rate of an AOA test based on APCPs. We illustrate each step with a simple, discrete example (sections 3.2, 3.4, and 3.5). (Note that the definition of multiplex ascertainment used here differs somewhat from others in the literature; e.g., see Morton 1959.)

3.1. *Random Ascertainment of APCPs*

We can picture random ascertainment of APCPs by using the following contrivance: suppose that we are sampling from a universe consisting entirely of parent-child pairs (without additional relatives). Some of these pairs will have neither parent nor child affected, some will have one but not the other affected, and some will have both affected. If we take a random sample from the subspace consisting of all and only those pairs with both parent and child affected, then we say we have a *random sample of APCPs* (note that "random sampling" is not being used to mean ascertainment independent of phenotype). In other words, if this is the true ascertainment scheme, then the appropriate likelihood is just $L_c(\theta)$, and the only "ascertainment correction" to the distribution of (X_1, X_2) is the factor $1/P_\theta(X_1 \leqslant$

Table 1

Sampling Probabilities for All Possible Configurations of c (Current-Age Pairs) and x (Age-at-Onset Pairs) for APCPs, under Random Ascertainment and Generalized Single Ascertainment

		Random Ascertainment			Generalized Single Ascertainment:
c	x	$P(x, c,$ asc'd)	$P(x, c \mid$ asc'd)	$P(x \mid$ asc'd, $c)$	$P(x, c \mid$ asc'd)
c_2, c_1	x_1, x_1	1/32	1/15	1/3	4/60 = 1/15
c_2, c_1	x_2, x_1	2/32	2/15	2/3	8/60 = 2/15
c_3, c_2	x_1, x_1	1/32	1/15	1/12	4/60 = 1/15
c_3, c_2	x_2, x_1	2/32	2/15	2/12	8/60 = 2/15
c_3, c_2	x_3, x_1	1/32	1/15	1/12	4/60 = 1/15
c_3, c_2	x_1, x_2	2/32	2/15	2/12	8/60 = 2/15
c_3, c_2	x_2, x_2	4/32	4/15	4/12	16/60 = 4/15
c_3, c_2	x_3, x_2	2/32	2/15	2/12	8/60 = 2/15

Note.—Under random ascertainment, $P(\text{asc'd}) = P(\text{an APCP is ascertained}) \propto \sum_x \sum_c P(x, c) = 15/32$. The fourth column is obtained by dividing the third column by 15/32; the fifth column is obtained by dividing each entry in the fourth column by $P(c)$, which is 3/15 if $c = (c_2, c_1)$ or 12/15 if $c = (c_3, c_2)$. Under generalized single ascertainment, $P(\text{asc'd}) = P(\text{pedigree is ascertained}) \propto$ no. of APCPs in the pedigree $= 60\pi/12^3$, where π is an arbitrary scaling constant and where each ascertained pedigree contributes exactly one pair to the data set. See the text for computation of $P(x, c, X \leqslant C)$ under generalized single ascertainment.

$c_1, X_2 \leqslant c_2$), which corrects for the right truncation of the sample space due to observation of only those pairs in which both ages at onset are prior to the observed current ages. Thus, thinking in terms of this artificial sampling frame is one particular way to envision a sampling frame and ascertainment scheme such that, as the sample size $n \to \infty$, Z_{AO} will converge to its theoretical limiting distribution. We note that this is exactly the sampling frame modeled in our previous work (Huang and Vieland 1997) and in the work by Heiman et al. (1996).

3.2. *Ascertainment Example Part I: Random Ascertainment of APCPs*

We begin by constructing a wholly artificial sampling frame to serve as a point of reference in what follows. Consider a universe of parent-child pairs, each of which has one of two equally likely, but different, current-age structures. Let the possible current ages in the population be c_3, c_2, and c_1, and let each parent-child pair be either (c_2, c_1) or (c_3, c_2), where the first individual listed is always the parent. (Note that there is a change of notation from the previous section. There, the subscripts "1" and "2" denote parent and child, respectively; here, the first individual listed within the pair is always the parent, while the subscripts will index a specific age [current or onset].) Let there be three possible ages at onset for an individual: $x_1 = c_1$, $x_2 = c_2$, or $x_3 = c_3$, occurring with probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively. (Note that every individual in the population will be affected on or before age c_3 with probability 1. This simplifies the tables that follow.) Let the age at onset and the current age be independent for each individual, and let the ages at onset be independent for the two individuals within each parent-child pair.

Under these assumptions, it is easy to calculate the probability of any given APCP. For example, $P[(x_1, x_1), (c_2, c_1)] = (\frac{1}{4})(\frac{1}{4})(\frac{1}{2}) = \frac{1}{32}$, et cetera. There are eight possible APCP configurations. Table 1 shows all possible observed pairs and their probability distributions, for randomly ascertained samples of APCPs. The probabilities $P(x, c \mid X \leqslant C)$ correspond to the components of our likelihood $L(\theta, \gamma)$, given above, while the probabilities $P(x \mid X \leqslant C, c)$ correspond to $L_c(\theta)$.

Note that the (expected value of the) mean age at onset in the parents is higher than the mean age at onset in the children within the randomly ascertained, truncated sample space shown in table 1. For example, let $x_1 = 10$, $x_2 = 30$, and $x_3 = 50$. Then the mean difference between the parents' ages at onset and the childrens' in this example is 8 years, even though by design there is no AOA in this hypothetical population. This illustrates the cause of inflation in the type I error rate of the paired t-test for APCP data, as pointed out by Heiman et al. (1996). We return to this example in the next section.

3.3. *Generalized Single Ascertainment*

It is not immediately obvious how this particular way of defining random sampling can be applied to sampling in the context of human pedigrees, since people do not come in discrete parent-child pairs but, rather, in arbitrarily large extended pedigrees, some of which may contain multiple (nonindependent) APCPs. Often in practice, larger pedigrees are ascertained (e.g., for the

purposes of linkage analysis), and then, for purposes of AOA testing, a single APCP is selected at random from within each ascertained pedigree. In this case, the definition of the appropriate subspace of the sampling frame is defined by the joint probability of the pedigree being ascertained and of the APCP being selected from the ascertained pedigree, or P(pedigree is ascertained) × P(pair is sampled|pedigree is ascertained).

We recall that single ascertainment, in the sense familiar to human geneticists, can be defined as any ascertainment scheme such that P(pedigree is ascertained) = P(asc'd) is strictly proportional to the number of affected individuals the pedigree contains (Stene 1977). Hodge and Vieland (1996) defined a new class of ascertainment schemes, designated "generalized single ascertainment," which extend this definition to allow for sampling through sets of individuals. For example, if pedigrees are ascertained through APCPs in such a way that P(pedigree is ascertained) is strictly proportional to the number of APCPs in the pedigree, this would be generalized single ascertainment through parent-child pairs.

Hodge and Vieland (1996) showed, with application to the present context, that, when pedigrees are ascertained through APCPs under generalized single ascertainment, the correct likelihood is proportional to $P(\boldsymbol{x} \mid \boldsymbol{X} \leqslant \boldsymbol{C}, \boldsymbol{c}) = L_c(\theta)$. Thus, the likelihood is the same under both generalized single ascertainment through APCPs and random sampling as defined in the previous section; or, in other words, generalized single ascertainment through APCPs is mathematically equivalent to random sampling of APCPs as defined above.

3.4. *Ascertainment Example Part II: Generalized Single Ascertainment*

We continue with the previous example in all particulars except one: we now suppose that rather than coming in parent-child pairs, all families in our universe consist of three people, one of each current age c_3, c_2, and c_1 (grandparent-parent-grandchild triplets). Thus, each triplet can contain 0, 1, or 2 APCPs. We calculate the probability distribution of these triplets on the basis of the same assumptions as used previously; that is, we assume that the age-at-onset distribution is $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, for x_1, x_2, and x_3, respectively, and we assume independence of ages at onset for individuals within "pedigrees." However, we now assume that a pedigree with two APCPs is twice as likely to contribute a pair to the sample as a pedigree with only one (generalized single ascertainment).

We continue to assume that each ascertained pedigree contributes exactly one APCP to the sample. Thus, for any ascertained pedigree containing two APCPs, we assume that one of the two is picked at random (with a 50/50 probability) for inclusion in the sample. We note that this procedure for picking a single APCP to sample from a pedigree containing more than one APCP corresponds to common practice (e.g., see McInnis et al. 1993).

For example, there are two possible pedigree configurations that can contribute a (c_3, c_2), (x_1, x_1) pair: one in which the grandchild is unaffected, and one in which the grandchild is affected, with age at onset x_1. The former occurs with probability $3/12^3$; the latter with probability $1/12^3$. In the latter case, all three individuals are affected, yielding two APCPs, and we assume that this pedigree is twice as likely as the first to contribute a pair to the sample. At the same time, which of the two pairs it contributes is random. Thus, the resulting probability of a (x_1, x_1), (c_3, c_2) pair is $3\pi/12^3 + (\frac{1}{2})2\pi/12^3 = 4\pi/12^3$, where π is an arbitrary (small) scaling constant.

The final column of table 1 shows the probability distribution of APCPs under generalized single ascertainment. It is readily confirmed that both the joint distribution $P(\boldsymbol{x}, \boldsymbol{c} \mid \boldsymbol{X} \leqslant \boldsymbol{C})$ and the conditional distribution $P(\boldsymbol{x} \mid \boldsymbol{X} \leqslant \boldsymbol{C}, \boldsymbol{c})$ are identical to those obtained under random sampling. This illustrates the equivalence between random sampling from a universe of parent-child pairs and generalized single ascertainment from a universe of larger pedigrees.

We note again that, in the case of generalized single ascertainment, as in the equivalent case of random ascertainment, the (expected value of the) mean age at onset in parents will be higher than the mean in children, even in the absence of any true AOA.

(The proof of equivalence given above, in section 2, pertained to PI sampling [Vieland and Hodge 1995]. By contrast, the proof of equivalence given here pertains to proband-dependent [PD] sampling [Vieland and Hodge 1995] but is restricted to the special case of [generalized] single ascertainment. This argument is a direct application of the proof given by Hodge and Vieland [1996]. Note also that Hodge and Vieland [1996] offer a broader definition of generalized single ascertainment than the one we have used here.)

3.5. *Ascertainment Example Part III: Multiplex Ascertainment*

We now illustrate the bias away from the null hypothesis, which occurs when APCPs are drawn (at random) from families preferentially selected for the presence of multiple affected individuals (multiplex ascertainment). We continue with the previous example of three-person pedigrees, in which the true, underlying current-age structure of each pedigree is (c_3, c_2, c_1). We make the same assumptions regarding the age-at-onset distribution as in the previous example. However, we

now consider an extreme ascertainment scheme in which we (*i*) ascertain (all and) only pedigrees in which all three individuals are affected and then (*ii*) randomly select one APCP from each ascertained pedigree. Table 2 shows the sampling proportions under this multiplex ascertainment scheme.

It is readily apparent that the joint distribution $P(\boldsymbol{x}, \boldsymbol{c} \mid \text{asc'd})$ under multiplex ascertainment (table 2) is not the same as the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{c}$ under generalized single ascertainment (table 1). This illustrates that sampling of pairs from pedigrees that are themselves preferentially ascertained for the presence of multiple affected individuals is *not* equivalent to generalized single ascertainment or random ascertainment of APCPs.

It is readily confirmed that the mean parent-child difference in age at onset, under the multiplex ascertainment scheme (table 2), is inflated relative to the mean difference under generalized single ascertainment (table 1). For example, when $x_1 = 10$, $x_2 = 30$, and $x_3 = 50$, as before, the expected mean parent-child age-at-onset difference based on table 2 is 10 years, whereas table 1 yielded a mean difference of 8 years (again, recall that, by design, the true mean difference in the hypothetical population is 0).

Note however that the probability distribution $P(\boldsymbol{x} \mid \text{asc'd}, \boldsymbol{c})$, calculated conditionally on the observed current-age structure c, is identical in Tables 1 and 2. The reason that the mean parent-child age-at-onset difference is inflated under multiplex ascertainment is that multiplex ascertainment results in a downward bias in the mean *current ages* of the observed parent-child pairs. Selection of affected pairs tends to inflate the current-age distribution in the sample relative to the population current-age distribution, since older individuals are more likely to be affected. But ascertainment of multiplex pedigrees produces as an artifact a lessening of this inflation of the current-age distribution: pedigrees that happen to have some younger affected individuals are more likely to have multiple affected individuals, and the result is a sample with a higher representation of younger affected individuals. It is readily confirmed in our example that when, for instance, $c_1 = 10$, $c_2 = 30$, and $c_3 = 50$ the mean current ages of parents and children, under our multiplex example, are 40 and 20, respectively, which are identical to the "population" means; however, under generalized single ascertainment (table 1), these means are 46 and 26, respectively.

Since younger affected individuals represent greater truncation of the age-at-onset distribution than older individuals, any ascertainment scheme that *deflates* the representation of older individuals, relative to the distribution expected under generalized single ascertainment, will tend to *inflate* the mean parent-child age-at-onset distribution beyond what is corrected for by $L_c(\theta)$.

Table 2

Sampling Probabilities for All Possible Configurations of c (Current-Age Pairs) and x (Age-at-Onset Pairs) for APCPs, under Multiplex Ascertainment

		Multiplex Ascertainment	
$\boldsymbol{c}$	$\boldsymbol{x}$	$P(\boldsymbol{x}, \boldsymbol{c} \mid \text{asc'd})$	$P(\boldsymbol{x} \mid \text{asc'd}, c)$
c_2, c_1	x_1, x_1	4/24	1/3
c_2, c_1	x_2, x_1	8/24	2/3
c_3, c_2	x_1, x_1	1/24	1/12
c_3, c_2	x_2, x_1	2/24	2/12
c_3, c_2	x_3, x_1	1/24	1/12
c_3, c_2	x_1, x_2	2/24	2/12
c_3, c_2	x_2, x_2	4/24	4/12
c_3, c_2	x_3, x_2	2/24	2/12

Note.—$P(\text{asc'd}) = 1$ if and only if all three individuals within a pedigree have age at onset prior to current age, that is, only if all three are affected.

The resulting estimate of the mean parent-child difference in age at onset will be upwardly biased.

(This example represents PD sampling and an ascertainment scheme that is not [generalized] single. In this case, we have shown that the current-age distribution carries information about the parameters of the age-at-onset distribution, and conditioning on current ages introduces asymptotic bias into estimates of those parameters. This is a special application of the general result proved by Vieland and Hodge [1995]—namely, when the observed pedigree structure depends on the ascertainment event [PD sampling] and when ascertainment is not [generalized] single, then conditioning on observed pedigree structure will produce errors in genetic parameter estimates.)

3.6. *Effects of Multiplex Ascertainment on the Type I Error Rate of Our Test*

We have confirmed, using simulations, that this bias in estimation will indeed result in an inflation of the type I error rate of our test (see Appendix B for details of the simulation procedures). In brief, we simulated random APCPs with one of two current-age structures and then enriched our sample for the presence of "young" pairs, in order to mimic the effects of multiplex ascertainment on the current-age distribution. Table 3 shows results for one generating model, with true mean age at onset $\mu = 55$ for both parents and children.

The first row of table 3 gives results under generalized single ascertainment. We note that, as expected, the proportion (4.2%) of young pairs in ascertained samples is quite a bit smaller than the generating proportion (50%). Accordingly, the average age ($\bar{c} = 72$) in ascertained samples is high relative to the generating current-age distribution ($\bar{c} = 60$). Also as expected, the observed mean ages at onset are only very slightly downwardly biased in the parents, for whom there is relatively little

Table 3

Simulation of Effects of Observed Current-Age Distribution on Type I Error Rate α

Percentage Young	$\bar{c}$	$\bar{x}_1, \bar{x}_2$	$\bar{x}_1 - \bar{x}_2$	$\hat{\mu}_1, \hat{\mu}_2$	$\hat{\mu}_1 - \hat{\mu}_2$ (SE)	α
4.2	72	55, 50	5.2	55, 55	.0 (3.2)	.056
10	71	54, 48	6.5	55, 55	.0 (3.4)	.065
25	66	54, 45	8.7	55, 55	.0 (3.6)	.069
50	60	52, 40	11.9	55, 55	.2 (4.1)	.072
75	54	51, 36	15.2	55, 54	.4 (5.2)	.077
100	48	50, 31	18.6	54, 50	3.6 (7.1)	.149

NOTE.—1,000 replicates of $n = 40$ APCPs per data set were generated with $\mu = 55$ and $\sigma = 10$, under the null hypothesis of no AOA, and a test based on Z_{AO} was applied at the 5% significance level. "Percentage Young" represents the proportion of young pairs in every sample (fixed by design); $\bar{x}_1$ and $\bar{x}_2$ are the observed mean age at onset in parents and children, respectively; $\hat{\mu}_1$ and $\hat{\mu}_2$ are the estimated mean ages at onset; SE = standard error; and α is the observed type I error rate of the test over the 1,000 replicates. The results for $\bar{x}$ and $\hat{\mu}$ have been rounded to whole numbers.

right truncation of the age-at-onset distribution, but are somewhat more biased in the children (sample averages have been rounded to whole numbers). This results in an observed parent-child difference in age at onset of 5.2 years, under generalized single ascertainment, when in fact the true difference is 0. However, the corrected mean ages at onset ($\hat{\mu}_1, \hat{\mu}_2$), obtained from maximization of $L_c(\theta)$, are unbiased. Furthermore, the type I error rate is very close to the nominal 5%. (Any bias here is small-sample bias and is not related to the ascertainment scheme, per se. We have confirmed that the type I error rate is exactly 5% for large samples, under generalized single ascertainment [results not shown].) These results confirm the theory presented above.

The anticipated effect inherent in multiplex ascertainment schemes will be to inflate the proportion of young pairs, relative to the 4.2% obtained under generalized single ascertainment. The remaining rows of table 3 show the impact on the type I error rate of increasing this proportion. We find that, as the preferential selection of young pairs increases, the observed mean age-at-onset difference between parents and children also increases, as expected. Moreover, although the "corrected" mean difference, based on maximization of our conditional likelihood, $L_c(\theta)$, is less than the observed mean difference, there nevertheless is residual bias, particularly in the estimate of the child age-at-onset mean μ_2. As a result, the true type I error rate of the test does in fact increase as the mean current age of the sample decreases. However, in stark contrast to the ordinary paired *t*-test, the test appears to be fairly robust. Even at the most extreme level of distortion shown here, when *all* of the pairs are young pairs, the type I error rate is only 15%.

This simulation is not intended to represent an accurate model of the complexity of true current-age distributions in natural human populations. For this reason, even though sampling 100% young pairs might appear extreme, this model in no way represents the maximum effect of ascertainment bias on the real type I error rate of the test. (However, note that, when APCPs are drawn from pedigrees preferentially selected for the presence of affected individuals in multiple generations, the resulting proportion of younger [vs. older] pairs may well be >50%.) In the model shown here, the overall mean observed current age $\bar{c}$ remains well above the mean age at onset (55 years), until the two final rows of table 3. Increasing the mean age at onset or decreasing the true mean current age will produce higher type I error rates in the second to sixth rows. Because the effects of ascertainment on the type I error rate are mediated by the impact on the observed current-age distribution, effects may be highly variable from application to application, and, in some cases, error rates may be substantially higher than those shown here. Note also that attempting to remedy this effect by testing at a more stringent significance level will reduce the power to detect true AOA. As the next section suggests, this may not be prudent.

4. A Realistic Model under the Alternative Hypothesis and the Power of Any AOA Test

In order to assess the power of our AOA test, it is necessary to specify an alternative hypothesis. However, consideration of the underlying biology suggests that, under the hypothesis that an ETNR is causing AOA, the appropriate mathematical model is not a single bivariate (normal) distribution, as it is under the null hypothesis of no AOA. Rather, the appropriate model is a *mixture* of bivariate normal distributions.

Suppose that there really is AOA and that the cause is a microsatellite-repeat that tends to expand in length during transmission from affected parent to child. We may think of each parent as belonging to one of k classes of repeat length, at the locus in question. (Here, k is arbitrary and may represent one class for every possible number of repeats in the sequence, or groupings of repeat lengths; e.g., all individuals with 20–30 repeat lengths might be grouped together into a single class. See below for further comments.) Associated with each repeat-length class i is an age-at-onset distribution with some mean μ_{i1} and some variance σ_{i1}, where the subscript "1" reminds us that these are parameters of the marginal parental distributions. Under the alternative hypothesis, the greater the repeat length, the lower the associated mean age at onset μ_{i1} will be. Thus, one effect of an ETNR will be the tendency for the k parental means μ_{i1} to be spread out along the X_1 plane, as a function of repeat length. We call this the "interdistributional" ETNR effect.

Table 4

Simulation of Interdistributional ETNR Effects on Power to Detect AOA, under a Mixture Model

Nonmixture Model			Mixture Model				
			Distribution 1		Distribution 2		
μ_1	μ_2	Power	μ_1	μ_2	μ_1	μ_2	Power
35	30	.98	40	35	30	25	.32
45	40	.92	50	45	40	35	.30
40	35	.96	45	40	35	30	.34
40	35	.96	50	45	30	25	.06

Note.—1,000 replicates of $n = 40$ randomly ascertained APCPs per data set were generated under the alternative hypothesis of true AOA, and a test based on Z_{AO} was applied at the 5% significance level. Mixture models are generated by mixing two bivariate normal distributions, with given marginal means μ_1 and μ_2 in equal proportions; corresponding nonmixture models have the same overall mean parent-child age-at-onset difference. All models shown are generated with marginal SDs $\sigma = 5$, and intradistributional correlation coefficients $\rho = 0$.

Then, under the hypothesis that there is a tendency for deleteriously large repeat lengths to expand when transmitted, associated with each class i is a bivariate distribution describing the joint behavior of the ages at onset for an APCP, in terms of μ_{i1} and μ_{i2}, the mean age at onset in children for the ith distribution, and any relationship (correlation) within parent-child pairs. Note that there may be a tendency for the difference in mean parent-child ages at onset, *within* each of the k constituent bivariate distributions, to change as a function of repeat length; similarly, the marginal within-distribution variances also may be affected by repeat length. We will call these the "intradistributional" effects of repeat length on AOA.

Because, in the usual applications of AOA testing—that is, prior to cloning of the relevant ETNR—we cannot observe repeat lengths directly, we cannot assign individuals to particular repeat-length classes. When we sample APCPs, therefore, we are actually sampling from a mixture of, say, bivariate normal distributions (one for each underlying parental repeat-length class i). Thus, the overall probability density function of the age-at-onset pair $\boldsymbol{X} = (X_1, X_2)$ is a mixture distribution of the form $f_x(\theta) = \alpha_1 f_1(\theta_1) + \alpha_2 f_2(\theta_2) + \ldots + \alpha_k f_k(\theta_k)$, where α_i is the unknown mixing proportion and where each constituent density f_i has its own parameter vector $\theta_i = (\mu_{i1}, \mu_{i2}, \sigma_{i1}, \sigma_{i2}, \rho_i)$. Hence, there are $6k - 1$ parameters in the full model (one vector θ for each of the k classes, plus the $k - 1$ mixing proportions α_i, where $\alpha_k = 1 -$ sum of the remaining α_i's).

Truncation of the age-at-onset distributions greatly complicates the picture. For example, suppose that the sole effect of the ETNR on AOA is a tendency for the mean age at onset in parents to decrease with increasing repeat length, with the relationship between parents' ages at onset and their children's ages at onset constant across the mixing distributions; that is, we suppose that there are interdistributional effects of an ETNR but no intradistributional effects. Then, in the absence of truncation, the distribution of the parent-child difference in age at onset, $D = X_1 - X_2$, is not in fact a mixture model, since each of the constituent distributions has, by stipulation, the same mean parent-child difference D and the same variance of this difference (we thank Scott Zeger for bringing this point to our attention). In this case, the larger the impact of repeat length on D, the greater the power of any AOA test.

However, when the constituent distributions are differentially right truncated for parents and their children, this is no longer the case. Even in the absence of any intradistributional effects, the resulting model is truly a mixture distribution with respect to D, and the variance of the estimate of D may escalate rapidly with increasing effect size. The effects on the power of our test may be devastating.

We illustrate these truncation effects on power under a mixture model with a simple simulation. In brief, we simulated random APCPs from mixtures of two bivariate normal age-at-onset distributions, holding the intradistributional effects constant but allowing the intradistributional distance $\mu_{11} - \mu_{21}$ between the parental mean ages at onset in the two mixing distributions to increase (then, by stipulation, $\mu_{12} - \mu_{22}$ increases at exactly the same rate). For each mixture model, we also simulated a "corresponding" single-distribution model, as described below (see Appendix B, for details of the simulation procedures).

Table 4 shows characteristic results for one particular generating model. The first three rows show the relative stability of the power, under these generating conditions, across a range of generating mean ages at onset, under both the single-distribution model and the mixture model to which it corresponds. Within each row, by design, the overall mean parent-child age-at-onset difference is the same under the mixture model and the corresponding nonmixture (single-distribution) model, and the intradistributional parent-child differences are constant as well. Therefore, comparison of the power under the nonmixture model to the power under the corresponding mixture model gives the reduction in power due to interdistributional distances. Under these simulation conditions, when the interdistributional difference is 10 years, the power under the mixture model is only 30%–35%, compared with the power under the corresponding single-distribution model, which is 92%–98%, depending on the generating mean age at onset. The final row of table 4 shows the impact of increasing intradistributional distance: when the interdistributional distance between parents with one class

of repeat length and parents with another class of repeat length is increased to 20 years, the power plummets to only 6%. Bearing in mind that we are modeling only one source of increased variability under the bivariate right-truncated mixture model (i.e., interdistributional effects), we conclude that the power of these tests may reasonably be supposed to be low in many applications. Moreover, we would generally expect power to increase with effect size; but, in this case, it is precisely as the interdistributional effect size increases that the power decreases.

This simulation is artificial in that it involves only two particularly simple constituent distributions and only one source of added variance under a mixture model. However, it illustrates clearly that, when both the underlying biological hypothesis and the right truncation of the age-at-onset distributions are taken into consideration, there is good reason to suppose that our power to detect AOA may be low.

The underlying biological model involves several sources of variability, including possible intradistributional effects on both means and variances, and it can be shown that, in the presence of intradistributional effects, the variance of the overall parent-child difference D also tends to increase as the intradistributional effect size increases. Whether these additional effects would cause further deflation of the power, or whether they might to some extent offset the impact of interdistributional effects on power, is not a question that can be addressed, in a comprehensive manner, until the biological effects of ETNRs on age-at-onset distributions are better understood (see Discussion).

5. Interpretation of Statistically Significant AOA Tests

Thus far, we have presented an AOA test that is correct under specified sampling conditions; have shown that its real type I error rate will tend to be inflated in realistic (linkage) samples; and have argued that, when the underlying biological hypothesis—that is, that an ETNR mechanism is responsible for any observed AOA—is taken into consideration, there is good reason to believe that the power of the test may be low. At this point, a reasonable question might be, can we establish guidelines for investigators regarding the true type I and II error rates of these tests, for realistic data? We believe that the answer to this question is no. Given the numerous complicating factors, such an enterprise would need to rely on simulations, but we have established that the behavior of the test statistic will depend critically on the true underlying current-age structure of the population, the true ascertainment scheme, and the true underlying age-at-onset distribution(s). Because all three of these conditions may be unknown and/or impossible to model with any accuracy with respect to any given application, general guidelines cannot be established. Therefore, in real applications, the true underlying distribution of the test statistic will remain unknown.

In sections 3 and 4 above, we presented some simulations of type I and type II error rates for our test. The simulated models are overly simplistic but also probably overly conservative in estimating the true inflation of error rates in applications (see Discussion). However, they do suggest that, in practice, for a nominal 5% test and typical data-set sizes, both the true type I error rate and the power might very well be in the 10%–20% range.

When the type I error rate of a test exceeds its power, the test is called "biased" (Stuart and Ord 1991, p. 839), and bias in a test has a serious implication for interpretation of statistical significance. Consider the posterior probability of true AOA, given a statistically significant test result. Define $\alpha = P(\text{significant test result} \mid \text{no AOA})$ = the type I error rate of the test, $(1 - \beta) = P(\text{significant test result} \mid \text{AOA})$ = the power of the test, and $\lambda = P(\text{AOA})$ = the "prior" probability of AOA for the disease being tested; then, a simple application of Bayes's theorem yields the following:

$$\begin{aligned} &P(\text{AOA} \mid \text{significant test result}) \\ &= [P(\text{significant test} \mid \text{AOA})P(\text{AOA})]/ \\ &\quad [P(\text{significant test} \mid \text{AOA})P(\text{AOA}) \\ &\quad + P(\text{significant test} \mid \text{no AOA})P(\text{no AOA})] \\ &= (1 - \beta)\lambda/[(1 - \beta)\lambda + \alpha(1 - \lambda)] \\ &= 1/[1 + \alpha(1 - \lambda)/(1 - \beta)\lambda] \ . \end{aligned}$$

As this formula shows, a biased test, for which $\alpha > (1 - \beta)$, has the unfortunate property that the posterior probability of an effect, given a statistically significant finding, is actually *smaller* than the prior probability, no matter what the prior probability may be. In this situation, even after collecting data, performing the test, and obtaining a statistically significant result, we apparently are *less* warranted in drawing the conclusion that there is an effect than we were prior to any data collection whatsoever.

The two complicating issues we have examined—ascertainment and mixture models—lead us to conclude that, in any given application, it is plausible to suppose that we find ourselves in the unfortunate situation $\alpha > (1 - \beta)$. In this case, testing for AOA for human data is futile, in that, regardless of the outcome of the test, we will have learned nothing about the underlying biological phenomenon of interest.

6. Discussion

We have shown that it is possible to design a test for AOA based on APCPs that properly accounts for both the right truncation of the age-at-onset distribution and the age differences between parents and children. However, we have also found compelling evidence that, in real applications, the type I error rate of this test will tend to be inflated because of ascertainment bias, while consideration of the underlying biological model suggests that the power of any AOA test may be expected to be quite low when the underlying mechanism is an ETNR. Under these circumstances, achieving a statistically significant AOA test result does not necessarily mean that the data support the hypothesis of an ETNR mechanism for the disease in question.

Our consideration of the true type I error rate of the test is conservative, insofar as we have not considered complications arising from selection bias, such as those enumerated in Penrose (1948). For example, the logistics of recruiting families through clinics could, under some circumstances, lead to (inadvertent) preferential selection of parents with later age at onset and children with earlier age at onset. Insofar as selection biases of this type are present in any actual data set, they will tend to further inflate the type I error rate of any AOA test (e.g., see McInnis et al. 1993; Bassett and Husted 1997). In assuming throughout that data are obtained without selection biases of this sort, we have ignored a second important cause of inflation in the type I error rate of an AOA test based on APCPs, in real applications. Similarly, we have not considered such additional complications as fertility bias or measurement error.

Our treatment of power is also conservative, insofar as the power simulations considered only one source of additional variability under a mixture model, compared with a single bivariate distribution—that is, what we call "interdistributional effect size." However, as noted above, it is plausible to suppose that ETNRs exert effects not only on the mean ages at onset across distributions, but also on the mean parent-child difference within repeat-length classes, on the variance of the age-at-onset distribution within repeat-length classes, and on the degree of correlation between parents and children within classes. All of these effects further inflate the variance of the mean parent-child age-at-onset difference and, accordingly, may tend to reduce the power of any AOA test.

We note parenthetically that, although each constituent bivariate distribution may have $\rho_i = 0$, it is readily confirmed that the parent-child age-at-onset pairs *will* be correlated in the resulting mixture distribution. Because it is almost certainly the case that repeat lengths themselves are correlated within parent-child pairs, any model that allows for sampling of pairs across repeat-length classes will entail overall parent-child age-at-onset correlations, under the alternative hypothesis. Thus, although it is true that AOA may exist in the absence of parent-child correlations in age at onset (Hodge and Wickramaratne 1995), this fact may be irrelevant when testing specifically for ETNR mechanisms. The extent of this correlation also may influence the power of our paired test, under certain circumstances.

Another way in which our mixture model represents a conservative estimate of the loss of power under a plausible biological model is that we have considered only the case $k = 2$ (two constituent distributions). As noted, the choice of k is arbitrary (in fact, the most appropriate mathematical model might be a continuous mixture model). However, the greater the number of distributions involved in the mixture (up to the limiting case of a continuous range of distributions), the greater the variance of the parent-child age-at-onset difference D, holding all other variables constant, because $\mathrm{Var}(D)$ then will involve the sum (integral) of the individual variances from each of the constituent distributions, as well as terms involving the pairwise distances. Thus, the greater the number of the constituent distributions contributing to the overall mixture, the greater the number of sources of variability, and we might expect, in general, to see a corresponding reduction in power. (However, it is also possible that, for example, ascertainment bias, while inflating the type I error rate of the test, might actually increase the power.) Again, we have not taken into consideration additional factors that could lead to loss of power, such as measurement error.

It might be argued that, although our test based on Z_{AO} may fare badly, perhaps other, better tests could be developed. One way in which we might achieve greater power to detect AOA would be to utilize more-informative data. Possibly, other configurations of affected-relative sets would be more efficient. For example, using trios consisting of a parent with an affected sib pair, rather than a single offspring, might be a more efficient strategy (with some allowance for correlation within sib pairs incorporated into the likelihood). Certainly, a more informative type of data would be complete pedigrees, including unaffected relatives. The difficulty here is that any AOA test based on full pedigrees would need to incorporate both a genetic model (in order to assign conditional genotypic probabilities to unaffected individuals) and appropriate techniques from survival analysis, in order to correctly handle censoring and truncation of the age-at-onset distribution(s).

This represents a terrifically complicated undertaking. Among other difficulties, if it is necessary (as surely it would be for a complex disorder) to simultaneously estimate parameters of the genetic model along with parameters of the age-at-onset distribution, then it will probably prove necessary to handle ascertainment in a

reasonably robust way. Use of families collected for purposes of linkage analysis would probably not lead to reliable results, because of the effects of unsystematic ascertainment.

While power may remain a fundamental difficulty for any AOA test, it might be possible to at least work out a better ascertainment correction, in order to curb the inflation of the type I error rate that we have found. However, it is worth noting in this context that one immediate consequence of ascertainment under any scheme other than generalized single ascertainment would be that $L(\theta, \gamma)$ and $L_c(\theta)$ would no longer be equivalent, and exact correction for ascertainment would involve parametric modeling of the current-age distribution (Vieland and Hodge 1995). This, taken together with the unsystematic nature of most ascertainment schemes for linkage studies, means that exact ascertainment corrections are, for all practical purposes, impossible.

It is possible of course that robust approximate ascertainment corrections could be developed for APCPs taken from linkage pedigrees. One such approximation is to simply use $L_c(\theta)$, that is, to correct *as if* ascertainment were generalized single, even though it is not. On the basis of our limited simulations, this approach did indeed seem to work well if there was only moderate deflation of the mean observed current age. While it may be possible to improve somewhat on the performance of this particular approach to approximating an ascertainment correction, we are skeptical that much improvement in the type I error performance of the test could be obtained by tinkering with this likelihood.

Another option would be to forego the use of linkage samples and to obtain APCPs for AOA testing based on true generalized single ascertainment. This would seem to require either a complete enumeration of the sampling frame (including diagnoses) beforehand or a willingness to sample pairs of people at random, regardless of phenotype, retaining only those that, following clinical assessment, are determined to be APCPs. Either procedure could be prohibitively expensive, and, if the disease being studied is even moderately rare, either scheme can fail to yield adequate numbers of pairs, even for coverage of relatively large catchment areas. This is, of course, the reason that researchers prefer using linkage samples as sources of APCPs: linkage samples offer the only ready supply of substantial numbers of appropriate pairs.

It should also be noted that generalized single ascertainment is not the same as single ascertainment, through affected individuals, combined with the additional requirement that there be at least one APCP in the pedigree (Morton 1959; Hodge and Vieland 1996). Thus, data collected for purposes of segregation analysis, which may be obtained under single ascertainment, are also, strictly speaking, not appropriate for AOA testing based on APCPs. However, it is possible that obtaining pairs from such samples would produce less bias than drawing pairs from linkage samples. In any event, the difficulty of obtaining a critical mass of APCPs from such samples would remain, unless the recurrence risks to parents/offspring were quite high.

Finally, we return to the question of whether we could gain an assessment of type I and type II error rates, over a broad range of models, by performing more extensive and realistic simulations. There are several reasons why we feel that, for any particular case, it is not possible to say what the appropriate simulation would be. Among these are the following: when ascertainment is unsystematic, there is no way to accurately simulate ascertainment effects; the true age structure of the population being studied generally is complex, unknown, and not amenable to accurate simulation; behavior of the tests depends on the true underlying age-at-onset distribution(s), which generally are unknown and may be complicated to estimate accurately (Heimbuch et al. 1980); and, under the alternative hypothesis, very little is known about ETNR effects on age-at-onset distributions for actual ETNR diseases, so that, at present, it is not possible to confirm that any particular mixture model is a realistic representation of a true ETNR effect on age at onset (however, see Brinkman et al. 1997, for an example of work on this issue).

It could be argued that these kinds of considerations apply equally to all efforts to quantify behavior of test statistics via simulations, but there are key features that distinguish AOA tests from other tests of interest in human genetics. The behavior of the LOD score, for example, can be highly dependent on pedigree structures in small samples, but only on the *observed* pedigree structures. Thus, we have the option of conducting simulations conditional on the observed pedigree structures within any given sample, in order to obtain relevant empirical distributions of the LOD score. However, the behavior of Z_{AO} depends crucially, even in large samples, on the true, underlying pedigree structures (including current ages) in the sampling frame, which we do not know. (Strictly speaking, the LOD score under multiplex ascertainment schemes also is dependent on the true, underlying pedigree structure; this result follows the same reasoning as above [Vieland and Hodge 1996], but in contrast to AOA tests, numerical effects on the distribution of the LOD scores appear to be negligible [Slager and Vieland 1997].) Similarly, although power estimates for LOD scores are always conditional on the specifics of the generating models, at least we are able to simulate models with sound biological underpinnings. However, until more is learned about the true effects of repeat lengths on age-at-onset distributions for true ETNR diseases, it is not possible to specify an appro-

priate mixture distribution for use in calculating power for AOA tests. Therefore, although we cannot conclude that the power of an AOA test in any given application *is* smaller than its type I error rate, there does not appear to be any reliable way to establish that it is *not*.

In conclusion, despite the fact that AOA testing based on APCPs taken from linkage samples is a virtually cost-free enterprise and that we have provided an AOA test that does not appear to be subject to the extreme bias in type I error rates shown by the ordinary paired *t*-test and related procedures, we reluctantly have come to the opinion that the results of any AOA test based on APCPs, in human applications, may be largely uninterpretable. We would not be inclined to interpret a statistically significant test as providing evidence that an expanding microsatellite-repeat mechanism underlies the disease being studied.

Acknowledgments

We wish to thank Gary Heiman and Chantal Mérette, for helpful discussion; Susan Hodge, for helpful comments on the manuscript; and Scott Zeger, for calling our attention to an error in the original paper. This work was supported by National Institutes of Mental Health grants R01-55378, R01-48858, and K02-01432 (to V.J.V.) and R01-01541 (to J.H.).

Appendix A

Computational Details

Let F_1' and F_2' be the partial derivatives of F_θ with respect to μ_1 and μ_2, respectively. Some calculation shows that $\hat{\mu}_1$ and $\hat{\mu}_2$ satisfy the conditional likelihood equations

$$\mu_1 = \bar{x}_1 + \left[-\sigma_1^2 \sum_{i=1}^{n} \frac{F_1'(c_{1i}, c_{2i})}{F_\theta(c_{1i}, c_{2i})} - \rho\sigma_1\sigma_2 \sum_{i=1}^{n} \frac{F_2'(c_{1i}, c_{2i})}{F_\theta(c_{1i}, c_{2i})} \right], \tag{A1}$$

and

$$\mu_2 = \bar{x}_2 + \left[-\sigma_2^2 \sum_{i=1}^{n} \frac{F_2'(c_{1i}, c_{2i})}{F_\theta(c_{1i}, c_{2i})} - \rho\sigma_1\sigma_2 \sum_{i=1}^{n} \frac{F_1'(c_{1i}, c_{2i})}{F_\theta(c_{1i}, c_{2i})} \right]. \tag{A2}$$

For unknown σ_1^2, σ_2^2, and ρ, three more estimating equations, in addition to equations (A1) and (A2), are needed to compute the estimators. Take partial derivatives of $\log[L_{cn}(\theta)]$ with respect to σ_1^2, σ_2^2, and ρ, where $L_{cn}(\theta)$ is the conditional likelihood, and then set these partial derivatives to zero. Some algebra yields

$$\sigma_1^2 = \frac{n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)^2 - \rho\sigma_2^{-1}\sigma_1 n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)(x_{2i} - \mu_2)}{(1 - \rho^2)(1 + 2\sigma_1^2 H_3)}, \tag{A3}$$

$$\sigma_2^2 = \frac{n^{-1}\sum_{i=1}^{n}(x_{2i} - \mu_2)^2 - \rho\sigma_1^{-1}\sigma_2 n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)(x_{2i} - \mu_2)}{(1 - \rho^2)(1 + 2\sigma_2^2 H_4)}, \tag{A4}$$

and

$$\rho = \frac{\sigma_1^{-1}\sigma_2^{-1} n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)(x_{2i} - \mu_2) - (1 - \rho^2)H_5}{2(1 + \sigma_1^2 H_3 + \sigma_2^2 H_4) - 1}, \tag{A5}$$

where

$$H_j = n^{-1}\sum_{i=1}^{n}[h_j(c_{1i}, c_{2i})/F_\theta(c_{1i}, c_{2i})], \; j = 3, 4, 5 .$$

Here h_1, h_2, and h_3 are the partial derivatives of $F_\theta(c_1, c_2)$, with respect to σ_1^2, σ_2^2, and ρ, and (c_{1i}, c_{2i}), $i = 1,...,n$. are the ages at interview of the observed APCPs.

It is clear that there is no explicit expression for $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho})$. However, $\hat{\theta}$ can be computed iteratively via equations (A1)–(A5). Specifically, computation can proceed as follows: (*i*) Start from an initial guess for $\hat{\theta}$. A convenient choice is the sample means, sample variances, and sample correlation coefficient of the observed ages at onset. (*ii*) Use equations (A1)–(A5) to obtain an updated value of the initial guess and repeat this process until the desired convergence criterion is satisfied.

We now give expressions for equations (A1)–(A5) in terms of standard normal distributions, which are convenient to compute. Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the density function and cumulative distribution function of the univariate standard normal distribution, respectively. Let $\phi(\cdot, \cdot)$ and $\Phi(\cdot, \cdot)$ be the density function and cumulative distribution function of the bivariate standard normal distribution, respectively. Define

$$k_1 = \frac{c_1 - \mu_1}{\sigma_1} \ , \ k_{1i} = \frac{c_{1i} - \mu_1}{\sigma_1} \ ,$$

$$k_2 = \frac{c_2 - \mu_2}{\sigma_2} \ , \ k_{2i} = \frac{c_{2i} - \mu_2}{\sigma_2} \ ,$$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{\phi(k_{1i})\Phi\left[(k_{2i} - \rho k_{1i})/\sqrt{1-\rho^2}\right]}{\Phi(k_{1i}, k_{2i})} \ ,$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\phi(k_{2i})\Phi\left[(k_{1i} - \rho k_{2i})/\sqrt{1-\rho^2}\right]}{\Phi(k_{1i}, k_{2i})} \ ,$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n} \frac{k_{1i}\phi(k_{1i})\Phi\left[(k_{2i} - \rho k_{1i})/\sqrt{1-\rho^2}\right]}{\Phi(k_{1i}, k_{2i})} \ ,$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n} \frac{k_{2i}\phi(k_{2i})\Phi\left[(k_{1i} - \rho k_{2i})/\sqrt{1-\rho^2}\right]}{\Phi(k_{1i}, k_{2i})} \ ,$$

and

$$m_5 = \frac{1}{n}\sum_{i=1}^{n} \frac{\phi(k_{1i}, k_{2i})}{\Phi(k_{1i}, k_{2i})} \ .$$

After some straightforward but tedious calculations, it can be verified that equations (A1)–(A5) can be written as

$$\mu_1 = \bar{x}_1 + \sigma_1(m_1 + \rho m_2) \ ,$$

$$\mu_2 = \bar{x}_2 + \sigma_2(m_2 + \rho m_1) \ ,$$

$$\sigma_1^2 = \frac{n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)^2 - \rho\sigma_2^{-1}\sigma_1 n^{-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)(x_{2i} - \mu_2)}{(1-\rho^2)(1-m_3)} \ ,$$

$$\sigma_2^2 = \frac{n^{-1}\sum_{i=1}^{n}(x_{2i}-\mu_2)^2 - \rho\sigma_1^{-1}\sigma_2 n^{-1}\sum_{i=1}^{n}(x_{1i}-\mu_1)(x_{2i}-\mu_2)}{(1-\rho^2)(1-m_4)} ,$$

and

$$\rho = \frac{\sigma_1^{-1}\sigma_2^{-1}n^{-1}\sum_{i=1}^{n}(x_{1i}-\mu_1)(x_{2i}-\mu_2) - (1-\rho^2)m_5}{1-m_3-m_4} .$$

If it is assumed that $\sigma_1 = \sigma_2 = \sigma$, then the likelihood equations become

$$\mu_1 = \bar{x}_1 + \sigma(m_1 + \rho m_2) , \tag{A6}$$

$$\mu_2 = \bar{x}_2 + \sigma(m_2 + \rho m_1) , \tag{A7}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\mu_1)^2 - 2\rho\sum_{i=1}^{n}(x_{1i}-\mu_1)(x_{2i}-\mu_2) + \sum_{i=1}^{n}(x_{2i}-\mu_2)^2}{n(1-\rho^2)(2-m_3-m_4)} , \tag{A8}$$

and

$$\rho = \frac{\sigma^{-2}n^{-1}\sum_{i=1}^{n}(x_{1i}-\mu_1)(x_{2i}-\mu_2) - (1-\rho^2)m_5}{1-m_3-m_4} . \tag{A9}$$

When the sample size is small, the correlation coefficient from the above updating scheme may not be in the interval $(-1, 1)$. In such a case, we can use a profile-likelihood approach, using a grid in $(-1, 1)$, the range of ρ. For each point in this grid, we maximize the likelihood L_{cn} with respect to (μ_1, μ_2, σ) using equations (A6)–(A8). Let $\hat{\rho}$ be the value at which the maximum of these likelihood values is attained. Then the maximum likelihood estimate (MLE) of the correlation coefficient is $\hat{\rho}$ and the MLE of (μ_1, μ_2, σ^2) is the corresponding $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)$ that maximizes $L_{cn}(\hat{\rho}, \mu_1, \mu_2, \sigma^2)$, with respect to (μ_1, μ_2, σ^2).

Appendix B

Simulation of Type I and Type II Error Rates

Type I Error Simulation Procedures

Random grandparent-parent-child current-age triplets (c_1, c_2, c_3) were generated. Loosely following the procedures described by Heiman et al. (1996), the grandparent's current age, c_1, was generated as a uniform $U(80, 90)$ random variable. Then, a random difference d_1 was generated from a $U(20–30)$ distribution, and the parent's current age, c_2, was assigned as $c_1 - d_1$. The same procedure was repeated to obtain a random difference d_2, and the child's current age, c_3, was set to $c_2 - d_2$. Again, following Heiman et al., we generated ages at onset from a single bivariate normal distribution, that is, under the null hypothesis of no AOA, with $\rho = 0$. Thus, there were two types of pairs with respect to current ages: grandparent-parent, or "old," pairs (c_1, c_2) and parent-child, or "young," pairs (c_2, c_3).

In order to generate generalized single ascertainment (true random sampling of APCPs) from this universe of three-person pedigrees, a current-age pair was selected at random, with equal probability that it was an old or young pair, and the pair was included in a data set if and only if both the parent's and child's ages at onset were less than the corresponding current ages. Note that, under this scenario, even though 50% of the generated pairs are young, the proportion of young pairs in the data is expected to be far less than 50%, since younger individuals are less likely to be affected than older ones, with the extent of this effect dependent on the generating parameters of the age-at-onset distribution.

Then, in order to model the effects of oversampling young pairs (deflating the current age distribution), we used a slightly different procedure. The same generating distributions were used, but for these models, the ascertainment procedure was as follows: (*i*) an old or young pair of current ages was selected; (*ii*) a random pair of ages at onset was generated for the current-age pair; and (*iii*) if both the parent and child ages at onset were less than the respective current ages (i.e., both were affected), the pair was included in a data set; otherwise, a new pair of ages at onset was generated for this same current-age pair. This last process was repeated until an APCP with the given current ages was obtained, and the entire process was repeated until enough pairs were generated. Thus by fixing the probability of an old versus young current-age pair at step (*i*), we were able to control the proportion of young pairs in the ascertained data, while maintaining the level of truncation of the age-at-onset distribution appropriate to the given observed current ages.

Power Simulation Procedures

We generated random parent-child current-age pairs (c_1, c_2) and assigned phenotypes (affected/unaffected) on the basis of a mixture of two bivariate normal age-at-onset distributions, with mixing proportion $\frac{1}{2}$. Data shown are for $\sigma_1 = \sigma_2 = 5$ and for intradistributional correlation coefficients $\rho_1 = \rho_2 = 0$ (however, see Discussion; this does not mean that parent-child ages at onset are uncorrelated in the resulting data). A randomly generated pair was included in a data set if and only if both the parent's and child's ages at onset were less than the corresponding current age, that is, the data exhibit bivariate right truncation but are not subject to ascertainment bias. The two constituent distributions were allowed to differ only in terms of the interdistributional distance $\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22}$. "Corresponding" single-distribution models were generated under a single age-at-onset distribution with the same overall mean parent-child age-at-onset difference and with marginal variances equal to those under each of the mixing distributions (i.e., in the model shown, all marginal SDs were 5).

References

Bassett AS, Honer WG (1994) Evidence for anticipation in schizophrenia. Am J Hum Genet 54:864–870

Bassett AS, Husted J (1997) Anticipation or ascertainment bias in schizophrenia? Penrose's familial mental illness sample. Am J Hum Genet 60:630–637

Bonifati V, Fabrizio E, Vanacore N, Demari M, Meco G (1995) Familial Parkinson disease: a clinical genetic analysis. Can J Neurol Sci 22:272–279

Brinkman RR, Mezei MM, Theilmann J, Almqvist E, Hayden MR (1997) The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. Am J Hum Genet 60:1202–1210

Ewens WJ (1982) Aspects of parameter estimation in ascertainment sampling schemes. Am J Hum Genet 34:853–865

Heiman GA, Hodge SE, Wickramaratne P, Hsu H (1996) Age-at-interview bias in anticipation studies: computer simulations and an example with panic disorder. Am J Med Genet (Neuropsychiatr Genet) 6:61–66

Heimbuch RC, Matthysse S, Kidd K (1980) Estimating age-of-onset distributions for disorders with variable onset. Am J Hum Genet 32:564–574

Hodge SE (1985) Family-size distribution and Ewens' equivalence theorem. Am J Hum Genet 37:166–177

——— (1988) Conditioning on subsets of the data: applications to ascertainment and other genetic problems. Am J Hum Genet 43:364–373

Hodge SE, Vieland VJ (1996) The essence of single ascertainment. Genetics 144:1215–1223

Hodge SE, Wickramaratne P (1995) Statistical pitfalls in detecting age-of-onset anticipation: the role of correlation in studying anticipation and detecting ascertainment bias. Psychiatr Genet 5:43–47

Horwitz M, Goode EL, Jarvik GP (1996) Anticipation in familial leukemia. Am J Hum Genet 59:990–998

Huang J, Vieland VJ (1997) A new statistical test for age of onset anticipation: application to bipolar disorder. In: Goldin LR, Bailey-Wilson JE, Borecki IB, Falk CT, Goldstein AM, Suarez BK, MacCluer JW (eds) Genetic Analysis Workshop 10: detection of genes for complex traits. Genet Epidemiol 14:1091–1096

Kalbfleisch JD, Sprott DA (1970) Application of likelihood methods to models involving large numbers of parameters. J R Stat Soc 32:175–208

McInnis MG (1996) Anticipation: an old idea in new genes. Am J Hum Genet 59:973–979

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR Jr (1993) Anticipation in bipolar affective disorder. Am J Hum Genet 53:385–390

Morton NE (1959) Genetic tests under incomplete ascertainment. Am J Hum Genet 11:1–16

Myers RH, Cupples LA, Schoenfeld M, D'Agostino RB, Terrin NC, Goldmakher N, Wolf PA (1985) Maternal factors in onset of Huntington disease. Am J Hum Genet 37:511–523

Paterson AD, Kennedy JL, Petronis A (1996) Evidence for genetic anticipation in non-Mendelian diseases. Am J Hum Genet 59:264–268

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugen 14:125–132

Slager SL, Vieland VJ (1997) Investigating the numerical effects of ascertainment bias in linkage analysis: development of methods and preliminary results. In: Goldin LR, Bailey-Wilson JE, Borecki IB, Falk CT, Goldstein AM, Suarez BK,

MacCluer JW (eds) Genetic Analysis Workshop 10: detection of genes for complex traits. Genet Epidemiol 14: 1119–1124

Stene J (1977) Assumptions for different ascertainment models in human genetics. Biometrics 33:523–527

Stuart A, Ord JK (1991) Kendall's advanced theory of statistics. Vol 2. Oxford University Press, New York

Vieland VJ, Hodge SE (1995) Inherent intractability of the ascertainment problem for pedigree data: a general likelihood framework. Am J Hum Genet 56:33–43

——— (1996) The problem of ascertainment for linkage analysis. Am J Hum Genet 58:1072–1084

Wald A (1943) Tests of statistical hypothesis concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, et al (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. Am J Hum Genet 56:99–105