Genetic survival analysis of age-at-onset of bipolar disorder: evidence for anticipation or cohort effect in families

Peter M. Visscher, M. Hossein Yazdi, Alan D. Jackson, Martin Schalling, Kerstin Lindblad, Qui-Ping Yuan, David Porteous, Walter J. Muir and Douglas H.R. Blackwood

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh, UK; Department of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK; Neurogenetics Unit, Department of Molecular Medicine, Karolinska Institutet, Karolinska Hospital, Stockholm, Sweden; Department of Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK

Correspondence to Dr Peter Visscher, The University of Edinburgh, Institute of Cell, Animal and Population Biology, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: peter.visscher@ed.ac.uk

Received 19 January 2001; accepted 10 July 2001

Age-at-onset (AAO) in a number of extended families ascertained for bipolar disorder was analysed using survival analysis techniques, fitting proportional hazards models to estimate the fixed effects of sex, year of birth, and generation, and a random polygenic genetic effect. Data comprised the AAO for 171 affecteds or age when last seen (ALS) for 327 unaffecteds, on 498 individuals in 27 families. ALS was treated as the censored time in the statistical analyses. The majority of individuals classified as affected were diagnosed with bipolar I and II (n = 103) or recurrent major depressive disorder (n = 68). In addition to the significant effects of sex and year of birth, a fitted ‘generation’ effect was highly significant, which could be interpreted as evidence for an anticipation effect. The risk of developing bipolar or unipolar disorder increased twofold with each generation descended from the oldest founder. However, although information from both affected and unaffected individuals was used to estimate the relative risk of subsequent generations, it is possible that the results are biased because of the ‘Penrose effect’. Females had a twofold increased risk in developing depressive disorder relative to males. The risk of developing bipolar or unipolar disorder increased by approximately 4% per year of birth. A polygenic component of variance was estimated, resulting in a ‘heritability’ of AAO of approximately 0.52. In a family showing strong evidence of linkage to chromosome 4p family 22, the ‘affected haplotype’ increased the relative risk of being affected by a factor of 46. In this family, there was strong evidence of a time trend in the AAO. When either year of birth or generation was fitted in the model, these effects were highly significant, but neither was significant in the presence of the other. For this family, there was no increase in trinucleotide repeats measured by the repeat expansion detection method in affected individuals compared with control subjects. Proportional hazard models appear appropriate to analyse AAO data, and the methodology will be extended to map quantitative trait loci (QTL) for AAO. Psychiatr Genet 11:129–137 © 2001 Lippincott Williams & Wilkins.

Keywords: bipolar disorder, unipolar disorder, anticipation, age-at-onset, survival analysis, proportional hazard

INTRODUCTION

Anticipation describes a pattern of inheritance that includes earlier age-at-onset and increased severity of symptoms in younger generations and is a feature of at least nine neurodegenerative diseases, including Huntington’s disease, fragile X, and myotonic dystrophy. The underlying dynamic mutations are unstable tri-nucleotide repeat sequences whose increasing size in succeeding generations or through sex-specific meiosis causes greater disruption to gene function, leading to increased severity of clinical symptoms and a reduction in age-at-onset (Paulson and Fischbeck, 1996). Several studies have reported anticipation in bipolar families selected to take part in linkage studies after controlling for several possible sources of bias (McInnis et al., 1993; Nylander et al., 1994; Grigoriou-Serbanescu et al., 1997; Mendlewicz et al., 1997; Ohara et al., 1998). However, Merette et al. (2000) found no evidence for anticipation in a group of Canadian bipolar families after controlling for an information bias that results from the availability of higher quality of information from younger generations compared with older subjects. Reports of anticipation have led to several studies of CAG repeat expansion in bipolar disorder using the repeat expansion detection (RED)
Age-at-onset (AAO) of a disease is a complex trait for genetic analysis because: (i) it is influenced by both the environment and by multiple genes; and (ii) it is measured on affected individuals only. There have been several quantitative genetic analyses proposed. The most common approach in linkage and association studies is to ignore age data, and to concentrate on affected versus unaffected individuals. A second method is to adjust the affected/unaffected scores for age (Zhu et al., 1997; Hanson et al., 1998). A third method (Daw et al., 1999) is to sample the AAO conditional on the actual age and inferred major genotype in segregation analysis. All these methods have drawbacks, because they do not use the information efficiently by ignoring the nature of the censored records. For example, the rationale for the method by Hanson et al. (1998) appears to be to adjust the binary trait affected/unaffected for age [whether AAO or age when last seen (ALS)], and thereby creating a quantitative trait. The authors use the population incidence to adjust for AAO or ALS, but for a large data set their method is equivalent to including age (AAO or ALS) in the model of analysis as a covariate. Depending on the true (unknown) genetic architecture underlying affected status (will a person get the disease) and the AAO (when will an affected person show the disease), adjusting for age may or may not be appropriate. Daw et al. (1999) assume that AAO is normally distributed (which is most likely not true for most late onset disorders or for bipolar disorder). Survival analysis methods, which properly take into account that ALS data are censored observations, have been used to model AAO for segregation analysis to detect a major gene (Abel and Bonney, 1990; Essioux et al., 1995) or to estimate residual effects to be used in subsequent linkage analysis (Zhu et al., 1997). These studies did not consider the general case of correlated AAO between family members due to polygenic effects, nor take phenotypic relationships between certain pairs of relatives (e.g. sister-sister and mother-daughter relationships in the breast cancer study of Essioux et al. (1995)) into account by fitting additional regressions in the model of analysis. Bivariate survival models based on correlated frailties have been applied to twin data to estimate a genetic component of longevity (Yashin and Iachine, 1995; Petersen, 1998). To our knowledge, quantitative genetic analyses of AAO to estimated variance components in complex pedigrees have not been reported previously.

In addition to the problem of the nature and distribution of the AAO trait, which makes a genetic analysis difficult, data from families may be subject to several types of bias (for example, Penrose, 1948; Hodge and Wickramaratne, 1995). These include reduced fertility in individuals who develop illness early, an information bias that results from the availability of higher quality of information from younger generations compared with older subjects (Merette et al., 2000), and a censoring bias when the younger generations are assessed over a shorter period of risk than older generations (Vieland and Huang, 1998). An ascertainment bias in favour of younger AAO offspring of affected parents is likely to be a feature of families collected for the purpose of linkage analysis (Huang and Vieland, 1997). These types of bias are likely to influence the inference on anticipation strongest when sampling affected parents and affected offspring only, but should have a smaller effect when dealing with extended pedigrees that are followed over time (Vieland and Huang, 1998).

The aim of the present study was to apply novel statistical methods based on survival analysis to genetically analyse AAO data for bipolar disorder in a number of extended families. The purpose of such analyses is to quantify the amount of genetic variation that exists in populations for AAO, and to estimate environmental risk factors simultaneously. In addition, we test for a possible anticipation effect, i.e. for evidence that the AAO is decreasing each generation. The application of survival analyses to test for anticipation in complex pedigrees was
recently suggested by Vieland and Huang (1998). In single large, multiply affected pedigrees, genetic and non-genetic heterogeneity is likely to be reduced. Therefore, we have also looked for evidence of anticipation in a single large Scottish bipolar family showing linkage between a bipolar and unipolar phenotype and a series of polymorphic markers mapped to a region on the short arm of chromosome 4 (Blackwood et al., 1996). We have also used the RED method to compare the trinucleotide repeat expansion length in affected and unaffected members of this family.

MATERIALS AND METHODS

Family data

Twenty-seven extended families were recruited for linkage studies. A summary of the data is presented in Table 1. Probands with bipolar 1 disorder were in-patients or out-patients at hospitals in the south of Scotland. All family members gave their consent to participate in genetic studies, provided a blood sample for DNA analysis, and were interviewed by a trained psychiatrist using the Schedule for Affective disorders and Schizophrenia — Lifetime version (Endicott and Spitzer, 1978). Diagnoses were made according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and RDC criteria after a review of hospital case records. Families were selected in which two first-degree relatives were diagnosed with bipolar disorder, or bipolar disorder and recurrent major depression. Pedigrees with apparent bilineal descent were not included. Data on AAO (affected individuals), ALS (unaffected individuals), year of birth, sex, and family were used from 498 individuals from 27 unipolar- and bipolar-affected extended families. Pedigree information was known and a total of 582 individuals were included in the pedigree file. From the pedigree information, the number of generations since the oldest founder was calculated to test for an anticipation effect in the analysis. One hundred and seventy-one individuals, year of birth, sex, and family were used for the statistical analyses on Family 22 only, 27 affected individuals with known haplotypes were used.

<table>
<thead>
<tr>
<th>TABLE 1. Summary of data on age-at-onset and age when last seen</th>
</tr>
</thead>
<tbody>
<tr>
<td>All families</td>
</tr>
<tr>
<td>Individuals in pedigree</td>
</tr>
<tr>
<td>Individuals with observations</td>
</tr>
<tr>
<td>Number affected</td>
</tr>
<tr>
<td>Age-at-onset (years)</td>
</tr>
<tr>
<td>Number unaffected</td>
</tr>
<tr>
<td>Age last seen (years)</td>
</tr>
</tbody>
</table>

aFor the statistical analyses on Family 22 only, 27 affected individuals with known haplotypes were used.

Statistical methods

We applied survival analysis methodology to model AAO, using correlated frailty models (for example, Ducrocq and Casella, 1996; Petersen, 1998; Hougaard, 1999) to estimate covariates and a polygenic variance component. All available pedigree information was used in the estimation of the variance component. Survival analysis models the trait of interest (here, AAO) properly as a function of time, and takes into account that some records are censored (affected individuals) and some are censored (unaffected individuals). The effects of factors such as sex, environmental time trend and genotype are estimated in terms of the risk of becoming affected.
We assumed that the risk (or hazard) of developing a disease follows a proportional hazard model, with the proposed mixed linear model as the exponential part of such a model. For an individual $k$, the hazard function is:

$$h(t|b,u) = h_0(t) \exp(x_k' b + u_k)$$

with $h(t|b,u)$ the hazard at time $t$ for an individual with covariates $x_k$, and polygenic value $u_k$. The hazard function describes the instantaneous rate of
‘failure’ at time t, given that a person was unaffected before t. In our application, the hazard function measures the risk of becoming affected at age t, given that an individual was not affected at age (t – 1). \( h_0(t) \) is the baseline hazard function at time t, which is an ‘average’ hazard function pertaining to all individuals in the population, and is similar to the average probability of becoming affected at age t, conditional on not being affected at age (t – 1). B is the vector of fixed effects and covariates.

The aforementioned model is called a mixed survival model or frailty model because the polygenic effects are random variables. The model assumes that there is an average baseline hazard function, \( h_0(t) \), which represents the ageing process in the population (or, in our case, the average risk of becoming affected as a function of time), and that covariates and polygenic values act multiplicatively to increase or decrease an individual’s hazard throughout his life. The terms \( \exp(u_k) \) are frailty terms. The baseline hazard function is usually described by assuming that the survival function (which describes the probability of being unaffected at a particular point in time) has an exponential or Weibull distribution. We will use the latter, since the exponential distribution is a special form of the Weibull distribution. So that we could, if necessary, perform a statistical test for a Weibull versus exponential fit. The baseline hazard function under the assumption of a Weibull survival function is:

\[
 h_0(t) = \lambda \rho(t)^{\rho - 1}
\]

The Weibull parameters \( \rho \) and \( \lambda \) define the shape and scale of the hazard function, respectively. If \( \rho > 1 \) (respectively, \( \rho < 1 \)) the hazard of an individual increases (respectively, decreases) with time. For example, if \( \rho = 2 \), then the relative hazard (risk) of individuals aged 60 and 20 is \( 60^{2-1}/20^{2-1} = 3 \), i.e. a threefold difference in risk. For \( \rho = 1 \), the baseline hazard is constant and the Weibull model reduces to an exponential regression model. Weibull models can also be viewed as special cases of accelerated (\( \rho > 1 \)) or decelerated (\( \rho < 1 \)) failure time models. The Weibull model is a parametric proportional hazards model. Estimation of fixed and random effects and of genetic parameters is much less demanding with a Weibull model than with a semi-parametric model, such as the popular Cox (1972) model, for which the baseline hazard function is left arbitrary. The Weibull model is also very flexible; in particular, when time-dependent covariates are included in the exponential part of the model.

If we assume that the vector of polygenic values has variance \( \text{var}(u) = A \sigma_u^2 \), then we have the same linear model as for normally distributed traits but the effects now act multiplicatively in the mixed survival (frailty) model. The proposed model fits into the general class of multivariate frailty models discussed by Petersen (1998). Essentially, we propose to put a genetic structure (additive relationship matrices) derived from quantitative genetics to account for correlated frailties (and, therefore, correlated failure times or AAO) between related individuals. For longevity traits in animal (livestock) populations, these models have been applied before including the polygenic component in the model (for example, Ducrocq and Casella, 1996).

All analyses were performed using the software package Survival Kit (Ducrocq and Sölkner, 1994). To estimate parameters in the model a Bayesian analysis is used, with flat priors for the fixed effects, genetic variance and Weibull parameter \( \rho \) (Ducrocq and Casella, 1996).

Models
Sex, family and generation were treated as fixed factors in the model, and year of birth was treated as a linear covariate. The rationale for including year of birth in the model was to account for a possible environmental trend (e.g. due to better diagnosis and reporting) in the data. Family effects were fitted to account for non-genetic differences in AAO between families. Initial analyses of treating families as fixed gave convergence problems because of a strong confounding between individual (genetic) effects and year-of-birth effects and the parameter \( \rho \) from the Weibull model. Therefore, the shape parameter \( (\rho) \) was first estimated from a model excluding any individual genetic effect, i.e. a completely fixed model, and held constant for subsequent analyses.

Family 22 was analysed individually, fitting a model with sex, generation, haplotype, and year of birth. Haplotype was coded as the presence or absence of the ‘affected haplotype’ from the Blackwood et al. (1996) study. In addition to these models, two fixed models fitting either year of birth or generation were analysed because these effects are expected to be highly correlated.

RESULTS
The estimate of the Weibull shape parameter \( (\rho) \) from the fixed model was 2.48. Hence, for two individuals with the same covariates (sex, year of birth,
family), the relative hazard (risk) of becoming affected at ages 50 and 20 years of age is $50.148/20^{1.48}$ = 3.9. This value of $\rho = 2.48$ was held constant for subsequent analyses to allow a more robust estimation of the remaining parameters. The full model, i.e. including a random individual genetic component, resulted in an estimated genetic variance component of 1.10 (i.e. the mode of the posterior distribution of the estimated variance component). The mean and standard deviation of this variance component were 1.16 and 0.20, respectively (see Table 2). An approximation of the heritability for failure time data using the Weibull model was recently derived (Yazdi et al., 2001) as $h^2 = \sigma^2/(\sigma^2 + 1)$, with $\sigma^2$ being the estimate of the additive genetic variance. For the present data, the estimate of the heritability is 0.52, i.e. 52% of the variation in AAO is due to genetic risk factors. A separate analysis excluding Family 22 resulted in similar estimates of fixed effects and genetic variance (results not shown), so the amount of genetic variance estimated in the complete data set is not an artefact of the quantitative trait loci (QTL) of large effect detected by Blackwood et al. (1996) in Family 22.

The effects of families were significant ($P < 0.0001$) both in the fixed model and in the full model, indicating differences in the risk of developing bipolar disorder across families. The estimates of individual family effects are not shown because most families have only few (<5) affected individuals, so the standard errors of family effects are large. Significant differences in the risk of developing affective disorder between families may be due to genetic, environmental effects, or a combination thereof.

The effects of sex, generation, and year of birth were $-0.77 (\pm 0.20)$, 0.86 (± 0.20), and 0.04 (± 0.01), respectively (see Table 2). From these results we conclude that males have an $\exp(-0.77) = 47\%$ smaller risk than females in developing affective (bipolar and unipolar) disorder. This effect was due to a large excess of both unipolar and bipolar depression in women ($n = 48$ for unipolar and $n = 66$ for bipolar) compared with men ($n = 20$ for unipolar and $n = 37$ for bipolar). A separate analysis in which either unipolars or bipolars were analysed revealed a significant increased risk for females of threefold in unipolars and twofold in bipolars (results not shown). Each generation from the oldest founder increases the risk by $\exp(0.86) = 2.4$-fold, which possibly points to an anticipation effect. The risk of becoming affected increases by $\exp(0.04) = 4\%$ each year.

The full fixed model for an analysis of Family 22 only resulted in a large estimated effect of the affected haplotype, 3.84 ± 0.75. Out of the 27 affected individuals with scored haplotypes, 25 carried the haplotypes determined by the Blackwood et al. (1996) study. This effect of haplotype on the hazard means that the risk of developing bipolar or unipolar disorder increased $\exp(3.84) = 46$-fold for those individuals carrying the identified haplotype from the analysis of Blackwood et al. (1996). The effect of sex was not significant for this (small) data set. When both year of birth and generation were fitted, neither of these highly correlated covariates were significant in the presence of the other (Table 2). However, fitting a fixed model with generation only (in addition to sex and haplotype) resulted in a highly significant effect ($P < 0.001$) of 1.21 for generation, i.e. an increased risk ratio of $\exp(1.21) = 3.4$ per generation. Similarly, when fitting year of birth but not generation, the significant effect of year of birth was 0.058 per year, i.e. an increased risk of $\exp(0.058)$, or 6% per year. These results suggest an average generation interval of $1.21/0.058 = 21$ years.

### Table 2. Estimates of fixed effects and variance components and their standard errors

<table>
<thead>
<tr>
<th></th>
<th>All families</th>
<th>Family 22</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects/covariates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (males versus female)</td>
<td>$-0.77 \pm 0.20$</td>
<td>$-0.57 \pm 0.40$ (NS$^a$)</td>
</tr>
<tr>
<td>Generation</td>
<td>$0.86 \pm 0.20$</td>
<td>$0.52 \pm 0.61$ (NS)</td>
</tr>
<tr>
<td>Year of birth</td>
<td>$0.04 \pm 0.01$</td>
<td>$0.04 \pm 0.03$ (NS)</td>
</tr>
<tr>
<td>Chromosome 4p haplotype</td>
<td>$0.57 \pm 0.03$</td>
<td>$3.84 \pm 0.56$</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additive genetic variance</td>
<td>$1.16 \pm 0.20$</td>
<td>$0.82 \pm 0.51$ (NS)</td>
</tr>
</tbody>
</table>

$^a$Non significant ($P > 0.05$).
type. Reactions were performed and analysed as described by Lindblad et al. (1995) using a (CTG)10 oligonucleotide giving a representation of bands on a polyacrylamide gel at an interval of 30 nucleotides. Reactions used 15 U Ampligase per reaction. Expanded RED products (120 base pairs or over) were observed in 5/7 bipolar and 4/8 control subjects, indicating no association of bipolar illness with expanded repeats in this family.

DISCUSSION

We have shown that survival analysis methods with random effect accounting for all genetic relationships can be used to analyse AAO data in complex pedigrees. It should be possible to extend these methods to QTL mapping for AAO, fitting both a polygenic and a QTL random component in the model of analysis, i.e. to perform a variance component analysis that properly takes account of the non-normal distribution of the data and of censoring.

One assumption of our model is that all individuals are in principle at risk of developing either unipolar or bipolar disorder. The same assumption was made by Daw et al. (1999) in their analysis of AAO of Alzheimer’s disease. This assumption may be reasonable in our study, because families were selected for their high incidence of the disorder. Nevertheless, in samples of the population there may be individuals that would never become seriously depressed, regardless of the environmental conditions. In that case, it would be better to attempt to model a proportion of individuals at zero risk in the data. Estimates of covariates and variance components would then only be relevant for the at-risk group. However, such an analysis would be difficult or impossible, because there is little (if any) information to distinguish between unaffected individuals who are at risk but have not developed the disorder (yet) and those that have a zero risk. Furthermore, even if a zero-risk proportion could be estimated, it may not affect our conclusions because the estimated risk of unaffected individuals with high censored time (ALS) would be very small (i.e. close to zero) anyway.

Both year of birth and generation number were highly significant in the analysis of all families. This could point to an environmental (year of birth or cohort effect) or biological time trend (generation or anticipation effect) in the reduction of AAO of bipolar disorder, or to a selection bias due to the ascertainment of families for linkage studies (Penrose, 1948). In addition to the possibility of a selection bias, the fitted effects of cohort (year of birth) and generation are highly confounded because older individuals are, on average, in the same generation group. In time, through the accumulation of additional data, this confounding should decrease.

Two of the most consistent findings in the epidemiology of depression are that the prevalence of depression is about twice as high among women as men and prevalence rates are higher among younger than older people (Kessler et al., 1993; Weissman et al., 1993). A year of birth effect (cohort effect) for bipolar and unipolar disorders has been well described in several countries (Klerman et al., 1985; Gershon et al., 1987; Wickramaratne et al., 1989; Joyce et al., 1990). These studies have detected a consistent trend for the AAO of affective disorders to become steadily younger over the past 50 years. This could be due to cultural factors such as more tolerant public attitudes to mental illness coupled with increased availability of effective treatments leading to earlier diagnosis. An epidemiological study that compared the prevalence of depression in 1952, 1970 and 1992 in a population within a single catchment area in Canada by Murphy et al. (2000) measured a stable overall prevalence of depression at the three time points, but within the overall constant rate there was a significant redistribution of cases by sex and age, with a recent increase in prevalence of depression in younger women. In the families included in the present study, the increased rate of affective disorder in women was found for both unipolar major depression and bipolar depression, and the men: women ratio of approximately 3:1 (unipolar disorder) and 2:1 (bipolar disorder) is in keeping with all the aforementioned studies.

The effects of sex and generation were not significant when year of birth was fitted in the model for the analysis of Family 22. However, the effects of year of birth and generation are highly confounded, and fitting either of these effects showed a highly significant effect. It is not clear whether the reduction in AAO illustrated in Figures 1 and 2 is due to censoring bias, an environmental time trend or a real biological phenomenon. This extended family has been studied over a long period of time, and the possible bias due to selection effects should diminish because all contemporary relatives are included in the analysis. There is no clear indication from national studies that the population incidence of bipolar disorder has increased over time. Hence, it is possible that the data on Family 22 (and indeed the other families) indicate real evidence of anticipation. There was no increase in trinucleotide repeat expansions detected by the RED method. It is probable that...
bipolar disorder shows genetic heterogeneity and that anticipation is found in only a proportion of families. However, the sample size that was used to test the hypothesis of an increased proportion of trinucleotide repeats in affected individuals was very small, so we cannot rule out that real differences exist between the affected and unaffected individuals. Nevertheless, we can conclude that there is no evidence in this family that all affecteds have a larger number of trinucleotide repeats than unaffected individuals.

For linkage analysis, the application of the survival model allows one to test the hypotheses that some genetic loci influence the risk of developing a disorder (e.g., certain genotypes are more susceptible, at any age), independent of the AAO, while other loci influence the AAO. This can be achieved by expanding the models to include stratification (according to, for example, genotype) and time-dependent covariates.

In conclusion, we have successfully applied survival analysis techniques to estimate genetic variation of AAO while simultaneously estimating other risk factors. There was some evidence of anticipation, but we cannot reject a possible bias due to the well-known ‘Penrose effect’. We plan to extend the analysis to incorporate marker information and estimate QTLs for AAO.

Acknowledgements

The authors thank the UK Biotechnology and Biological Sciences Research Council, Chief Scientist Office of the Scottish Executive, EC (Biomed2 grant EC BMH4-CT97-2466) and Akzo Nobel (Or ganon) for support, and Vincent Ducrocq and Robin Thompson for helpful discussions. They thank two reviewers for useful comments.

REFERENCES


