Psychiatric Genetics (1996), **6**, 61–66

Age-at-interview bias in anticipation studies: computer simulations and an example with panic disorder

G.A. Heiman, S.E. Hodge, P. Wickramaratne and H. Hsu

Columbia University and New York State Psychiatric Institute, Division of Clinical and Genetic Epidemiology, New York, NY, USA

Correspondence to: Gary A. Heiman, MSc, New York State Psychiatric Institute, Clinical and Genetic Epidemiology, Unit 14, 722 West 168th Street, New York, NY 10032, USA

The phenomenon of anticipation has received considerable interest recently, especially among researchers investigating the genetics of psychiatric disorders. Anticipation can involve an earlier age at onset, greater severity, and/or a higher number of affected individuals in successive generations within a family. There is some controversy concerning detection of age-at-onset anticipation, due to problems of sampling bias, which may account for the phenomenon by preferentially sampling either *later*-onset parents or *earlier*-onset children. One source of bias that has not been explicitly investigated is differential age at interview between parent and child, such that parents have passed through more of the risk period than their offspring. We conducted a computer simulation study of affected parent–child pairs to determine whether, for realistic age-at-onset and age-at-interview distributions, this source of bias is a serious one. Our results show that the timing of diagnostic assessment can strongly affect the ascertainment of parent–child pairs, to produce a severely biased sample exhibiting apparent anticipation. Under realistic assumptions, an investigator may face a greatly increased risk of false positives (i.e. detecting anticipation when none exists). For example, a nominal 5% significance level may correspond to true p values as high as 50% or even approaching 100%. We conclude with an application to existing data on panic disorder.

Keywords: Age at onset – Anticipation – Ascertainment bias – Panic disorder

INTRODUCTION

The term anticipation has been given to the phenomenon that in a few disorders there seem to be an earlier age at onset, greater severity, and a higher number of affected individuals in successive generations within a family. This term was first used in 1910 by Mott in describing his observation that the children of the 'insane' seemed themselves to have an earlier onset of insanity (Mott, 1910). Since that time, there has been considerable controversy as to the existence and nature of this phenomenon (Penrose, 1948; Höweler *et al.*, 1989; Harper *et al.*, 1992; Petronis and Kennedy, 1995; Hodge and Wickramaratne, 1995).

Due to the peculiar inheritance pattern produced by anticipation, researchers investigating the genetics of psychiatric disorders have also begun to look at this phenomenon. Since finding the gene(s) for psychiatric disorders has been particularly difficult, researchers are beginning to suspect that anticipation might help account for the 'complex' modes of inheritance. In the last few years a number of studies have been published purporting to demonstrate anticipation in bipolar disorder and schizophrenia (McInnis *et al.*, 1993; Asherson *et al.*, 1994; Bassett and Honer, 1994; Nylander *et al.*, 1994). Some of these studies have mentioned the potential biases in assessing anticipation, but they either minimized their importance or could not describe their magnitude of effect.

We will not review the history and controversy in detail, as that has been done elsewhere by others (Harper *et al.*, 1992; Petronis and Kennedy, 1995). For the purposes of this paper, two issues are relevant. One is whether anticipation, as a biological phenomenon, really does exist. In recent years, this question has been answered in the affirmative. That is, a biological mechanism, namely, unstable trinucleotide repeat mutation, has been identified which can give rise to anticipation, and the phenomenon has been confirmed for several diseases such as myotonic dystrophy, fragile X syndrome, Huntington

disease, facioscapulohueral muscular dystrophy, and Machado-Joseph disease (Ashizawa *et al.*, 1992; Caskey *et al.*, 1992; Ross *et al.*, 1993; Willems, 1994; Maciel *et al.*, 1995; Petronis and Kennedy, 1995; Zatz *et al.*, 1995). The second issue is that sampling bias still remains a concern. Such bias can occur in a number of ways (Penrose, 1948; Harper *et al.*, 1992; Petronis and Kennedy, 1995), but in the context of age-at-onset anticipation, the underlying principle is always that any ascertainment scheme which preferentially samples either *later*-onset parents or *earlier*-onset children may lead to false evidence of anticipation.

There exists one mechanism, which no one to our knowledge has addressed explicitly, that may lead to a preferential ascertainment of pairs exhibiting false evidence of anticipation. This mechanism involves the timing of the interviews or diagnostic assessment. Usually most of the individuals participating in a study are interviewed during the same specified time interval. Since parents are, by definition, one generation older than their children, this inevitably leads to a *difference* in age at interview between parent and child. *A priori*, clearly since children have had fewer years at risk, this difference in age at interview will induce an apparent difference in age at onset between parents and children, even when none exists. However, no one has quantified the magnitude of this effect, or determined its impact on significance levels when testing for anticipation.

We investigated to what extent the differential timing of the diagnostic assessment of parents and children will lead to a biased sample in anticipation studies. We confined our analysis to only one aspect of anticipation, that of a progressively earlier age at onset in successive generations, and we considered only studies of affected parent–child pairs (as have been performed by, for example, Asherson *et al.*, 1994; Deighton *et al.*, 1994; Sano *et al.*, 1994; Loyd *et al.*, 1995).

METHODS

We simulated datasets consisting of 100 affected parent–child pairs each. For each set of input parameters (see below), 100 such datasets were simulated. Input parameters: a probability density function was assigned to each of the following input variables: the parent's age at onset (PAO), the child's age at onset (CAO), the parent's age at interview (PAI) and the parent–child age difference (AD). The PAO and CAO were assumed to be normally distributed. The mean PAO was fixed at 20, 30 or 40 years. The mean CAO was also fixed at 20, 30 or 40 years, but never exceeded the mean PAO. The standard deviation (S.D.) of age at onset was fixed at 5 or 10 years and was always equal for parent and child. The PAI was assigned a uniform distribution: $U(30,75)$, $U(40,75)$, or $U(50,75)$*. The AD between parent and child in each pair was assigned a uniform distribution $U(20,35)$. In total, there were 36 different combinations (6 PAO–CAO combinations × 2 S.D. × 3 PAI distributions). For each combination, 100 datasets were generated, consisting of 100 pairs each. If either the parent or the child had an age at onset greater than that individual's age at interview, then that pair was excluded, and the computer continued to generate pairs until there were 100 pairs. Since individuals with an onset after being interviewed would not be detected, these individuals would not be included in anticipation studies. (Due to space limitations, we omit tables showing the number of pairs that were excluded for each combination. This information is available upon request.)

For each of the 36 combinations of input variables, we tabulated the observed mean PAO and CAO, averaged over the 100 datasets. This was done in order to evaluate the separate impact of the age at interview on the observed parent and child ages at onset. For the 18 combinations in which the true mean PAI equalled the true mean CAO (i.e. not true anticipation), we also performed a paired *t*-test of anticipation. This was performed in order to estimate the probability of finding a significant outcome when there is actually no anticipation. There was no need to correct for multiple tests since the analysis was done only once for each simulation.

RESULTS

Two types of parent–child pairs were generated in this study. In half of the simulations, the true means of the underlying ages at onset were the same for both parent and child (i.e. no anticipation). In the other half, anticipation did exist between the ages at onset in the parent and the child. The results of these simulations are presented in Table I. The table shows the average of the observed parental and child age at onset in the 100 datasets for each combination of input parameters. In every set of 100 datasets generated, the observed mean PAO was equal to or only slightly less than the true PAO, whereas the observed mean CAO was always depressed from its true value.

*The uniform distribution $U(a,b)$ on the interval $[a,b]$ is defined as follows:

$$f(x) = \begin{cases} 1/(b-a) & \text{for } x \in [a,b] \\ 0 & \text{elsewhere} \end{cases}$$

TABLE I. Observed parental age at onset (**bold type**) and child age at onset (regular type) – mean of sample means from 100 replicates of 100 parent–child pairs each

True mean age at onset (parent/child)	True standard deviation	True age-at-interview distribution for parents		
		$U(30,75)$	$U(40,75)$	$U(50,75)$
A. When the true situation is no anticipation				
20/20	$\sigma = 5$	**20.0** 19.1	**20.0** 19.2	**20.0** 19.6
	$\sigma = 10$	**20.0** 17.4	**19.9** 17.9	**20.1** 18.8
30/30	$\sigma = 5$	**30.0** 28.5	**30.0** 28.6	**29.9** 28.7
	$\sigma = 10$	**29.8** 24.9	**29.9** 25.1	**29.9** 25.8
40/40	$\sigma = 5$	**40.0** 37.3	**39.9** 37.3	**40.0** 37.3
	$\sigma = 10$	**39.7** 31.8	**39.7** 31.9	**39.6** 32.2
B. When the true situation is anticipation				
30/20	$\sigma = 5$	**30.1** 19.2	**30.0** 19.2	**30.0** 19.5
	$\sigma = 10$	**29.5** 17.4	**29.7** 17.9	**29.9** 18.7
40/20	$\sigma = 5$	**39.8** 19.3	**39.9** 19.3	**40.0** 19.7
	$\sigma = 10$	**38.7** 18.0	**39.1** 18.0	**39.5** 18.7
40/30	$\sigma = 5$	**40.0** 28.6	**39.9** 28.6	**40.0** 28.8
	$\sigma = 10$	**39.5** 25.2	**39.4** 25.2	**39.5** 25.9

TABLE II. Number of datasets, out of 100, with nominal p-value ≤ 0.05, when the true situation is no anticipation

True mean age at onset (parent/child)	True standard deviation	True age-at-interview distribution for parents		
		$U(30,75)$	$U(40,75)$	$U(50,75)$
20/20	$\sigma = 5$	28	15	3
	$\sigma = 10$	47	32	18
30/30	$\sigma = 5$	54	49	42
	$\sigma = 10$	94	95	91
40/40	$\sigma = 5$	98	96	98
	$\sigma = 10$	100	100	100

Hence, CAO *always* appeared to be less than PAO, whether or not the true CAO was less than the true PAO. Interestingly, the age-at-interview distribution appeared to have relatively little impact on the observed PAO and CAO.

Table II shows the number of times the p-value is 0.05 or less, i.e. the number of times the null hypothesis is rejected, from the 100 paired t-tests, for each combination in which there was no true anticipation. Only combinations in which no true anticipation existed are presented, so as to demonstrate the risk of a false positive conclusion. If the test really had size 0.05, we would expect to observe approximately 5 out of 100 rejections in each cell in Table II. However, in all but one of these cells, the numbers of rejections were much greater than 5, thus indicating that the true significance level greatly exceeded the nominal one. The magnitude of this phenomenon tends to increase as the true mean ages at onset, the standard deviations, and the range of the parents' age at interview increase.

AN EXAMPLE: PANIC DISORDER

To illustrate how difficult it may be to assess the impact of this phenomenon in a real dataset, we ascertained parent–child pairs from two existing datasets on panic disorder and performed a paired t-test on the observed ages at onset. We chose panic disorder because, like bipolar disorder and schizophrenia, panic disorder is believed to be familial (Crowe *et al.*, 1983; Weissman *et al.*, 1993), and there is evidence that at least some cases of panic disorder have a genetic etiology (Kendler *et al.*, 1993; Torgersen, 1983). Additionally, the mode of inheritance is not well understood (Vieland *et al.*, 1993).

The two datasets, one from a family study and the other from a linkage study, were collected as part of an effort to investigate the genetics of panic disorder. The datasets have been described elsewhere (Knowles *et al.*, 1993; Weissman *et al.*, 1993). Briefly, in the family study, probands with and without panic disorder were ascertained from both clinical and epidemiological samples, and their first-degree relatives were interviewed at about the same time for the presence of panic disorder. A best-estimate diagnosis was made on each individual. The best-estimate diagnosis on uninterviewed relatives was based on family history information. In the linkage study, families were chosen for being 'highly loaded' with panic disorder. All first-degree relatives of an individual with panic disorder were interviewed and a best-estimate diagnosis was made.

We analysed the samples separately for the presence of anticipation. All parent–child affected pairs were identified from each study for inclusion into the anticipation sample. Each sample was partitioned in four ways: full vs partial samples and sibship averaged vs sibship not averaged. In the full sample, both directly interviewed and uninterviewed subjects were included. In the partial sample, only directly interviewed subjects were included. When more than one sibling in a sibship was affected, the siblings' ages at onset were either averaged and treated as a single parent–child pair ('sibship averaged') or were treated as multiple independent parent–child pairs ('sibship not averaged'). The results of these analyses are presented in Table III. In every breakdown of the two samples, the PAO was greater than the CAO. For example, the mean ages at onset for parents and

TABLE III. Mean differences, in years, between parental and child ages at onset observed in panic disorder dataset

	Full sample	Partial sample
A. Family study sample:		
Sibship not averaged	9.3* (n = 34)	6.4* (n = 22)
Sibship averaged	8.4* (n = 25)	5.9 (n = 17)
B. Linkage study sample:		
Sibship not averaged	8.3* (n = 71)	6.8* (n = 55)
Sibship averaged	6.1* (n = 29)	4.1 (n = 23)

*$p \leq 0.05$ (paired t-test).

children in the family study, using the full sample and sibship not averaged, were 30.1 and 20.8 years, respectively. In six out of eight analyses, the difference between the parents and the children was significant at the 0.05 level. In the other two analyses, the mean age differences were in the same direction (i.e. exhibiting anticipation) but were based on small numbers. *Prima facie*, our results in Table III could be taken as providing suggestive evidence for age-at-onset anticipation in panic disorder. However, the findings in Table II caution us that this apparent anticipation could be merely an artifact arising from differential age at interview between parents and children. The mean age differences observed in our panic disorder data are consistent with some of the realistic values in Table I, part A (Klerman *et al.*, 1991). We want to reiterate that we have included the panic data only for purposes of illustration. The goal of this illustration is to point out the difficulty in assessing anticipation when there is differential timing of diagnostic assessment of parents and children.

DISCUSSION

The results of this computer simulation show that the true probability of a type I error can be much greater than the nominal probability when samples of parent–child affected pairs are ascertained at the same time. When parent–child pairs are ascertained for an anticipation study, only families in which both members are affected are being identified. Parent–child pairs in which one member of the pair has not yet become affected will not be ascertained. The magnitude of this phenomenon tends to increase as the true mean ages at onset, the standard deviations, and the range of the parents' age at interview increase.

The samples used in anticipation studies are usually ascertained from existing linkage or family study datasets. Investigators select parent–child affected pairs from these datasets and analyse this subsample for anticipation. Therefore, the selected pairs are interviewed before they are identified for inclusion in the anticipation study subsample, so that only pairs in which both members are affected are identified. Since both members of the pair are interviewed at approximately the same time, the parents have had many more years at risk than their children. The probability is therefore greater for a late-onset parent to be affected at the time of interview than a late-onset child.

We have shown that ascertaining and interviewing parent–child pairs at the same time may lead to an increase in type I error. In theory, there are two ways to adjust for this error. The first would be to ascertain only parent–child pairs in which *both* members are through the age at risk. In practice, however, it would be difficult to find a sufficient sample meeting this requirement. However, the results of our simulation studies indicate that for disorders with an early mean age at onset (e.g. 20 years) coupled with a small variance, this method of sampling appears to introduce very little bias if the lower bound of the age at interview of the child is greater than the mean age at onset. The second would be to ascertain probands independent of family history. These probands would be followed prospectively for many years until both the parents and children were through the age at risk. This long-term design would be very expensive.

Within the range of realistic age-at-interview distributions that we examined, the actual limits of the uniform distribution had relatively little impact on the observed PAO and CAO. We speculate that changes in the lower limit of PAI (30, 40, 50) have less of an effect on the observed PAO and CAO than the upper limit (75) does.

Our findings do not apply to other types of anticipation studies, for example, studies of increased severity in subsequent generations, or studies involving whole families where unaffected siblings of both the parent's and children's generation are also included in the analysis. In this latter situation, the use of survival analysis can overcome the bias due to differential timing of the diagnostic assessment. However, the use of survival analysis may not overcome the other potential biases such as recall bias or underreporting of remote events (Simon and VonKorff, 1992). Therefore, due to the many potential biases inherent in anticipation studies, we are skeptical about assessing anticipation purely on statistical grounds (Hodge and

Wickramaratne, 1995). Evidence of anticipation can best be demonstrated after the genes have been found (Ashizawa *et al.*, 1992).

In summary, we have documented that the timing of diagnostic assessment can affect the ascertainment of parent–child pairs to produce a biased sample that exhibits anticipation. We have also illustrated the magnitude of this effect and have demonstrated that under realistic assumptions, an investigator may face a greatly increased risk of false positives (i.e. detecting anticipation when none exists). Our aim in this paper is to point out that the timing of the diagnostic assessment may affect anticipation studies, not to set guidelines for conducting this analysis or to suggest which significance level should be attained.

Acknowledgements

We thank Abby J. Fyer, MD and Myrna M. Weissman, PhD for the use of their panic disorder data. This work was supported in part by grants DK-31813, MH-48858, MH-43878, MH-36197, MH-55376, MH-37592 and MH-28274 from the National Institutes of Health.

REFERENCES

Asherson P, Walsh C, Williams J, Sargeant M, Taylor C, Clements A, Gill M, Owen M and McGuffin P (1994) Imprinting and anticipation: are they relevant to genetic studies of schizophrenia. *British Journal of Psychiatry*, **164**, 619–624.

Ashizawa T, Dunne CJ, Dubel JR, Perryman MB, Epstein HF, Boerwinkle E and Hejtmancik JF (1992) Anticipation in myotonic dystrophy. Statistical verification based on clinical and haplotype findings. *Neurology*, **42**, 1871–1883.

Bassett A and Honer W (1994) Evidence for anticipation in schizophrenia. *American Journal of Human Genetics*, **54**, 864–870.

Caskey CT, Pizzuti A, Fu Y, Fenwick RG and Nelson DL (1992) Triplet repeat mutations in human disease. *Science*, **256**, 784–789.

Crowe RR, Noyes R, Pauls DL and Slyman DJ (1983) A family study of panic disorder. *Archives of General Psychiatry*, **40**, 1065–1069.

Deighton C, Heslop P, McDonagh J, Walker D and Thomson G (1994) Does genetic anticipation occur in familial rheumatoid arthritis? *Annals of Rheumatic Disease*, **53**, 833–835.

Harper PS, Harley HG, Reardon W and Shaw DJ (1992) Review article: anticipation in myotonic dystrophy: new light on an old problem. *American Journal of Human Genetics*, **51**, 10–16.

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

Höweler CF, Busch HFM, Geraedts JPM, Niermeijer MF and Staal A (1989) Anticipation in myotonic dystrophy: fact or fiction? *Brain*, **112**, 779–797.

Kendler KS, Neale MC, Kessler RC, Heath AC and Eaves LJ (1993) Panic disorder in women: a population based twin study. *Psychological Medicine*, **23**, 397–406.

Klerman GL, Weissman MM, Ouellette R, Johnson J and Greenwald S (1991) Panic attacks in the community. Social morbidity and health care utilization. *Journal of the American Medical Association*, **265**, 742–746.

Knowles JA, Vieland VJ, Fyer A, Heiman GA, Rassnick H, Fine L, Austin TL, Adams P, Hodge SE, Klein DF, Ott J, Weissman MM and Gilliam TC (1993) Genetic linkage analysis of panic disorder. Poster: American Society of Human Genetics, New Orleans, October 1993. *American Journal of Human Genetics*, **53**, 1023A.

Loyd JE, Butler MG, Foroud TM, Conneally PM, Phillips JA and Newman JH (1995) Genetic anticipation and abnormal gender ratio at birth in familial primary pulmonary hypertension. *American Journal of Respiratory Critical Care Medicine*, **152**, 93–97.

Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, Sudarsky L, Guimarães J, Loureiro JEL, Nezarati MM, Corwin LI, Lopes-Cendes I, Rooke K, Rosenberg R, MacLeod P, Farrer LA, Sequeiros J and Rouleau GA (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph Disease. *American Journal of Human Genetics*, **57**, 54–61.

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA and DePaulo JR (1993) Anticipation in bipolar affective disorder. *American Journal of Human Genetics*, **53**, 385–390.

Mott FW (1910) Hereditary aspects of nervous and mental diseases. *British Medical Journal*, **2**, 1013–1020.

Nylander P, Engström C, Chotai J, Wahlström J and Adolfsson R (1994) Anticipation in Swedish families with bipolar affective disorder. *Journal of Medical Genetics*, **31**, 686–698.

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. *Annals of Eugenics*, **14**, 125–132.

Petronis LS and Kennedy JL (1995) Unstable genes unstable mind? *American Journal of Psychiatry*, **152**, 164–172.

Ross CA, McInnis MG, Margolis RL and Li S (1993) Genes with triplet repeats: candidate mediators of neuropsychiatric disorders. *Trends in Neuroscience*, **16**, 254–260.

Sano A, Yamauchi N, Nakimoto Y, Komure O, Kawai J, Hazama F, Kuzume K, Sano N and Kondo I (1994) Anticipation in hereditary dentatorubral-pallidoluysian atrophy. *Human Genetics*, **93**, 699–702.

Simon GE and VonKorff M (1992) Reevaluation of secular trends in depression rates. *American Journal of Epidemiology*, **135**, 1411–1422.

Torgersen S (1983) Genetic factors in anxiety disorders. *Archives of General Psychiatry*, **40**, 1085–1089.

Vieland VJ, Hodge SE, Lish JD, Adams P and Weissman MM (1993) Segregation analysis of panic disorder. *Psychiatric Genetics*, **3**, 63–71.

Weissman MM, Wickramaratne P, Adams PB, Lish JD, Horwath E, Charney D, Woods SW, Leeman E and Frosch E (1993) The relationship between panic disorder and major depression: a new family study. *Archives of General Psychiatry*, **50**, 767–780.

Willems PJ (1994) Dynamic mutations hit double figures. *Nature Genetics*, **8**, 213–215.

Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, Padberg G and Frants R (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *American Journal of Human Genetics*, **56**, 99–105.