Psychological Medicine, 2000, **30**, 1315–1325. Printed in the United Kingdom

Genetic and environmental influences on juvenile antisocial behaviour assessed on two occasions

K. C. JACOBSON,[1] C. A. PRESCOTT AND K. S. KENDLER

From the Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Medical College of Virginia of Virginia Commonwealth University, Richmond, VA, USA

ABSTRACT

Background. There is conflicting evidence concerning the magnitude of genetic and shared environmental influences on juvenile antisocial behaviour (AB). The use of more than one assessment of AB may yield more accurate estimates of these influences.

Methods. Retrospective reports of antisocial behaviour prior to age 18 were obtained on two occasions from a population-based sample of 3522 adult males from male–male twin pairs: phone interviews (wave 1) and self-report questionnaires obtained 19 months later (wave 2). Structural equation modelling estimated the genetic and environmental influences on reliably-measured AB. Factors related to participation of co-twin at wave 1, attrition between waves 1 and 2, and reliability of wave 1 and wave 2 assessments were also investigated.

Results. Twin analyses revealed that genetic, shared environmental, and non-shared environmental influences accounted for approximately 33% (95% CI = 9–57%), 31% (95% CI = 10–51%) and 36% (95% CI = 29–44%) of the variance of reliably measured AB, respectively. We also found significant occasion-specific genetic influences on wave 1 AB. Wave 1 AB did not predict wave 1 participation of co-twin or attrition, but was related to reliability. Co-twins of MZ twins and younger twins were more likely to participate at wave 1; attrition was predicted by being a DZ twin, lack of initial participation of co-twin, fewer years of education, and fewer children. Being older, being unmarried, and having less psychopathology were associated with greater reliability.

Conclusions. When measurement error is taken into account, both genetic and shared environmental factors are significant influences on juvenile AB, accounting for approximately one-third of variation. The origin of the specific genetic influences on wave 1 AB is unclear, but may be due to factors related to measurement.

INTRODUCTION

There is considerable debate over the role that genetic factors play in the aetiology of juvenile antisocial behaviour (AB). A number of studies have suggested that genetic influences are the primary source of familial resemblance (Rowe, 1983, 1986; Graham & Stevenson, 1985; Ghodsian-Carpey & Baker, 1987; Silberg *et al.* 1996; Gjone & Stevenson, 1997; Slutske *et al.* 1997), whereas other studies suggest that the primary source of familial resemblance is common rearing environment (Lyons *et al.* 1995; Thapar & McGuffin, 1996). Still other studies of adolescents find that both genetic and shared environmental factors contribute to variation in juvenile AB (Edelbrock *et al.* 1995; Simonoff *et al.* 1995; Eley *et al.* 1999). Differences in sample composition and size may be partly responsible for these inconsistencies. For example, there is likely to be a selection against antisocial behaviour in studies that rely on volunteer or selected samples (Lyons *et al.* 1995; Slutske *et al.* 1997).

[1] Address for correspondence: Dr Kristen C. Jacobson, Virginia Institute for Psychiatric and Behavioral Genetics, MCV/VCU, 800 E. Leigh Street, PO Box 980126, Richmond, VA 23298–0126, USA.

A second possibility is that differences in the measurement of juvenile AB may account for the discrepant results. In social science research, the variance of any given observed measure is composed of the true score variance and measurement error. Thus, error of measurement increases the discrepancy between the true score variance and the observed, phenotypic variance. The closer the phenotypic variance is to the true score variance, the more accurate the estimates of genetic and environmental influences on the true score variance will be. Studies of juvenile AB often differ on the instruments and raters used to assess AB, and certain instruments and certain raters may yield more reliable measures of AB than others. One way to obtain more accurate estimates of genetic and environmental influences is to combine results across many studies in a meta-analyses. In an analysis based on 24 twin and adoptions studies of aggressive behaviour, Miles & Carey (1997) concluded that genes and common rearing environment each contribute approximately equally to variation in juvenile aggression.

However, meta-analyses cannot differentiate between reliable and unreliable variation; thus, estimates of genetic and environmental influences may still be attenuated if there is substantial measurement error. Measurement models that use more than one occasion of assessment can differentiate between variation that is specific to a given occasion (such as variation due to measurement error) and the reliable variation of the underlying latent phenotype. When the measurement model is used in combination with genetically-informative data, estimates of the heritability of the latent phenotype can be contrasted with heritability estimates from a single measure. Studies of major depression, fears, and phobias have found that the heritability of the underlying latent phenotype was greater than the heritability obtained from a single measure (Foley *et al.* 1998; Kendler *et al.* 1993, 1999).

There have been at least two studies employing a measurement-model approach to examine the effects of rater bias on the heritability of juvenile AB (Hewitt *et al.* 1992; Simonoff *et al.* 1995). Both studies revealed that the heritability of the latent phenotype was greater than the heritability of AB as assessed by any one given rater. However, measurement models that use different raters raise more complex issues, as there may be systematic reasons for differences in heritability estimates across raters, such as genetically-influenced characteristics that are related to rater perception and bias.

The primary purpose of the present study is to use a measurement model to estimate the heritability of reliable juvenile AB in a sample of adult twins, using retrospective self-reports of juvenile AB from two different waves of data. To address potential biases in our sample, we also investigate factors associated with study participation, attrition and reliability.

METHOD

Sample and procedure

Twins were part of the Medical College of Virginia Stress and Coping Project, a longitudinal study of the genetic and environmental risk factors for common psychiatric and substance use disorders among adult male and female twins. Twins were ascertained from the population-based Virginia Twin Registry, which was formed by a systematic search of all Virginia birth certificates since 1918. Subjects from multiple births were matched by name and birth date to state records to obtain addresses and phone numbers. Twins were eligible for participation in the study if they met the following four conditions: one or both twins were successfully matched to state records; they were a member of a multiple birth which included at least one male; they were Caucasian; and they were born between 1940 and 1974 ($N = 9417$ eligible twins). Data for the present study come from the wave 1 interview and the wave 2 self-report questionnaire (SRQ).

Of the 9417 twins eligible for participation, 72·7% ($N = 6847$) participated at the wave 1 interview. Wave 2 interviews were obtained from 82·5% ($N = 5651$) of twins who participated at wave 1, and 94·2% of the wave 2 respondents also completed an SRQ. Wave 2 interviews took place an average of 19·3 (S.D. = 8·8) months after the wave 1 interview. Of wave 1 interviews 98·1% were conducted over the phone; 79·8% of wave 2 interviews were conducted face-to-face (see Kendler & Prescott, 1999, for details). The SRQ was completed during the wave 2 interview for 80·2% of the sample, with 1·8% of respondents completing it

an average of 6·9 days (S.D. = 11·8) prior to the wave 2 interview and 18·0% returning the SRQ a median of 1·5 (S.D. = 7·0) months after the wave 2 interview.

Although the SRQ was designed to be a written questionnaire, some of the twins interviewed via telephone answered the SRQ orally. Due to inadequate reading comprehension levels, in a small number of face-to-face interviews the interviewer read the SRQ aloud and the respondent answered orally. Thus, the response mode for the SRQ included both oral and written self-reports. The SRQ response mode was missing from twins from a random sample of twins studied earlier in the project. We were able to identify the response mode for the SRQ for 3955 twins (74·3%). Of these 3955 twins, 74·2% ($N = 2934$) provided written responses to the SRQ.

Two waves of data on AB from same-sex female twins were not available; thus, the present analyses focus solely on same-sex male twin pairs. At wave 1 3540 individual twins from male–male twin pairs were interviewed. Our final sample was 3522 twins with valid data for AB. This sample included 2996 twins from complete twin pairs (852 MZ, 646 DZ pairs), 15 individual male–male–male triplets from five triplet families (creating 10 MZ and 5 DZ pairs) and 511 unpaired male twins whose co-twins did not participate in the study (235 MZ, 276 DZ). Zygosity of complete twin pairs was determined by a combination of twins' responses to standard questions regarding twin similarity, photographs, and DNA typing. Assignment of zygosity for twins whose co-twins did not cooperate was done using a discriminant function analysis of items regarding physical similarity and twin self-report of zygosity, with DNA-typed twins ($N = 540$) as the comparison group.

The average age of twins at the wave 1 interview was 35·1 (S.D. = 9·3; range = 19–56 years). Twins had a mean of 13·3 years of education (S.D. = 2·7). This project was approved by our local Institutional Review Board. Subjects were informed about the goals of the study and provided verbal consent prior to telephone interviews and written informed consent prior to face to face interviews and collection of DNA samples.

Measures

Antisocial behaviour (*AB*)

In the wave 1 interview, twins were asked whether they had engaged in 11 specific antisocial behaviours prior to age 18.[1]† The wording of each item corresponded closely to the criteria used to define 11 of the 13 symptoms of conduct disorder in the DSM-III-R (APA, 1987; e.g. 'played hooky a lot from school', 'physically hurt other people a number of times', 'used a weapon in a fight more than once').[2] Responses for each item were 'yes' or 'no'. A summary score of symptoms was computed by summing the 'yes' items. In the wave 2 SRQ, the 11 questions were repeated for the period prior to age 15, and 9 of the 11 items were asked for the period age 15–17.[3] Twins were asked to indicate the frequency with which they had engaged in these behaviours, ranging from 0 = 'never' to 3 = '6 or more times' or 'often'. The frequencies for each item were summed across the two time periods and a computer algorithm was applied to the summed score to create a cut-off criterion for whether that particular symptom was present (1) or absent (0). The algorithm was designed to match the wording of the wave 1 AB items. A summary score of number of symptoms present was then created. Because each of the two scales was positively skewed (skewness = 1·91, for wave 1 AB; 1·40, for wave 2 AB), a square-root transformation was employed for use in the twin analyses.

Demographic variables

Demographic characteristics used in the participation and reliability analyses included age, number of years of education (as a proxy for social class) and stability characteristics such as number of children and marital status at wave 1 (coded as currently married *versus* all other marital statuses).

Other psychiatric disorders

Two other psychiatric disorders were used to predict reliability: wave 1 measurement of lifetime diagnosis of major depression (based on DSM-III-R criteria; see Kendler & Prescott, 1999, for details), and diagnosis of alcohol

† The notes will be found on p. 1324

https://doi.org/10.1017/S0033291799002780 Published online by Cambridge University Press

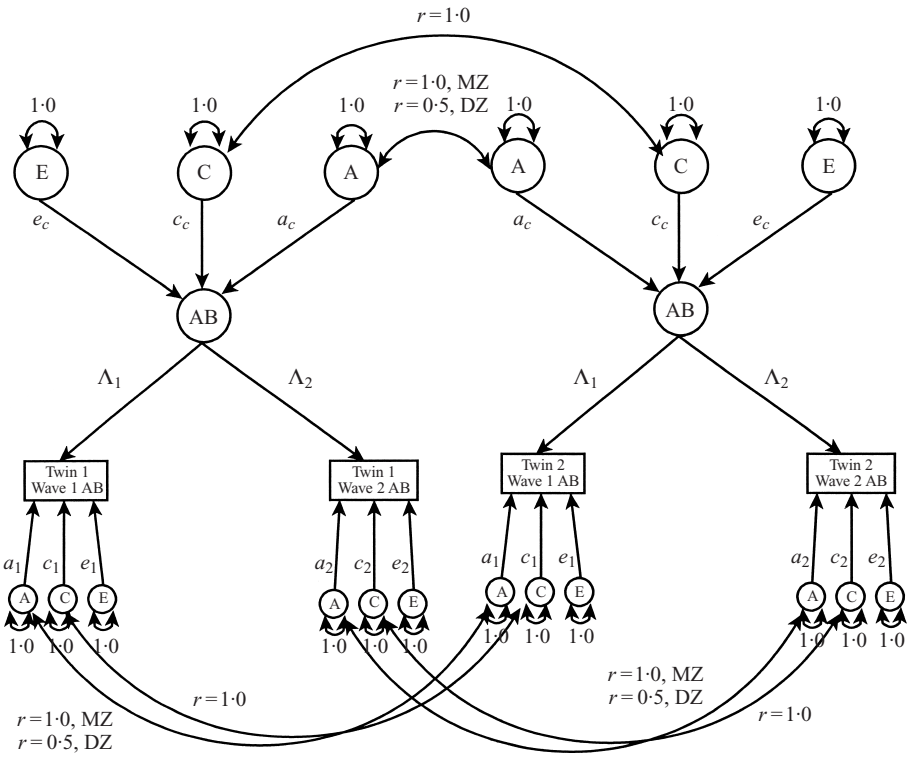

FIG. 1. Bivariate path model for twin analyses (A, Additive genetic variance; C, common (shared) environment; E, non-shared environment; AB, latent variable representing the phenotype antisocial behaviour). Genetic influences are correlated $r = 1{\cdot}0$, for MZ twins and $r = 0{\cdot}5$, for DZ twins. Shared environmental influences are correlated $r = 1{\cdot}0$ for both MZ and DZ twins. Non-shared environmental influences are not correlated across twins. The variances of all latent variables are assumed to be 1·0. Constraints were imposed so that $a_c^2 + c_c^2 + e_c^2 = 1{\cdot}0$.

dependence or alcohol abuse during the period of heaviest drinking (based on DSM-IV criteria; see Prescott & Kendler, 1999, for details).

Participation analyses

We examined whether zygosity, wave 1 AB, age, and education level predicted co-twin participation at wave 1 using standard chi-square analyses and univariate logistic regressions with the PROC GENMOD procedure in SAS. PROC GENMOD takes into account the statistical dependency arising from having more than one twin per family in the analysis by using generalized estimating equation (GEE) methodology (SAS Institute Inc., 1997). We also examined the impact of zygosity, co-twin participation at wave 1, and age, number of children, marital status, and AB at wave 1 on individual attrition using multivariate logistic regression in PROC GENMOD. Six twins were excluded from the participation analyses because of missing data on one or more of the demographic variables. Co-twin participation at wave 1 was coded as '1' (yes) or '0' (no), and attrition was coded as '1' for twins who did not participate at wave 2, and '0' for twins who participated at both waves.

Reliability

A total of 2758 individual twins had valid demographic data, psychiatric diagnosis data and AB data at both wave 1 and wave 2. Reliability of reporting was indexed by taking the absolute difference between the wave 1 and 2 AB symptom scores. This variable was then used as the dependent variable in a multivariate, stepwise regression analysis, using zygosity, wave 1 co-twin cooperation, age, education

https://doi.org/10.1017/S0033291799002780 Published online by Cambridge University Press

level, number of children, marital status, number of days between wave 1 and wave 2 assessments, wave 1 AB, and the two psychiatric diagnoses as independent variables.

Twin analyses

Fig. 1 presents the bivariate measurement model that was used in the genetic analyses. In this model, the latent variable, AB, symbolizes the underlying phenotype of AB, and is indexed by the measured square-root transformed AB scores at waves 1 and 2. The path coefficients, λ_1 and λ_2, are square-root estimates of the amount of variance in wave 1 and 2 AB that is common to both measures. The full model allows for common additive genetic (A), shared environmental (C), and non-shared environmental influences (E) on the latent phenotype (represented by the parameters a_c, c_c, and e_c), as well as for specific genetic and environmental influences on the individual measures of AB at each wave (parameters a_1–e_2). These latter influences are occasion-specific influences because they are not correlated across waves and they do not share variance with the underlying latent phenotype. To identify the model, the variance of the genetic and environmental influences was assumed to be 1·0, and a constraint was imposed such that the variance of the latent phenotype was also unity (i.e. $a_c^2 + c_c^2 + e_c^2 = 1{\cdot}0$). This model also estimates mean AB scores (not shown in Fig. 1).

The full model and subsequent submodels were fit to raw data using the structural equation modelling program, *Mx* (Neale, 1997). Twin correlations based on raw data are shown in Appendix A. The relative fit of the submodels was evaluated using the likelihood ratio chi-square statistic (LRC), which is calculated by twice the difference in log-likelihood between two nested models.

All 3522 individual twins were included in the genetic analyses. Of the 1513 twin pairs in which both twins participated at wave 1, 1075 pairs also had wave 2 data from both twins (641 MZ, 434 DZ pairs), 323 pairs had wave 2 data from one twin only (161 MZ, 162 DZ pairs), and 115 pairs did not have any wave 2 data (60 MZ, 55 DZ pairs). Of the 511 twins whose co-twins did not participate at wave 1, 306 participated at wave 2 (142 MZ, 164 DZ), and 205 did not have wave 2 data (93 MZ, 112 DZ). The use of all twins reduces biases in estimates of mean level AB and variance due to nonparticipation. In addition, twins pairs with AB data for at least one wave for each twin also contribute to estimates of covariance.

RESULTS

Study participation

Participation of co-twin at wave 1 was predicted by monozygosity ($\chi^2(1) = 22{\cdot}91$, $P < 0{\cdot}001$), higher education (OR = 1·12, 95% CI = 1·08, 1·16), and younger age (OR = 0·98, 95% CI = 0·97, 0·99). Wave 1 AB did not predict wave 1 co-twin participation. Multivariate logistic regression found significant effects of zygosity, wave 1 cooperation of co-twin, years of education, and number of children on attrition from wave 1 to wave 2. MZ twins and twins whose co-twin participated at wave 1 were more likely to participate at wave 2 (OR = 1·24, 95% CI = 1·04, 1·49; OR = 2·91, 95% CI = 2·36, 3·60, respectively), and twins with fewer children and fewer years of education were less likely to participate at wave 2 (OR = 0·93, 95% CI = 0·86, 0·99; OR = 0·90, 95% CI = 0·87, 0·94, respectively). No effects on attrition were found for wave 1 AB, age, or marital status.

Reliability

Short-term reliability of wave 2 AB was estimated with a randomly-selected subsample of 131 male twins who completed a second SRQ a median of 29 days after the initial wave 2 interview. The intraclass test–retest *r* for square-root transformed AB among this sample was 0·72. Long-term reliability, as indexed by the intraclass correlation between waves 1 and 2 AB (square-root transformed), was 0·51 ($N = 2767$).

According to results from the stepwise regression, older twins and unmarried twins were more reliable ($b = 0{\cdot}010$, S.E. = 0·002, $t = 4{\cdot}08$, $P < 0{\cdot}001$, for age; $b = 0{\cdot}09$, S.E. = 0·04, $t = 2{\cdot}08$, $P < 0{\cdot}05$, for marital status) and higher number of AB symptoms at wave 1 ($b = 0{\cdot}19$, S.E. = 0·01, $t = 12{\cdot}93$, $P < 0{\cdot}001$) and history of alcohol dependence or abuse ($b = 0{\cdot}17$, S.E. = 0·04, $t = 4{\cdot}21$, $P < 0{\cdot}001$) were associated with lower reliability. Zygosity, co-twin participation

https://doi.org/10.1017/S0033291799002780 Published online by Cambridge University Press

Table 1. *Results from model-fitting of the twin data*

Model	Absolute model fit			Relative model fit*		
	−2 LL	df	P†	LRC	df	P†
1 Full model	12893·55	6305	0·34	—	—	—
2 No a_1, c_1, a_2, c_2	12930·42	6309	0·001	36·87	4	0·001
3 **No c_1, a_2, c_2**	**12894·45**	**6308**	**0·48**	**0·90**	**3**	**0·83**
4 Model 3 + No a_c	12901·95	6309	0·16	7·50	1	0·006
5 Model 3 + No c_c	12902·57	6309	0·15	8·12	1	0·004
6 Model 3 + No e_c	13003·51	6309	0·001	109·06	1	0·001

−2 LL, Log-likelihood function. Parameters with the subscript: 1, refer to occasion-specific influences on wave 1 antisocial behaviour; 2, refer to occasion-specific influences on wave 2 antisocial behaviour; and c, represent genetic and environmental influences on the underlying latent phenotype.

* The full model is used as the comparison model for Models 2 and 3; Model 3 is used as the comparison model for Models 4–6.

† The *P* value for the absolute model fit is based on a comparison of the −2 LL of each model with the −2 LL of a fully-saturated model (not shown) and reflects the degree to which the model fits the observed data. The *P* value for the relative model fit is the *P* value associated with the LRC. The best-fitting model is in boldtype.

at wave 1, number of children, years of education, length of time between wave 1 and wave 2 assessments, and lifetime history of major depression were not associated with reliability. The full model explained 9·1% of the variance in reliability.

Genetic analyses

Prior to testing the significance of the genetic and environmental influences on AB, we examined: (1) whether mean AB scores could be equated across twin 1 and twin 2 and across MZ and DZ twins; and (2) whether estimates of the amount of variance in wave 1 and 2 AB common to both measures could be made equal for the wave 1 and wave 2 assessments (i.e. $\lambda_1 = \lambda_2$). The LRC for question 1 was 9·91 (df = 6, P = 0·128), indicating that mean level of AB did not differ across twins or across zygosity. Setting the λ parameters to be equal also did not result in a significant deterioration in fit (LRC = 0·15, df = 1, P = 0·698); thus, these constraints were imposed in the following analyses.

Table 1 presents the results from the model-fitting analyses. Model 1 is the full model, portrayed in Fig. 1. Model 2 set the occasion-specific genetic and shared environmental influences on wave 1 and wave 2 AB to 0. This model fit the data significantly more poorly than the full model (LRC = 36·87, df = 4, P = 0·001). Examination of the 95% confidence intervals (CI) from the full model indicated that the specific genetic influence on wave 1 AB was significantly greater than 0. Thus, Model 3 allowed for measurement error at both waves and for specific genetic influences on wave 1 AB. This model fit the data about as well as the full model (LRC = 0·90, df = 3, P = 0·83). This result indicates that although the occasion-specific genetic influences on wave 1 AB were significant, occasion-specific genetic influences on wave 2 AB and occasion-specific shared environmental influences on both wave 1 and wave 2 AB were not statistically significant. Models 4, 5, and 6 then tested the significance of the genetic, shared environmental, and non-shared environmental influences on the latent variable, respectively, by setting the corresponding parameter to zero and examining the deterioration in model fit compared to Model 3. Using the LRC test, all three models fit the data more poorly than Model 3. Thus, Model 3 was considered the best-fitting model.

The standardized parameter estimates from Model 3 are shown in Fig. 2. Although the unstandardized parameters for λ_1 and λ_2 were constrained to be equal, variance differences in the 2 waves resulted in slightly different standardized estimates. Estimates of the amount of variance in wave 1 and 2 AB common to both measures were 0·55 and 0·52, respectively. The expected correlation between wave 1 and wave 2 AB was 0·53, similar to the observed intraclass correlation. Genetic and shared environmental influences on the latent phenotype accounted for 31·4% (95% CI = 10·1–51·2%) and 30·3% (95% CI = 9·2–56·9%) of the variance, respectively, with non-shared environmental influences accounting for the remaining 35·9% (95% CI = 28·9–43·9%). Total genetic, shared environ-

https://doi.org/10.1017/S0033291799002780 Published online by Cambridge University Press

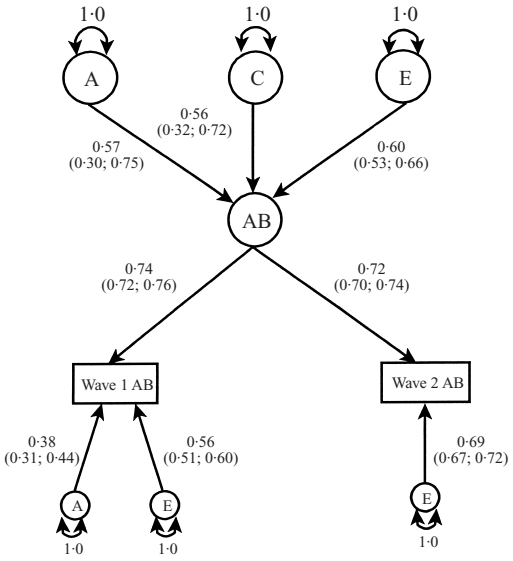

FIG. 2. Standardized parameter estimates from the best-fitting model (Model 3). Only one twin is shown in the diagram (95% confidence intervals are presented in parentheses). (A, Additive genetic variance; C, common (shared) environment; E, non-shared environment; AB, latent variable representing the phenotype antisocial behaviour). The variances of all latent variables are assumed to be 1·0. Constraints were imposed so that the variance of the underlying latent variable, AB, equalled unity. Proportions of variance are obtained by squaring the estimates shown.

mental and non-shared environmental influences on wave 1 AB, including both occasion-specific and common factors, accounted for 32·3% (95% CI = 19·3–45·6%), 17·2% (95% CI = 5·5–28·1%) and 50·5% (95% CI = 46·1–55·3%) of the variance, respectively and the corresponding percentages for wave 2 AB were 16·9% (95% CI = 4·8–29·6%), 16·3% (95% CI = 5·2–26·7%) and 66·8% (95% CI = 62·4–71·3%). Of the 0·32 heritability estimate for wave 1 AB, approximately 45% (0·14) came from the occasion-specific genetic influences that were not related to the underlying latent phenotype.

DISCUSSION

The present study suggests that retrospective measures of antisocial behaviour (AB) are modestly reliable. When the same instrument was administered to a group of 131 twins approximately 1 month after the initial administration, the estimate of reliability was 0·72. The correlation between AB assessed at wave 1 and AB assessed at wave 2 (an average of 19 months apart) was approximately 0·50.

This modest reliability suggests that the use of more than one measure of AB should yield more accurate estimates of genetic and environmental influences than studies that rely on a single measure. In the present study, genetic and shared environmental influences each accounted for approximately one-third of the reliable variation in juvenile AB, and both of these influences were statistically significant. This finding is consistent with the conclusions based on Miles & Carey's (1997) meta-analysis of aggressive behaviour. Estimates of shared environmental influences on the latent phenotype were approximately twice as great as the corresponding estimates for either measured variable, and the heritability of the latent phenotype was greater than the heritability of wave 2 AB (but not wave 1; see discussion below).

The measurement-model approach to studying variation is also useful because it can differentiate between nonshared environmental influences that are due to measurement error, and reliable non-shared environmental influences. In this study, in addition to occasion-specific non-shared environmental influences, which are likely measurement error, non-shared environmental influences on the latent phenotype were also statistically significant, and accounted for approximately one-third of the variation in juvenile AB. This reinforces the notion that environmental factors that are not shared by twins during adolescence, such as peer groups or differential parental treatment, are important in the aetiology of juvenile AB (Patterson & Stouthamer-Loeber, 1984; Patterson & Dishion, 1985; Snyder *et al.* 1986; Pike *et al.* 1996).

Occasion-specific genetic influences on antisocial behaviour

A surprising finding from the present study was significant occasion-specific genetic influences on wave 1 reports of AB. Examination of the twin correlations reveals that this unique genetic influence was specific to wave 1 because the difference between the MZ and DZ twin correlations for wave 1 AB (0·48, MZ; 0·32 DZ) was greater than for wave 2 AB (0·35, MZ; 0·27, DZ) and because the MZ twin correlation for wave 1 AB (0·48) was greater than the MZ cross-twin, cross-wave correlations (0·29 and 0·37; see Appendix A).

https://doi.org/10.1017/S0033291799002780 Published online by Cambridge University Press

There are at least three possible explanations for this result. The first possibility is that this is a stochastic effect, simply a chance finding that has been interpreted as a 'real' result. The LRC associated with a model that set only the wave 1 specific genetic parameter to zero was 5·13 (df = 1), with a P value of 0·023 (results not shown). Thus, in approximately one out of 50 studies, such a unique genetic effect would be found solely by chance.

A second possibility has to do with selection effects. If AB was related to study participation, there could be a higher proportion of antisocial individuals at the wave 1 assessment, relative to the wave 2 assessment. The selection against antisocial individuals at wave 2 might attenuate genetic influences. In the present study, however, wave 1 AB scores were not related to attrition or to initial participation of co-twin, suggesting that estimates of genetic influence were not biased by selection effects.

A third possibility is that the significant occasion-specific genetic influence on recall of AB at wave 1 reflects genetic influences on characteristics that are related to inherent differences in the measurement of wave 1 and wave 2 AB. For example, retrospective reports of AB at wave 1 were assessed via items that used a yes/no response format to items based on a specific frequency (e.g. 'used a weapon in a fight more than one time') whereas in wave 2, the frequency of the AB items was assessed directly via a Likert scale. There is evidence that factors such as question wording and response format are related to item endorsement (Schwarz, 1999) and these factors may also influence estimates of genetic and environmental influences. In a recent twin study of personality using three separate measures (trait ratings, a standard personality inventory and adjective scales), significant measure-specific genetic components were found, indicating that there were genetic influences on each measure that were not related to the underlying latent personality trait (Loehlin *et al.* 1998). In the present study, because the type of measure is confounded with wave of measurement, it is impossible to test the hypothesis that genetic factors influence yes/no responding to a greater extent than they do frequency-based responding.

Another difference between the two waves is in primary response mode. In wave 1, questions concerning AB were answered verbally over the telephone. In wave 2, the majority of respondents gave written reports of AB while the interviewer was present. It is possible that personality characteristics, such as social desirability, influence willingness to disclose sensitive information in different response settings (Siemiatycki, 1979). There is considerable evidence that major aspects of personality, including neuroticism, conscientiousness and agreeableness, are heritable (see Eaves *et al.* 1989; Loehlin, 1992, for reviews), and there is some evidence that responses on the Lie Scale from the Eysenck Personality Questionnaire, a measure of social desirability response, may also be genetically influenced (Eaves *et al.* 1989; Macaskill *et al.* 1994). Thus, the unique genetic influence found for wave 1 AB may reflect genetic influences on willingness to disclose sensitive information over the telephone.

Strengths and limitations

Although our sample was drawn from a population-based twin registry, there was some evidence that the sample was not completely representative. For example, the present sample was restricted to Caucasian male twins from male–male twin pairs who were born in Virginia. Thus, it is possible that these results may not generalize to females, or to non-Caucasian individuals. Like other twin studies (Lykken *et al.* 1978, 1987), we also found evidence for systematic differences among twins who participated in the study *versus* twins who did not: MZ twins, younger twins, and more educated twins were more likely to have co-twins who participated at wave 1, and MZ twins and twins whose co-twins participated in the first wave were less likely to drop out of the study. Also consistent with prior research was the finding that certain stability characteristics, such as lower education level and fewer children, predicted attrition (Spoth *et al.* 1999). However, wave 1 AB score did not predict either co-twin participation at wave 1 or attrition, suggesting that our sample was not biased with respect to the phenotype in question.

The use of retrospective self-reports of juvenile AB has both strengths and limitations. On the one hand, there is evidence that individuals are more willing to report negative behaviours in self-report questionnaires than in structured

face-to-face interviews (Siemiatycki, 1979). Moreover, the use of retrospective data ensures that all individuals have passed through the age of risk, so developmental differences in rates of CD cannot bias results. On the other hand, retrospective reports of juvenile AB may be less reliable than contemporaneous measures, and reliability may be further influenced by the age of the respondent. In our study, short-term reliability ($r = 0{\cdot}72$), based on an average 1-month interval, was somewhat greater than long-term reliability ($r = 0{\cdot}51$), based on an average 19-month interval. However, measured length of time between wave 1 and wave 2 assessments was not related to reliability in a stepwise multiple regression. In addition, older twins were more reliable than younger twins, suggesting that our retrospective measures were not biased by length of time in recall.

Finally, reliability was predicted by both number of AB symptoms at wave 1 and diagnosis of alcohol abuse or dependence, implying that individuals with greater psychopathology are less reliable reporters of AB. However, these results might also be due to distributional properties of AB. In particular, 46·8% of the sample reported no AB symptoms at wave 1. In longitudinal studies, initial level of behaviour and magnitude of change are often correlated. Thus, the finding that higher levels of wave 1 AB were associated with a larger absolute difference between waves 1 and 2 could be due to this phenomenon. To determine whether the distribution of AB could have biased our estimates of genetic and environmental influences, we also ran the genetic models using actual number of symptoms as an ordinal variable and estimated thresholds. Results from those analyses did not differ from those presented here. The finding that history of alcohol abuse or dependence was also associated with less reliability could be due to possible memory impairment in heavy drinkers, or could be a function of the fact that twins with history of alcohol abuse or dependence also had higher wave 1 AB scores (mean = 0·72, S.D. = 1·04, for twins without alcohol abuse or dependence; mean = 1·62, S.D. = 1·71, for twins with history of alcohol abuse or dependence, $F(1, 3516) = 377{\cdot}35$, $P < 0{\cdot}001$).

In conclusion, this study found that genetic, shared environmental, and non-shared environmental influences contributed about equally to variation in a latent phenotype represented by two measures of AB. The origin of the significant genetic influences on wave 1 AB could not be determined; it is possible that differences in measurement at the two waves is responsible. Nevertheless, the study accords with findings from the Miles & Carey (1997) meta-analysis, and indicates that both genetic and environmental factors are important in the aetiology of juvenile AB.

This work was supported by a National Institutes of Health grant MH/AA-49492 and a Research Scientist Award (MH-01277) (Dr Kendler). The Virginia Twin Registry, established by Dr W. Nance and maintained by Dr L. Corey, and Dr Lenn Murrelle is supported by the US National Institutes of Health grants HD-26746 and NS-31564.

We thank Sarah Woltz, Frank Butera and Patsy Waring for their assistance with data collection and maintenance.

APPENDIX A: TWIN CORRELATIONS FOR ANTISOCIAL BEHAVIOUR

	MZ twins		DZ twins	
	N	*r*	*N*	*r*
Within-twin correlations				
Twin 1, Wave 1 AB – Twin 1, Wave 2 AB	801	0·57***	605	0·54***
Twin 2, Wave 1 AB – Twin 2, Wave 2 AB	784	0·50***	589	0·50***
Cross-twin correlations				
Twin 1, Wave 1 AB – Twin 2, Wave 1 AB	862	0·48***	650	0·32***
Twin 1, Wave 2 AB – Twin 2, Wave 2 AB	641	0·35***	435	0·27***
Cross-twin, cross-wave correlations				
Twin 1, Wave 1 AB – Twin 2, Wave 2 AB	723	0·29***	517	0·24***
Twin 1, Wave 2 AB – Twin 2, Wave 1 AB	720	0·37***	513	0·25***

AB, Antisocial behaviour.
*** $P < 0{\cdot}001$.

NOTES

[1] In the initial stages of wave 1, respondents were asked whether they had engaged in these behaviours before age 15. The low endorsement of the items caused us to change the wording of the items to include behaviours prior to age 18. Approximately 5·7% ($N = 202$) of the male twins in the present sample completed the earlier version of the wave 1 interview. All analyses reported in the paper were redone without these twins, and the pattern of results was identical to the results reported here.

[2] The items in the SRQ differ from the DSM-III-R in the following ways. First, the item regarding forced sex was eliminated, as in other epidemiological samples its endorsement is very infrequent and could be offensive to some respondents. Secondly, two of the DSM-III-R items ('has broken into someone else's house, building, or car' and 'has stolen without confrontation') were combined into a single item: 'stole things from a store or from someone I knew'.

[3] Two of the items (frequencies of lying and starting physical fights) were only asked for the period prior to age 15.

REFERENCES

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders* (*3rd edn Revised*). APA: Washington, DC.

Eaves, L. J., Eysenck, H. J. & Martin, N. G. (1989). *Genes, Culture, and Personality: An Empirical Approach*. Academic Press: San Diego, CA.

Edelbrock, C., Rende, R., Plomin, R. & Thompson, L. A. (1995). A twin study of competence and problem behavior in childhood and early adolescence. *Journal of Child Psychology and Psychiatry* **56**, 775–785.

Eley, T. C., Lichtenstein, P. & Stevenson, J. (1999). Sex differences in the etiology of aggressive and non-aggressive antisocial behavior: results from two twin studies. *Child Development* **70**, 155–168.

Foley, D. L., Neale, M. C. & Kendler, K. S. (1998). Reliability of a lifetime history of major depression: implications for heritability and comorbidity. *Psychological Medicine* **28**, 857–870.

Ghodsian-Carpey, J. & Baker, L. A. (1987). Genetic and environmental influences on aggression in 4- to 7-year old twins. *Aggressive Behavior* **13**, 173–186.

Gjone, H. & Stevenson, J. (1997). A longitudinal twin study of temperament and behavior problems: common genetic or environmental influences? *Journal of the American Academy of Child and Adolescent Psychiatry* **36**, 1448–1456.

Graham, P. & Stevenson, J. (1985). A twin study of genetic influences on behavioral deviance. *Journal of the American Academy of Child Psychiatry* **24**, 33–41.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves, L. J. & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics* **22**, 293–317.

Kendler, K. S. & Prescott, C. A. (1999). A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry* **56**, 39–44.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. (1993). The lifetime history of major depression in women. *Archives of General Psychiatry* **50**, 863–870.

Kendler, K. S., Karkowski, L. M. & Prescott, C. (1999). Fears and phobias: reliability and heritability. *Psychological Medicine* **29**, 539–553.

Loehlin, J. C. (1992). *Genes and Environment in Personality Development*. Sage: Newbury Park, CA.

Loehlin, J. C., McCrae, R. R., Costa, P. T. & John, O. P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. *Journal of Research in Personality* **32**, 431–453.

Lykken, D. T., Tellegen, A. & DeRubeis, R. (1978). Volunteer bias in twin research: the rule of two-thirds. *Social Biology* **25**, 1–9.

Lykken, D. T., McGue, M. & Tellegen, A. (1987). Recruitment bias in twin research: the rule of two-thirds reconsidered. *Behavior Genetics* **17**, 343–362.

Lyons, M. L., True, W. R., Eisen, S. A., Goldberg, J., Meyer, J. M., Faraone, S. V., Eaves, L. J. & Tsuang, M. T. (1995). Differential heritability of adult and juvenile traits. *Archives of General Psychiatry* **52**, 906–915.

Macaskill, G. T., Hopper, J. L., White, V. & Hill, D. J. (1994). Genetic and environmental variation in Eysenck Personality Questionnaire scales measured on Australian adolescent twins. *Behavior Genetics* **24**, 481–491.

Miles, D. R. & Carey, G. (1997). Genetic and environmental architecture of human aggression. *Journal of Personality and Social Psychology* **72**, 207–217.

Neale, M. C. (1997). *Mx: Statistical Modeling, 4th edn*. Department of Psychiatry, Box 126 MCV: Richmond, VA 23298.

Patterson, G. R. & Dishion, T. J. (1985). Contributions of families and peers to delinquency. *Criminology* **23**, 63–79.

Patterson, G. R. & Stouthamer-Loeber, M. (1984). The correlation of family management practices and delinquency. *Child Development* **55**, 1299–1307.

Pike, A., Reiss, D., Hetherington, E. M. & Plomin, R. (1996). Using MZ differences in search of nonshared environmental effects. *Journal of Child Psychology and Psychiatry* **37**, 695–704.

Prescott, C. A. & Kendler, K. S. (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *American Journal of Psychiatry* **156**, 34–40.

Rowe, D. C. (1983). Biometrical genetic models of self-reported delinquent behavior: a twin study. *Behavior Genetics* **13**, 473–489.

Rowe, D. C. (1986). Genetic and environmental components of antisocial behavior: a study of 265 twin pairs. *Criminology* **24**, 513–532.

SAS Institute, Inc. (1997). *SAS/SYSTAT® Software: Changes and Enhancements through Release 6.12*. SAS Institute, Inc.: Cary, NC.

Schwarz, N. (1999). Self-reports. How the questions shape the answers. *American Psychologist* **54**, 93–105.

Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health* **69**, 238–245.

Silberg, J., Rutter, M., Meyer, J., Maes, H., Hewitt, J., Simonoff, E., Pickles, A., Loeber, R. & Eaves, L. (1996). Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. *Journal of Child Psychology and Psychiatry* **37**, 803–816.

Simonoff, E., Pickles, A., Hewitt, J., Silberg, J., Rutter, M., Loeber, R., Meyer, J., Neale, M. & Eaves, L. (1995). Multiple raters of disruptive child behavior: using a genetic strategy to examine shared views and bias. *Behavior Genetics* **25**, 311–326.

Slutske, W. S., Heath, A. C., Dinwiddie, S. H., Madden, P., Bucholz, K. K., Dunne, M. P., Statham, D. J. & Martin, N. G. (1997). Modeling genetic and environmental influences in the etiology of conduct disorder: a study of 2,682 adult twins pairs. *Journal of Abnormal Psychology* **106**, 266–278.

Snyder, J., Dishion, T. J. & Patterson, G. R. (1986). Determinants and consequences of associating with deviant peers during preadolescence and adolescence. *Journal of Early Adolescence* **6**, 29–43.

Spoth, R., Goldberg, C. & Redmond, C. (1999). Engaging families in longitudinal preventive intervention research: discrete-time survival analysis of socioeconomic and social-emotional risk factors. *Journal of Consulting and Clinical Psychology* **67**, 157–163.

Thapar, A. & McGuffin, P. (1996). A twin study of antisocial and neurotic symptoms in childhood. *Psychological Medicine* **26**, 1111–1118.