Developmental Science 4:2 (2001), pp 195–208

REPORT

Comorbidity between verbal and non-verbal cognitive delays in 2-year-olds: a bivariate twin analysis

Shaun Purcell,[1] Thalia C. Eley,[1] Philip S. Dale,[2] Bonamy Oliver,[1] Steven A. Petrill,[3] Thomas S. Price,[1] Kimberly J. Saudino,[4] Emily Simonoff,[1] Jim Stevenson,[5] Eric Taylor[1] and Robert Plomin[1]

1. Institute of Psychiatry, London, UK
2. University of Washington, Seattle, USA
3. Wesleyan University, Middletown, Connecticut, USA
4. Boston University, USA
5. University of Southampton, UK

Abstract

The purpose of this paper is to investigate the genetic and environmental aetiology of the comorbidity between verbal delay and non-verbal delay in infancy. For more than 3000 pairs of 2-year-old twins born in England and Wales in 1994, we assessed verbal (vocabulary, V) and non-verbal (non-verbal, P) performance. V delay probands were selected who were in the lowest 5% of V; P delay probands from the lowest 5% of P. We assessed the comorbidity of delay both categorically, using twin cross-concordances, and dimensionally, by applying a bivariate extension of DeFries and Fulker (DF) group analysis. Both approaches are bidirectional, in that probands can be selected for either V delay (and analysed in relation to their co-twin's P score) or P delay (analysed in relation to their co-twin's V score). From a categorical perspective, twin cross-concordances indicated that comorbidity between V delay and P delay is substantially due to genetic factors whether probands are selected for V delay or for P delay. MZ and DZ cross-concordances were 24% and 8%, respectively, for probands selected for V delay and 27% and 6% for probands selected for P delay. From a dimensional perspective using bivariate DF analysis, selecting for V delay yielded high bivariate group heritability (0.59) and a genetic correlation of 1.0. In contrast, when selecting on P, DF analysis indicated lower bivariate group heritability (0.20) and only a modest genetic correlation with V assessed dimensionally (0.36). These results are discussed in terms of the difference between categorical and dimensional approaches to quantitative traits and the bidirectional nature of comorbidity. Such multivariate genetic results could lead to diagnostic systems that are based on causes rather than phenotypic descriptions of symptoms.

Childhood disorders often co-occur and understanding the aetiology of this comorbidity is a key issue for nosology. Several different sources of comorbidity have been considered (Caron & Rutter, 1991). Comorbidity could represent the co-occurrence of two independent disorders; it could be due to a causal relation between two disorders in the sense that one leads to the development of another; it could also be due to a common aetiology between two disorders. Additionally, ascertainment artefacts can produce apparent comorbidity. The first source of comorbidity needs to be considered for disorders of high incidence such as childhood disorders because frequent co-occurrence of independent disorders purely by chance is to be expected for common disorders. Longitudinal data are needed to address the second source of comorbidity. Multivariate genetic analysis is especially well suited to address the third source of comorbidity, shared aetiology. In this paper, we use multivariate genetic analysis to investigate genetic and environmental sources of comorbidity between verbal and non-verbal cognitive delays in 2-year-olds.

Address for correspondence: Shaun Purcell, Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK; e-mail: s.purcell@iop.kcl.ac.uk

Little is known about the aetiology of verbal and non-verbal delays in infancy or of their comorbidity. The only twin studies of general cognitive delay in infancy, which includes both verbal and non-verbal delay, suggest some genetic influence but their sample sizes were too small to yield reliable estimates of genetic and environmental influences (Wilson & Matheny, 1976; Nichols, 1984; Petrill *et al.*, 1997). No previous research on verbal delays in infancy has been reported but three studies of language impairment in childhood indicated high heritability (Lewis & Thompson, 1992; Bishop, North & Donlan, 1995; Tomblin & Buckwalter, 1998). In the first report of verbal and non-verbal delay in infancy defined as the lowest 5% of the distribution, we also found high heritability (73%) for delay in vocabulary (Dale *et al.*, 1998) and we found moderate heritability (40%) for delay in non-verbal performance (Eley *et al.*, 1999).

Co-occurrence between verbal and non-verbal delay seems likely for two reasons. First, mental retardation implies general cognitive disability assessed by IQ tests which include both verbal and performance subtests. Second, children with language impairment often show general cognitive disability, and for this reason researchers have focused on specific language impairment that excludes general mental retardation (Bishop, 1997; Rispens & van Yperen, 1997). The rationale for focusing on *specific* disorders such as specific language impairment is, first, that there is comorbidity with other disorders and, second, that avoiding overlap with these other disorders yields a 'purer' disorder. However, it is important that this reductionistic assumption be tested. For example, multivariate genetic analysis has shown that the same genetic factors affect depression and anxiety in adulthood (Kendler, Neale, Kessler, Heath & Eaves, 1992) and in childhood (Eley & Stephenson, 1999). Focusing on 'pure' depression independent of anxiety would eliminate genetic influence on depression, which would for example scupper attempts to identify genes for depression.

Although we are not aware of multivariate genetic analyses of verbal and non-verbal delays, there is an extensive literature on the normal range of variability for verbal and performance abilities which are two major components of general cognitive ability or *g* (Brody, 1992; Carroll, 1993). Verbal and performance abilities generally correlate about 0.4. Multivariate genetic research that analyses the covariance between traits suggests that the same genetic factors largely affect both domains (Plomin, DeFries, McClearn & Rutter, 1997), although our study of 2-year-olds suggests that genetic factors only show modest overlap (Price, Eley, Dale, Stevenson & Plomin, 2000). The few multivariate genetic studies of disorders also suggest that, like anxiety and depression, co-occurrence of disorders is substantially due to genetic factors, e.g. for hyperactivity and spelling disabilities (Stevenson, Pennington, Gilger, DeFries & Gillis, 1993) and for reading and mathematics disabilities (Knopik, Alarcón & DeFries, 1997).

Univariate DeFries and Fulker group analysis

Univariate DeFries and Fulker (DF) group analysis (DeFries & Fulker, 1985, 1988) addresses the fundamental issue of the nature of variability that places individuals at the extreme of a dimension, and how it is related to variability in the normal range. For example, it has been argued (Leonard, 1987) that children classified as language impaired represent the extreme of normal variation; i.e. they are simply slow learners, in contrast to the assumption that they manifest a qualitatively distinct clinical syndrome.

DF group analysis estimates the magnitude of genetic and environmental effects for an extreme group using a dimensional measure. For example, in a previous report (Dale *et al.*, 1998), we selected 2-year-old twins who were in the lowest 5% of the distribution of vocabulary scores. Simple pairwise concordances indicating the proportion of pairs in which both members of a twin pair were in the lowest 5% group suggested substantial genetic influence on this 'diagnostic' category. However, twin concordances provide only a rough index of genetic and environmental influence based on a dichotomization of what could be a dimension. Liability-threshold models have been used to convert dichotomous diagnostic data to correlations on the assumption of an underlying continuous liability (Falconer, 1965; Smith, 1974). However, instead of assessing a dichotomy (e.g. diagnosing a disorder) and then assuming a continuous dimension, DF group analysis assesses the continuous dimension directly.

The essence of univariate DF group analysis is that if vocabulary delay is associated genetically with the continuum of normal variation in vocabulary, the vocabulary scores of twin partners (co-twins) will be more similar to the vocabulary scores of language-disabled probands for monozygotic (MZ) co-twins than for dizygotic (DZ) co-twins. The vocabulary scores of the MZ and DZ probands and their co-twins, expressed as standardized deviation units from the mean of the total sample, indicate the extent to which genetics contributes to the mean vocabulary difference between the probands and the population.

Application of DF group model-fitting analysis (see Methods) to these vocabulary data yielded a significant

estimate of 0.73 (0.38–1.00 95% confidence interval, CI) for group differences heritability, indicating that most of the mean difference between the probands and the population assessed on a dimensional measure of vocabulary can be ascribed to genetic factors. This is called *group differences heritability* to distinguish it from the usual heritability estimate, which could be called 'individual differences heritability' because the latter refers to differences among individuals whereas the former denotes mean differences between an extreme group and the population. The standard individual differences heritability estimate for these data was 25%, significantly lower than the group heritability of 73%. These results make the point that heritability of an extreme group can differ from the heritability of variability in the normal range even when the same dimensional measure is used to define the extreme group and to assess variability in the normal range.

Heritabilities of disorders can differ from heritabilities of dimensions if there is a mixture of distributions. That is, although children are assessed using a single-dimensional measure of vocabulary, it is possible that children with the slowest vocabulary development differ from the rest of the population due to genetic or environmental factors that do not contribute to variability in the normal range. This suggests that when genes associated with vocabulary are identified, some of these genes will be specifically associated with vocabulary delay, and not with normal variability. Nonetheless, group differences heritability is usually similar to individual differences heritability (e.g. Deater-Deckard, Reiss, Hetherington & Plomin, 1997). Indeed, our result for vocabulary delay at 2 years is the first time that a significant difference has been reported between group differences heritability and individual differences heritability.

Not only can group differences heritability differ from individual differences heritability, group differences heritability can differ for low and high extremes. For example, in our study, although group differences heritability for the children with the lowest vocabulary scores at 2 years was substantial (Dale *et al.*, 1998), group heritability for the children with the highest vocabulary scores was modest and similar to individual differences heritability for vocabulary (Dale, Purcell & Plomin, in preparation).

Bivariate DeFries and Fulker group analysis

When disorders or dimensions are studied one at a time, many show genetic influence, but it is highly unlikely that each of these is influenced by a completely different set of genes. Multivariate genetic analysis, one of the most important advances in behavioural genetics during the past two decades, focuses on the aetiology of the covariance between traits rather than the variance of each trait considered separately (Eaves, Martin & Eysenck, 1977; Plomin & DeFries, 1979). Specifically, it assesses genetic and environmental factors responsible for the phenotypic correlation between two traits. If the same genes affect different traits, a correlation will be observed among the traits. One of the genetic causes of correlation is that the same genes influence both traits, an effect called *pleiotropy*. (Genetic correlations can also result from temporary linkages due to recent admixtures of populations or nonrandom mating.) Similar reasoning applies to the way in which environmental influences may lead to phenotypic correlation.

Multivariate genetic analysis is based on cross-twin correlations. That is, one twin's X is correlated with the co-twin's Y. In univariate twin analyses (i.e. where one twin's X is correlated with the co-twin's X) the variance of a trait is attributed to genetic factors to the extent that the MZ correlation exceeds the DZ correlation. In bivariate twin analysis, the covariance between two traits is attributed to genetic factors to the extent that the MZ cross-twin correlation exceeds the DZ cross-twin correlation. Bivariate heritability estimates the extent to which genetic factors account for the phenotypic correlation between two traits (Plomin & DeFries, 1979). Bivariate heritability is a function of the heritability of the two traits and a construct called the *genetic correlation* (see Figure 1). The genetic correlation estimates the extent to which two traits are influenced by the same genes regardless of the magnitude of their contribution to phenotypic variances of the two traits. That is, the heritability of two traits can be modest yet the genetic correlation between them can be high and vice versa. Thus there are two multivariate questions that address different issues. The first, bivariate heritability, addresses the extent to which the phenotypic correlation between two traits is mediated by genetic factors. The second question addresses the genetic correlation between two traits independent of their heritabilities.

The bivariate extension of DF group analysis addresses these same issues for the extremes of dimensions (DeFries, Olson, Pennington & Smith, 1991). Just as twin concordances have been used in the univariate case, cross-twin concordances have been used in the bivariate case to analyse comorbidity. Cross-twin concordance is the proportion of twin pairs in which probands for disorder X have co-twins who meet criteria for disorder Y. A higher cross-concordance of MZ twin

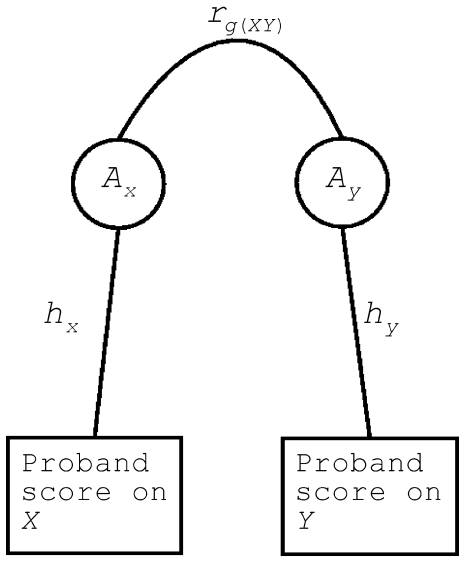

Figure 1 *The relation between bivariate heritability and genetic correlation: g represents genetic factors influencing either X or Y; $r_{g(XY)}$ represents the genetic correlation between these two factors; bivariate heritability equals the path $r_{g(XY)}h_Xh_Y$ where h_X equals the square root of univariate heritability for X and h_Y equals the square root of univariate heritability for Y.*

pairs compared to DZ twin pairs suggests a genetic contribution to the comorbidity between the two disorders. Such analyses are bidirectional in the sense that selecting probands for disorder *X* or disorder *Y* are separate studies that can yield different results. That is, results can differ for probands selected for disorder *X* whose co-twins are examined for disorder *Y* compared with probands selected for disorder *Y* whose co-twins are examined for disorder *X*. The reason for this possible asymmetry is that, although *X* and *Y* are comorbid which means that some of the same individuals are selected for both *X* and *Y*, many of the *X* probands are not *Y* probands and vice versa.

Bivariate DF group analysis goes beyond this dichotomization of disorders to a dimensional perspective. Rather than selecting probands as extreme on *X* and comparing the quantitative scores of their MZ and DZ co-twins on *X* as in univariate DF group analysis, bivariate DF group analysis selects probands on *X* and compares the quantitative scores of their co-twins on *Y*. The extent to which the cross-twin regression to the population mean is greater for DZ co-twins than MZ co-twins indicates the extent to which proband deficits in *X* are due to genetic factors that also influence the co-twins' quantitative scores on *Y*.

Bivariate DF group analysis is also bidirectional. Selecting probands on *X* and comparing co-twins on *Y* can yield different results from selecting probands on *Y* and comparing co-twins on *X*. This issue was first explored in an analysis of hyperactivity probands and spelling disabled probands identified independently (Stevenson *et al.*, 1993). In this case, the bivariate DF results were similar for hyperactivity probands and for spelling disabled probands, both indicating that about three-quarters of the overlap was due to shared genetic influences.

In the present study, bivariate DF group analysis is applied to the comorbidity between vocabulary (V) delays and non-verbal performance (P) delays using twin data from a large, epidemiologically recruited sample of 2-year-olds. The use of an epidemiological sample is important for studying comorbidity because clinically derived samples are likely to include the most severely affected children who may show greater comorbidity (Caron & Rutter, 1991). In univariate DF group analyses we found that V delay is highly heritable but P delay is only moderately heritable – group differences heritabilities were 73% and 40%, respectively – when the two types of delays were considered independently (Eley *et al.*, 1999). These differences in group heritability were even greater when comorbid cases were excluded: group differences heritabilities were 77% for V-not-P and 23% for P-not-V. We concluded that because the aetiologies of V and P delays differ they are better considered separately rather than combined in a composite measure of general cognitive delay. However, bivariate analyses are needed to address the issue of the aetiology of comorbidity. That is, although group differences heritability differs for V delay and P delay, genetic factors could completely account for the comorbidity of V delay and P delay and the genetic correlation between them could be substantial. Nonetheless, the greater group differences heritability for V delay than for P delay led us to consider the possibility that bivariate DF group analysis might show asymmetrical results when applied to V probands compared with P probands. This would suggest that while genes for delay on one variable also influence the other, the converse may not be true.

Method

Sample

The sampling frame for the present study, called the Twins Early Development Study (TEDS), consisted of all twins born in England and Wales in 1994. Parents of 7756 pairs of twins identified from their children's birth records were contacted by the Office for National

Statistics (ONS) after checking for infant mortality when the children were 1 year old. A total of 3442 families returned booklets sent at the twins' second birthday. The sample is representative of the general population in terms of maternal ethnicity and education when compared to ONS 1994 census data. The lowest 5% vocabulary (V) group and the lowest 5% non-verbal performance (P) group were similar to the rest of the sample in parental employment, marital status, family size and ethnicity but somewhat lower in maternal education (Dale *et al*., 1998).

A parent-rated instrument was used to assign twin zygosity, by asking parents, for example, about instances of confusion in identifying the twins. This approach yielded 95% certainty for same-sex pairs, a rate which is typical of other studies (Goldsmith, 1991). We excluded 70 pairs whose zygosity was uncertain based on the parent-rated instrument (from whom we are currently obtaining DNA). Pairs in which at least one of the children had specific medical problems that might impact upon V and P scores were excluded, as well as outliers on other measures. Forty-one pairs were excluded in which at least one twin had a hearing problem or specific medical syndrome such as Down's syndrome and other chromosomal anomalies, cystic fibrosis and cerebral palsy; 46 pairs who were extreme outliers for birth weight, time spent in hospital, special care after birth, gestational age or maternal alcohol consumption during pregnancy; 116 pairs in which English is not the basic language spoken in the family; 119 pairs in which the vocabulary measure was completed 6 months or more after the twins' second birthday; 11 pairs for whom one or other twin did not have a vocabulary test score; 33 pairs were excluded for whom there was no reported language, but for whom the parents described the child's language in terms of different sentence types in a subsequent section of the booklets (suggesting that the zero score on the vocabulary measure was likely to be due to missing data rather than a genuine lack of use of any of the words listed); 47 pairs for whom the non-verbal cognitive measure was completed 6 months or more after the second birthday; 89 pairs for whom there was no valid non-verbal cognitive measure score. As expected, proportionately more of the children excluded for medical or perinatal problems would have been included in the lowest 5% group (17% for medical problems and 12% for perinatal problems), and those excluded had substantially lower scores for both V and P.

The final sample consisted of 2870 pairs: 989 pairs of MZ twins, 948 pairs of same-sex DZ twins, and 933 pairs of opposite-sex DZ twins.

Measures

Vocabulary

Productive vocabulary was assessed at 2 years using a UK adaptation of the MacArthur Communicative Development Inventory (MCDI) (Fenson *et al*., 1994). Two years is a particularly appropriate age to study vocabulary development because it follows a period of rapid acceleration in the use of words at about 18 months and the beginning use of word combinations at about 20 months. In the MCDI, parents report on their children's production of monomorphemic root words and also grammatical development by checking the words that they have heard their child speak, and the number of positive responses is summed. Validity results include a correlation of 0.73 with a standard tester-administered measure of expressive vocabulary for unselected 2-year-olds, and 0.85 for a language-impaired sample of 3-year-olds (Fenson *et al*., 1994). Two lists of 100 words that predict the full MCDI list of 680 words with very high accuracy ($r = 0.98$) have been identified (Fenson *et al*., 1997). One of these short-form lists (Form A), with a few minor changes to 'anglicize' items, along with questions about other aspects of language not reported here, constituted the language measure of the present study (MCDI : UKSF, hereafter referred to as MCDI).

Non-verbal performance

Non-verbal performance was assessed when the children were 2 years old using the Parent Report of Children's Abilities (PARCA; Saudino *et al*., 1998). The PARCA is an hour-long test in which parents administer standard cognitive tasks including design copying, match-to-sample, block building and imitative action, and also report on specific behaviours (e.g. 'Does your child recognize himself/herself when looking in the mirror?'). In a sample of 107 2-year-olds, internal consistency as estimated by Cronbach's alpha coefficient was 0.75 for the parent-report items and 0.83 for the parent-administered items. Two weeks after parents administered the PARCA, a standard measure of general cognitive development, the Bayley Mental Development Index (Bayley, 1993), was administered by testers in the homes of the 2-year-olds. The PARCA predicted scores on the Bayley test nearly as well as the old version of the Bayley test predicts the new version, even though the PARCA specifically excludes items with primarily verbal content (Saudino *et al*., 1998). These findings indicate both the reliability and validity of the PARCA. A review of 23 studies relating parental ratings to

standard tester-administered measures also supports the validity of parental measures (Dinnebeil & Rule, 1994).

Analysis

Delay in vocabulary (V) and non-verbal performance (P) was defined in terms of a fifth percentile cut-off. This represents a deviation from the mean equivalent to an IQ score of 75. Two proband groups were identified: a V delay group consisting of children scoring in the lowest 5% on V; a P delay group consisting of the children scoring in the lowest 5% on P. The number of children in both V delay and P delay groups is a crude indicator of the extent of V delay and P delay comorbidity: a greater-than-chance level of co-occurrence is evidence of significant comorbidity. The appropriateness of a dimensional rather than a categorical approach to comorbidity between disorders X and Y can be tested by asking whether the X-related dimension functions as a risk factor for disorder Y at levels below the diagnostic threshold of X (Caron & Rutter, 1991).

It is possible to explore the role of genetic and environmental influences on the comorbidity of V delay and P delay by comparing probandwise cross-concordance rates (the number of probands in cross-concordant pairs divided by the number of probands) between MZ and DZ twins. However, using concordance assumes that both disorders are dichotomous rather than dimensional. As noted earlier, bivariate DF group analysis goes beyond this dichotomization of disorders to a dimensional perspective by selecting probands at the extreme of a dimension X and comparing the quantitative scores of their co-twins on a dimension Y. If the extreme of X is linked genetically with a dimension Y, the quantitative trait scores for Y of co-twins of X probands will be more similar to the population mean for Y for DZ twins than for MZ twins. DF group analysis uses this differential regression to the mean for the co-twins of MZ and DZ twin probands to estimate genetic influence on the mean quantitative trait difference between the probands and the population, called 'group differences heritability'.

The basic univariate DF model is represented as the regression $C_X = \beta_1 P_X + \beta_2 R + \alpha$, in which the co-twin's score on X (C_X) is predicted from the proband's score on X (P_X) and the coefficient of relatedness R, which is 1.0 for MZ and 0.5 for DZ pairs. Because the proband mean is transformed to a mean of 1.0 and the unselected population to a mean of 0.0, the transformed mean of the co-twin's score for MZ and DZ twins estimates their group differences familiality. The regression weight β_2 represents the extent to which twin resemblance differs for MZ and DZ twins independent of average MZ and DZ resemblance (β_1). Thus β_2 estimates group differences heritability, which is the genetic contribution to the difference on X between the probands and the population. Group shared environment, twin resemblance not explained by genetic factors, is estimated by subtracting group differences heritability from the MZ transformed co-twin mean. Transformed co-twin means represent group familiality, or total within-pair similarity including both genetic and shared environmental influences. The univariate estimates of group heritability, and group shared environment, were calculated for V and P using this basic DF model. All scores were first corrected for sex and age at testing using a regression procedure, as is standard in twin research because these variables can inflate twin similarity (McGue & Bouchard, 1984).

To assess the aetiology of the comorbidity between V and P delay, the basic univariate DF regression can be extended to the bivariate case. As in the univariate case, children who fall in the lowest 5% of the V, or P, distribution are selected as V, or P, probands. However, as bivariate DF analysis is bidirectional, we can select probands low on V and examine their co-twins' P scores (hereafter referred to as 'V$\rightarrow$P'); alternatively, we can select low P probands, examining co-twins' V scores (hereafter referred to as 'P$\rightarrow$V').

Both distributions have been previously standardized such that the population mean is zero. For the V$\rightarrow$P analysis, both probands' V and co-twins' P scores are transformed by dividing them by the zygosity-specific proband mean for V, to give zygosity-specific proband means of 1.0 and an unselected population mean of zero. The following regression model is fitted to the data:

$$C_P = \beta_1 P_V + \beta_2 R + \alpha$$

where C_P is the co-twin's P score, P_V is the proband's V score, R is the coefficient of relatedness and α is a constant term. β_1 therefore represents the partial regression of the co-twin's P score on the proband's V score, a measure of the average cross-variable twin resemblance across zygosity. β_2 is the partial regression of co-twin's P on the coefficient of relatedness. The transformation used means that β_2 estimates the genetic contribution to the mean difference on P between V probands and the population, which Figure 1 represents as the chain of paths $h_V h_P r_{g(VP)}$. As explained earlier, this estimates the extent to which genetic factors are responsible for the lowered P scores of V delay probands. Group shared environment ($c^2_{g(VP)}$) can be estimated as the difference between the MZ co-twin transformed mean and group heritability estimate (*not* from β_1 as has been assumed in some previous studies).

In a similar fashion, this regression model can be fitted for P probands using their co-twin's V scores (P→V). This bidirectionality of analysis represents an important conceptual point regarding the nature of comorbidity. Depending upon the nature of the causes of comorbidity, asymmetry in results can occur. That is, being low on *X* may not be a risk factor for being low on *Y* even if being low on *Y* is a risk factor for being low on *X*. Therefore, symmetry or asymmetry in bivariate heritabilities for V→P and P→V may help us to understand the nature of comorbidity between V and P delay.

As explained earlier, two important statistics can be derived from β_2. The first is the ratio between β_2 and the observed phenotypic association between V delay and the P dimension, which is assessed as the mean P difference between the V probands and the population. This ratio provides an index of the extent to which the phenotypic association between V delay and the P dimension can be attributed to genetic factors. The second construct is the genetic correlation which estimates the extent to which genetic factors that affect V delay also affect the P dimension regardless of the heritability of V delay or the P dimension. The regression weight β_2 estimates $h_V h_P r_{g(VP)}$, and thus r_g can be estimated by dividing β_2 by $h_V h_P$ (where $h_V h_P$ is the product of the square roots of univariate group heritabilities for V delay and P delay).

The main analyses are restricted to same-sex pairs only – i.e. opposite-sex DZ twin pairs are excluded. This is a standard procedure in behavioural genetic studies – any mean sex difference will necessarily make DZ opposite-sex pairs even less similar to each other than DZ same-sex pairs, thus inflating heritability estimates. However, DZ opposite-sex pairs can be entered to check for such effects: in our study, the same pattern of results was found. In order to simplify the presentation, only results for same-sex pairs are reported below.

Because probands were selected from a double-entered data set (whereby concordant pair members are entered twice in the selected sample, once as proband, once as co-twin) standard error estimates and significance tests were corrected for the inflated *N* (see the Appendix).

Results

Table 1 describes the two groups selected for V delay and P delay and compares them to the entire population.

Table 1 *Table of means, standard deviations and N for V delay probands, P delay probands and the entire sample on V and P by sex and zygosity*

	V probands		P probands		Entire sample	
	V	P	V	P	V	P
All same-sex twins	5.65 (*2.4*) 178	−0.89 (*0.9*) 178	24.40 (*17.5*) 192	−2.00 (*0.3*) 192	47.82 (*25.1*) 3874	0.00 (*0.9*) 3874
All twins	5.75 (*2.4*) 261	−0.86 (*1.0*) 261	25.57 (*18.0*) 289	−1.99 (*0.3*) 289	48.34 (*25.1*) 5740	0.00 (*0.9*) 5740
MZ twins	5.41 (*2.3*) 117	−0.97 (*0.94*) 117	23.00 (*18.2*) 105	−2.02 (*0.4*) 105	46.74 (*25.5*) 1978	−0.01 (*1.0*) 1978
DZ same-sex twins	6.09 (*2.6*) 61	−0.75 (*1.02*) 61	26.06 (*16.4*) 87	−1.98 (*0.3*) 87	48.96 (*24.6*) 1896	0.02 (*0.9*) 1896
DZ opposite-sex twins	5.97 (*2.3*) 83	−0.78 (*1.0*) 83	27.90 (*18.9*) 97	−1.98 (*0.4*) 97	49.40 (*25.05*) 1866	0.01 (*0.9*) 1866
MZ and DZ same-sex males	5.70 (*2.3*) 168	−0.94 (*1.0*) 168	22.50 (*16.9*) 173	−2.00 (*0.4*) 173	43.93 (*24.7*) 2769	−0.10 (*0.9*) 2769
MZ and DZ same-sex females	5.83 (*2.6*) 93	−0.71 (*0.9*) 93	30.10 (*18.7*) 116	−1.97 (*0.3*) 116	52.44 (*24.7*) 2971	0.10 (*0.9*) 2971
MZ males	5.17 (*2.1*) 80	−1.02 (*0.9*) 80	20.20 (*16.6*) 66	−2.03 (*0.4*) 66	40.90 (*24.9*) 896	−0.11 (*1.0*) 896
MZ females	5.92 (*2.7*) 37	−0.85 (*0.9*) 37	27.60 (*20.1*) 39	−2.00 (*0.4*) 39	51.60 (*24.9*) 1082	0.07 (*0.9*) 1082
DZ same-sex pair males	6.54 (*2.4*) 37	−0.78 (*1.0*) 37	23.60 (*16.6*) 48	−2.02 (*0.3*) 48	43.73 (*23.7*) 940	−0.11 (*0.9*) 940
DZ same-sex pair females	5.40 (*2.7*) 24	−0.69 (*1.0*) 24	29.10 (*15.9*) 39	−1.94 (*0.3*) 39	54.10 (*24.4*) 956	0.14 (*0.9*) 956
DZ opposite-sex males	5.94 (*2.3*) 51	−0.92 (*1.1*) 51	24.20 (*17.4*) 59	−1.97 (*0.4*) 59	47.10 (*25.2*) 933	−0.08 (*1.0*) 933
DZ opposite-sex females	6.03 (*2.4*) 32	−0.55 (*0.9*) 32	33.60 (*20.0*) 38	−1.99 (*0.4*) 38	51.70 (*24.7*) 933	0.10 (*0.9*) 933

Note: The mean and standard deviation (in italics in parentheses) are given on the first line of each row and the number of individuals is given on the second line.

Children with a V raw score of 9 or less were selected as V delay probands; children with a P score of −1.61 or less were selected as P probands. The 178 V probands were 1.7 standard deviation units below the mean for V; the 192 P probands were 2.0 units below the mean. As discussed in our previous papers (e.g. Dale *et al.*, 1998), the V score distribution indicated a wide range of variability in vocabulary at 2 years of age and showed slight but significant effects of twinning (compared to singleton means) and sex consistent with existing research. P scores also reflected a wide range of variability at 2 years of age – e.g. the average 2-year-old can copy a scribble, a horizontal line and possibly a circle, whereas children in the lowest 5% are only able to copy a scribble (Eley & Stevenson, 1999; Eley *et al.*, 1999).

Twelve percent of all probands (43 children) had scores in the lowest 5% for both V and P. These 'comorbid' children had V scores as low as the rest of the V probands but P scores that were even lower than the rest of the P probands. The proportion of V delay probands who were also P delay probands was 43/200 = 21.5%. The proportion of P delay probands who were also V delay probands was 43/192 = 22.4%. The 43 comorbid probands represent 1.1% of the entire population of 3874 same-sex twins, significantly greater ($z = 10.6$, $p < 0.002$) than the proportion expected to be comorbid on the basis of chance ($0.05 \times 0.05 = 0.25\%$). In the entire sample, V and P correlated 0.44.

As suggested by the phenotypic correlation of 0.44 between V and P scores, the relationship between V and P is seen throughout the distribution. A test of dimensionality of disorder confirms this, suggesting that a dimensional approach to V delay and P delay comorbidity is warranted. Table 2 shows the proportion of V probands falling into each of six score ranges of standardized P and vice versa. The risk of being a proband is linearly related to the subject's score on the other measure for both V proband membership across the range of P scores (χ^2 for trends = 118.14; degrees of freedom (df) = 1; $p < 0.0001$) and P proband membership across the range of V scores (χ^2 for trends = 156.26; df = 1; $p < 0.0001$).

Table 2 *A test of dimensionality: the proportions of V probands across the range of P scores and the proportions of P probands across the range of V scores (same-sex twins only)*

Score on standardized P	V probands	Score on standardized V	P probands
< −1.0	13.7%	< −1.0	14.2%
−0.99 to −0.5	5.7%	−0.99 to −0.5	6.1%
−0.49 to −0.01	3.8%	−0.49 to −0.01	4.9%
0 to 0.49	2.2%	0 to 0.49	1.7%
0.5 to 0.99	2.9%	0.5 to 0.99	0.7%
> 1.0	0.3%	> 1.0	0.5%

Univariate DF analyses

As the univariate results for both V and P have been reported previously (Dale *et al.*, 1998; Eley *et al.*, 1999), we will summarize them briefly. Using a similar 5% threshold to define V delay, probandwise concordances were 84% for MZ twins (68 pairs), 30% for same-sex DZ twins (52 pairs) and 51% for opposite-sex DZ twins (62 pairs), suggesting substantial genetic influence. Application of DF group analysis to these data for same-sex twins yielded a significant estimate of 0.77 (0.52–1.00 CI) for group differences heritability, indicating that most of the mean difference between the probands and the population can be ascribed to genetic factors. The estimate of group shared environment was 0.20 (0.00–0.71 CI). The remaining 3% of the difference was due to nonshared environmental influence and error of measurement. The estimates presented in Dale *et al.* (0.73 for h^2g and 0.18 for c^2g) differ marginally because a number of extra twin pairs have been excluded in the present analyses, on the basis of exclusion criteria specific to P.

For the P delay group, probandwise concordances were 61% for MZ twins (73 pairs), 32% for same-sex DZ twins (73 pairs) and 37% for opposite-sex DZ pairs (79 pairs), suggesting genetic influence. Application of DF group analysis to the data yielded a significant group heritability estimate of 0.40 (0.14–0.67 CI), and a shared environment influence of 0.41 (0.00–0.87 CI). The remaining 19% of the difference was due to nonshared environmental influence and error of measurement.

Several analyses adopted a univariate DF approach to V and P as components of general cognitive delay (Eley *et al.*, 1999). Excluding comorbid cases, specific (V not P versus P not V) analyses revealed an even greater difference in group heritabilities (78% versus 22%). Running DF univariate analyses for V and P on only the probands comorbid for both V and P delay yielded high group heritabilities for both V (77%) and P (93%). These results suggested that, as the genetic and environmental origins of V and P delay in infancy appear to differ, they are better considered separately rather than combined into a composite measure of general cognitive delay.

These univariate DF analyses suggest genetic influence on V delay and P delay that ranges from moderate to strong. However, these univariate analyses do not address the aetiology of the association between V delay

and P delay, which is the provenance of bivariate analyses.

Bivariate DF results

As explained earlier, bivariate DF analysis addresses the genetic and environmental origins of the overlap between disorder X and dimension Y rather than the aetiology of the disorder or the dimension considered separately as in univariate analysis. The results of bivariate analyses are described first for probands for low V whose co-twins are examined for P scores (V → P) and then for probands selected for low P whose co-twins are examined for V scores (P → V).

V → P

In the V delay group, 117 MZ probands (from 68 pairs) and 61 same-sex DZ probands (from 52 pairs) were selected from the lowest 5% of the distribution of V. As noted earlier, the within-individual comorbidity for V delay probands was 21.5%. MZ probandwise cross-concordance was 24%, which indicates that comorbidity across members of MZ twin pairs is just as great as within individuals. That is, of the 117 V delay MZ probands, 24% had a co-twin with a P score in the bottom 5% of the P distribution. Same-sex DZ probandwise cross-concordance was much lower (8%), suggesting genetic mediation of comorbidity. Opposite-sex DZ probandwise cross-concordance was 14% for 62 pairs, similar to that for same-sex DZ twins, suggesting that gender differences do not substantially affect these analyses of comorbidity.

The greater twin cross-concordance for V → P for MZ compared with DZ twins suggests genetic mediation of comorbidity for V delayed probands. We used bivariate DF analysis to investigate the association further for dimensional P scores of all co-twins of the V probands.

Table 3 lists the standardized means for MZ and same-sex DZ twins for V and their co-twins' standardized means for P. The much greater regression towards the population P mean for DZ co-twins compared with MZ co-twins suggests substantial genetic influence on the relationship between V delay and P. After transforming both proband V and co-twin P scores by dividing by the proband zygosity-specific mean (ensuring group means of 1.0 for MZ and DZ probands) the co-twin means, estimates of group familiality, were found to be 0.59 for MZ co-twins and 0.22 for DZ co-twins.

Table 3 *Proband V scores and co-twin P scores for V → P bivariate DF analysis with scores standardized for the entire sample (same-sex twins only)*

	Proband standardized V mean (SD)	Co-twin standardized P mean (SD)
MZ	−1.64 (0.219)	−0.97 (0.990)
DZ	−1.65 (0.271)	−0.37 (0.992)

Application of bivariate DF group analysis yields a β_2 of 0.73 (0.20–0.99 CI). Just as the MZ twin cross-concordance is somewhat greater than the within-individual comorbidity, this estimate of β_2 is larger than the MZ transformed co-twin mean (0.59) which is a logical upper bound for the β_2 estimate. Constraining the DF analysis to this upper bound yields a β_2 estimate of 59% (20%–98% CI). Group shared environment contributes nothing (0.0%, 0%–100%) to the relationship between V delay and P. Figure 2 shows the relationship between the unselected sample, probands and MZ and DZ co-twins in this bivariate DF group analysis. The scatter-plot represents the entire sample, before any selection. The shaded area of the V distribution on the Y axis represents the bottom 5% selected as V probands. The MZ and DZ co-twin distributions underneath the scatter-plot represent the distribution of scores for co-twins of V probands; the shaded areas represent the proportion of co-twins scoring in the bottom 5% for P.

As discussed earlier, the ratio between β_2 for V → P and the 'phenotypic' intra-individual association between V delay and the P dimension indicates the extent to which genetic factors account for the association. The intra-individual association can be conceptualized as the extent to which P scores of V probands regress towards the population mean. It can be assessed using DF analysis by regressing the transformed proband P scores on the transformed proband V scores in the V delay sample after excluding the genetic relatedness term from the model. This estimate was 0.54. Thus, the ratio of β_2 (0.59) to this estimate of intra-individual association slightly exceeds 1, suggesting that all of the phenotypic association between V delay and P can be attributed to genetic factors.

As bivariate heritability equals $h_V h_P r_{g(PV)}$, the genetic correlation ($r_{g(PV)}$) can be estimated. Univariate analyses supply values for h_V ($\sqrt{0.77}$) and h_P ($\sqrt{0.40}$). The genetic correlation can be computed as follows:

$$r_{g(VP)} = \frac{\beta_{2(VP)}}{\sqrt{\beta_{2(V)}}\ \sqrt{\beta_{2(P)}}}$$

$$= \frac{0.59}{\sqrt{0.77} \times \sqrt{0.40}} = 1.06 \approx 1.00$$

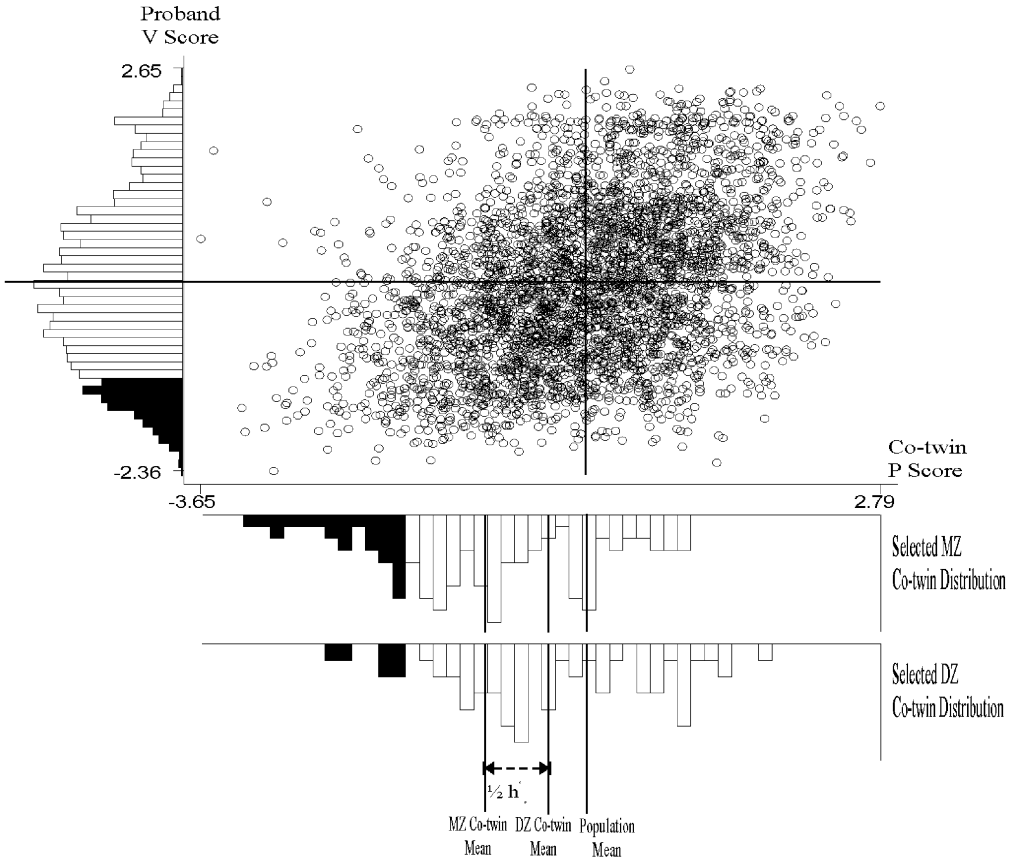

Figure 2 *V⟶P bivariate DF analysis (shaded areas represent delayed probands and cross-concordant co-twins; difference between MZ and DZ co-twin means estimates* h_g^2 *(after appropriate transformation)).*

The genetic correlation of 1.0 indicates that the genes that affect V delay are the same genes responsible for P scores for these V delay probands.

In summary, for V probands genetic factors govern comorbidity with P whether P is assessed dichotomously or dimensionally.

P⟶V

For probands selected in the lowest 5% of the P distribution, bivariate probandwise cross-concordance rates were similar to those for V⟶P. For the 105 MZ selected twin probands (73 pairs) cross-concordance was 27%, as great as comorbidity within individuals. For 87 same-sex DZ selected twin probands (73 pairs), cross-concordance was only 6%, suggesting substantial genetic mediation of comorbidity. For 79 opposite-sex twin pairs, cross-concordance was 12%.

For bivariate DF analysis, Table 4 lists the standardized means for MZ and same-sex DZ twins for P and their co-twins' standardized means for V. Unlike the V⟶P results in Table 3, these P⟶V results show that DZ co-twins regress to the population mean only moderately more than the MZ co-twins. This suggests that the genetic contribution to the relationship between P delay and V is only moderate in the case of P⟶V.

Table 4 *Proband P scores and co-twin V scores for P⟶V bivariate DF analysis with scores standarized for the entire sample (same-sex twins only)*

	Proband standardized P mean (SD)	Co-twin standardized V mean (SD)
MZ	−2.13 (0.44)	−0.88 (0.81)
DZ	−2.11 (0.38)	−0.66 (0.81)

Transforming proband P and co-twin V scores (dividing by the zygosity-specific proband mean) revealed transformed co-twin means of 0.41 for MZ co-twins and 0.31 for DZ co-twins, suggesting less group familiality than in the case of V⟶P. Application of bivariate DF group analysis gives a β_2 estimate of 20% (0%–43% CI) and group shared environment of 21% (0%–84% CI) for the relationship between P delay and

V. Figure 3 shows the relationship between groups for P → V.

The phenotypic association between V and P in the P proband group is 0.40. The ratio of the bivariate β_2 estimate to this observed correlation suggests that half of the phenotypic covariance between V and P in the P delay group can be attributed to genetic causes. The genetic correlation for P → V can be estimated as 0.36:

$$r_{g(PV)} = \frac{\beta_{2(VP)}}{\sqrt{\beta_{2(V)}}\ \sqrt{\beta_{2(P)}}}$$

$$= \frac{0.20}{\sqrt{0.77} \times \sqrt{0.40}} = 0.36$$

In other words, unlike V → P only about a third of the genetic effects on P delay also affect V scores in these P delayed probands.

In summary, P → V yields different results from V → P. Although P → V twin cross-concordances suggest that genetic factors are entirely responsible for comorbidity between P delay and V delay assessed dichotomously for P delay probands, bivariate DF analysis for P → V indicates that for P probands genetic factors play only a moderate role in the relationship with V assessed dimensionally.

Discussion

Verbal delay and non-verbal cognitive delay in infancy co-occur, as do many childhood disorders. In the entire population of 3874 same-sex twins, 43 twins (1.1%) were in the lowest 5% for both V and P, more than four times the rate expected by chance co-occurrence. Comorbidity is symmetrical in the sense that 22% of V delay probands were also P delay probands and 22% of P delay probands were also V delay probands. Moreover, some symmetry is also seen when co-occurrence is viewed dimensionally at the phenotypic level of analysis. The intra-individual DF 'phenotypic' association between V delay and the P dimension is 0.54 and the association between P delay and the V dimension is 0.40. The focus of the present paper is on the genetic and environmental sources of this phenotypic comorbidity viewed dimensionally as well as dichotomously.

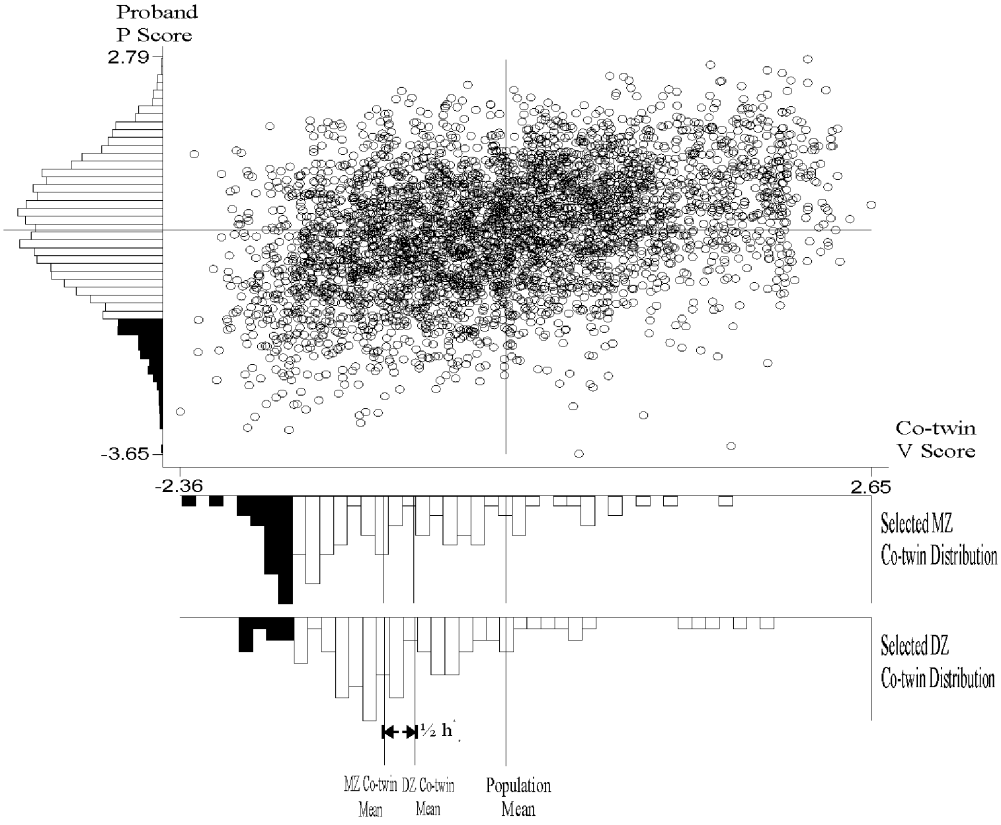

Figure 3 *P → V bivariate DF analysis (shaded areas represent delayed probands and cross-concordant co-twins; difference between MZ and DZ co-twin means estimates h_g^2 (after appropriate transformation)).*

Previous bivariate genetic analyses of this data set of individual differences throughout the normal range of variation rather than the extremes of the distributions yielded results that mirror the results of univariate genetic analyses (Price *et al.*, 2000). That is, both V and P scores are moderately heritable and genetic factors account for their overlap to a moderate extent. Heritabilities were 24% for V and 23% for P; V and P correlated 0.42, and genetic factors accounted for 19% of this phenotypic correlation.

In this context, the present categorical results are interesting because they suggest that almost all of the comorbidity between V delay and P delay is mediated genetically. That is, twin cross-concordances indicate that comorbidity within MZ twins (i.e. one twin's V delay and the co-twin's P delay) is just as great as comorbidity within individuals, whereas comorbidity within DZ twins is much lower. The same result was found when probands were selected for V delay and when probands were selected for P delay. This finding implies that V delay and P delay are largely the same disorder from a genetic perspective. That is, if specific genes were found that are associated with V delay, the same genes are likely to be associated with P delay, and vice versa. V delay and P delay are obviously not the same disorder phenotypically because only 22% of children in the lowest 5% of V are also in the lowest 5% of P and vice versa. The differences between V delay and P delay are largely environmental in origin but the overlap is largely genetic.

Similar results were found for V→P when the dimensional perspective of bivariate DF analysis was applied to V probands in the lowest 5% of the V distribution and their co-twin's scores on the P dimension. That is, for V probands the mean decrement of co-twins on P was entirely due to genetic factors. Moreover, bivariate DF analysis permits the calculation of the genetic correlation which was 1.0 for V→P, again indicating that the genes that affect V delay are the same genes responsible for lowered P scores for these V delay probands. In other words, for V probands, genetic factors govern comorbidity with P whether P is assessed dichotomously or dimensionally.

The present results for P→V indicate that bivariate genetic results are not necessarily symmetrical. Bivariate DF analysis applied to P probands and their co-twin's scores on the V dimension suggests that genetic factors account for half of the phenotypic comorbidity with V assessed dimensionally. The genetic correlation for P→V was only 0.36, i.e. genes responsible for P delay are largely different from genes responsible for lowered V scores for these P delayed probands. Thus, bivariate DF analyses suggest that the genetic correlation for V→P is very high whereas for P→V the genetic correlation is modest.

The bivariate DF result for P→V showing modest genetic overlap between P delay and V assessed dimensionally is surprising because genetic factors appear to explain completely the overlap between P delay and V assessed dichotomously as V delay. This finding makes the point that results at the extreme of a distribution need not correspond to those for the entire range of the distribution. That is, twin cross-concordance analysis for P delay probands only counts concordance when co-twin V scores are in the lowest 5% of V. In contrast, bivariate DF analysis for P delay probands considers the average decrement of V performance in all co-twins. These are different questions and the present results serve as a reminder that the answers can differ. This finding suggests that in P delay probands the genes that affect P delay are the same genes that affect V delay defined dichotomously as the lowest 5% of the V distribution, whereas different genes are involved when the entire distribution of V scores is considered. For example, a particular gene that contributes to severe P delay might be involved in general developmental delay and thus also lead to severe V delay. However, most genes responsible for severe P delay might not lead to decrements in V.

Despite this bivariate DF result for P→V, the major message from these findings is that comorbidity between V delay and P delay is largely genetic in origin whereas the differences are largely environmental. Even the bivariate DF result for P→V suggests that half of the overlap between P delay and V is genetic in origin. These results showing that the same genes but different environments are responsible for V delay and P delay are similar to findings for anxiety and depression in children (Eley & Stevenson, in press) and adults (Kendler *et al.*, 1992). These are good examples of the potential offered by genetic analysis to create diagnostic systems that are based on causes rather than phenotypic descriptions of symptoms.

The other analyses of the current data set (Eley *et al.*, 1999) that have focused on V delay and P delay as components of general cognitive delay indicated that V delay showed greater genetic influence than P delay and that different heritabilities imply that V delay and P delay are best considered separately in infancy rather than combined as one composite measure of general cognitive delay. However, it was also noted that although heritabilities of V delay and P delay differ, it is possible that they are both affected by the same genetic factors. This present genetic analysis of their comorbidity has suggested that this is indeed the case, that genes influencing V delay also influence P delay and

vice versa, although the magnitude of genetic influence for V delay is quite different from that for P delay.

These results are based on a large sample of twins using parental report measures of verbal and non-verbal delay. Potential sources of bias in any quantitative genetic analysis of this kind include bias arising from low participation rate, which increases the probability of the final sample not being representative of the population, as well as parental rater bias. We can be confident that the present results are not artefacts of such phenomena for several reasons. First, one could not realistically expect participation rates much greater than 50% for such samples. More importantly, less than perfect participation rates only increase the probability of the final sample not being representative – the TEDS sample has been shown to be representative, compared with census data. Additionally, the very essence of DF group analysis involves selecting a phenotypically extreme sample and examining whether or not this group is representative of the unselected sample. The 5% group has been selected to represent extremes of the variables pertinent to verbal and cognitive development; the unselected sample has not, and can be regarded as a control group. Another potential bias is parental rater bias. Specifically for the present bivariate analyses, rater bias is unlikely to affect the analysis of comorbidity. It is easy to conceive of biases which, within individuals, make V and P rate as more similar; also, biases might make MZ twins rate as more similar than DZ twins for V and P. To impact upon comorbidity, however, any bias would have to make parents rate two traits *within the same child* as more similar for individuals who are MZ twins than for individuals who are DZ twins – this is a less plausible form of bias.

These results have implications for general research strategies. For example, it is generally supposed that 'purer' disorders will yield cleaner results. However, this reductionistic assumption could seriously mislead research such as molecular genetic research. That is, these results suggest that selecting children with 'pure' V delay (i.e. without P delay) would select for cases largely environmental in origin, thus foreclosing the possibility of identifying genes.

Finally, data collected on the 1995 birth cohort of the TEDS sampling frame will soon be available. This will potentially allow a replication of the present findings for the aetiology of comorbidity of verbal and non-verbal delay in 2-year-olds. In addition, as these twins will also be assessed at 3 and 4 years of age, it will be possible to explore the ways in which these aetiological factors change throughout development, and to examine the relationship between early language delay and persistent language difficulties.

Appendix

Standard errors for c^2g can be calculated using the following formula:

$$SE_{c^2g} = \left[\frac{4(\sigma_{DZ})^2}{N_{DZ}} + \frac{\sigma_{MZ}^2}{N_{MZ}} \right]^{1/2}$$

where σ_{MZ} is the standard deviation of MZ co-twins' scores, σ_{DZ} is the standard deviation of DZ co-twins' scores, N_{MZ} is the number of MZ pairs and N_{DZ} is the number of DZ pairs.

Correction for standard errors necessary due to concordant proband pair double-entry is calculated with the following formula:

$$SE_{corrected} = SE_{obtained} \times \left(\frac{N_D - K - 1}{N_S - K - 1} \right)^{1/2}$$

where N_D is the number of double-entry probands, N_S is the number of single-entry probands and K is the number of coefficients in the regression equation.

Acknowledgements

We thank the parents of the twins in the Twins Early Development Study (TEDS) for making the study possible. TEDS is supported by a programme grant from the UK Medical Research Council.

References

Bayley, N. (1993). *Bayley Scales of Infant Development*. New York: The Psychological Corporation.

Bishop, D. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove, UK: Psychology Press.

Bishop, D., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Developmental Medicine and Child Neurology*, **37**, 56–71.

Brody, N. (1992). *Intelligence*. New York: Academic Press.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **32**, 1063–1080.

Carroll, J. (1993). *Human cognitive abilities*. New York: Cambridge University Press.

Dale, P., Simonoff, E., Bishop, D., Eley, T., Oliver, B., Price, T., Purcell, S., Stevenson, J., & Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience*, **1**, 324–328.

Dale, P., Purcell, S., & Plomin, R. (in preparation). A genetic analysis of precocity in language development in 2-year-olds.

Deater-Deckard, K., Reiss, D., Hetherington, E., & Plomin, R. (1997). Dimensions and disorders of adolescent adjustment: a quantitative genetic analysis of unselected samples and selected extremes. *Journal of Child Psychology and Psychiatry*, **38**, 515–525.

DeFries, J., & Fulker, D.W. (1985). Multiple regression analysis of twin data. *Behavior Genetics*, **15**, 467–473.

DeFries, J., & Fulker, D.W. (1988). Multiple regression analysis of twin data: etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae: Twin Research*, **37**, 205–216.

DeFries, J., Olson, R., Pennington, R., & Smith, S. (1991). Colorado Reading Project: An update. In D. Duane & D. Gray (Eds), *The reading brain: The biological basis of dyslexia* (pp. 53–87). Parkton, MD: York Press.

Dinnebeil, L., & Rule, S. (1994). Congruence between parents' and professionals' judgments about the development of young children with disabilities: a review of the literature. *Topics in Early Childhood Special Education*, **14**, 1–25.

Eaves, L., Martin, N., & Eysenck, S. (1977). An application of the analysis of covariance structures to the psychogenetical study of impulsiveness. *British Journal of Mathematical and Statistical Psychology*, **30**, 185–197.

Eley, T.C., & Stevenson, J. (1999). Exploring the covariation between anxiety and depression symptoms: a genetic analysis of the effects of age and sex. *Journal of Child Psychology and Psychiatry*, **40**, 1273–1284.

Eley, T., & Stevenson, J. (in press). Using genetic analyses to clarify the distinction between depressive and anxious symptoms in children and adolescents. *Journal of Abnormal Child Psychology*.

Eley, T., Bishop, D., Dale, P., Oliver, B., Petrill, S., Price, T., Purcell, S., Saudino, K., Simonoff, E., Stevenson, J., & Plomin, R. (1999). Genetic and environmental origins of verbal and performance components of cognitive delay in two-year-olds. *Developmental Psychology*, **35**, 1122–1131.

Falconer, D. (1965). The inheritance of liability to certain diseases estimated from the incidence among relatives. *Annals of Human Genetics*, **29**, 51–76.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59**, v–173.

Fenson, L., Pethick, S., Renda, C., Cox, J., Dale, P., & Reznick, J. (1997). Technical manual and user's guide for the MacArthur communicative development inventories: short form versions. Unpublished manuscript, San Diego State University.

Goldsmith, H. (1991). A zygosity questionnaire for young twins: a research note. *Behavior Genetics*, **21**, 257–269.

Kendler, K., Neale, M., Kessler, R., Heath, A., & Eaves, L. (1992). Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Archives of General Psychiatry*, **49**, 716–722.

Knopik, V., Alarcón, M., & DeFries, J. (1997). Comorbidity of mathematics and reading deficits: evidence for a genetic etiology. *Behavior Genetics*, **27**, 447–453.

Leonard, L. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics*, Vol. 1: *Disorders of first-language development* (pp. 1–39). New York: Cambridge University Press.

Lewis, B., & Thompson, L. (1992). A study of developmental speech and language disorders in twins. *Journal of Speech and Hearing Research*, **35**, 1086–1094.

McGue, M., & Bouchard, T. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics*, **14**, 325–343.

Nichols, P. (1984). Familial mental retardation. *Behavior Genetics*, **14**, 161–170.

Petrill, S., Saudino, K., Cherny, S., Emde, R., Hewitt, J., Fulker, D., & Plomin, R. (1997). Exploring the genetic etiology of low general cognitive ability from 14 to 36 months. *Developmental Psychology*, **33**, 544–548.

Plomin, R., & DeFries, J. (1979). Multivariate behavioral genetic analysis of twin data on scholastic abilities. *Behavior Genetics*, **9**, 505–517.

Plomin, R., DeFries, J., McClearn, G., & Rutter, M. (1997). *Behavioral genetics* (3rd edn), New York: Freeman.

Price, T.S., Eley, T.C., Dale, P.S., Stevenson, J., & Plomin, R. (2000). Genetic and environmental covariation between verbal and non-verbal cognitive development in infancy. *Child Development*, **71**, 948–959.

Rispens, J., & van Yperen, T. (1997). How specific are 'specific developmental disorders'? The relevance of the concept of specific developmental disorders for the classification of childhood developmental disorders. *Journal of Child Psychology and Psychiatry*, **38**, 351–363.

Saudino, K., Dale, P., Oliver, B., Petrill, S., Richardson, V., Rutter, M., Somonoff, E., Stevenson, J., & Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of two-year-olds. *British Journal of Developmental Psychology*, **16**, 349–363.

Smith, C. (1974). Concordance in twins: methods and interpretation. *American Journal of Human Genetics*, **26**, 454–466.

Stevenson, J., Pennington, B., Gilger, J., DeFries, J., & Gillis, J. (1993). Hyperactivity and spelling disability: testing for shared genetic aetiology. *Journal of Child Psychology and Psychiatry*, **34**, 1137–1152.

Tomblin, J., & Buckwalter, P. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language and Hearing Research*, **41**, 188–199.

Wilson, R., & Matheny, A. (1976). Retardation and twin concordance in infant mental development: a reassessment. *Behavior Genetics*, **6**, 353–356.

Received: 10 November 1999
Accepted: 22 May 2000