

**Do children with Williams syndrome really have good vocabulary knowledge?
Methods for comparing cognitive and linguistic abilities in developmental disorders**

Jon Brock, Macquarie Centre for Cognitive Sciences, Macquarie University

Christopher Jarrold, University of Bristol

Emily K. Farran, University of Reading

Glynis Laws, University of Bristol

Deborah M. Riby, University of Stirling

The comparison of cognitive and linguistic skills in individuals with developmental disorders is fraught with methodological and psychometric difficulties. In this paper, we illustrate some of these issues by comparing the receptive vocabulary knowledge and nonverbal reasoning abilities of 41 children with Williams syndrome, a genetic disorder in which language abilities are often claimed to be relatively strong. Data from this group were compared with data from typically developing children, children with Down syndrome, and children with non-specific learning difficulties using a number of approaches including comparison of age-equivalent scores, matching, analysis of covariance, and regression-based standardization. Across these analyses children with Williams syndrome consistently demonstrated relatively good receptive vocabulary knowledge, although this effect appeared strongest in the oldest children.

Studies of developmental disorders have played an increasingly important role in our understanding of language development. Of particular interest are disorders in which language appears to develop out of step with other cognitive faculties, either being specifically impaired or more advanced than predicted by global or nonverbal intellectual abilities. This 'de-coupling' of language and cognition has been interpreted by some as evidence for the independence of linguistic and cognitive development (e.g., Pinker, 1999) and by others in terms of developmental constraints that affect one domain more than another (e.g., Karmiloff-Smith, 1998). In many cases, however, strengths and weaknesses are relative rather than absolute and the process of quantifying impairments is less than straightforward. Here, we illustrate the psychometric issues involved in comparing performance across tasks using the example of receptive vocabulary knowledge in Williams syndrome, considering the advantages and limitations of various methodological and statistical approaches.

Williams syndrome is a genetic disorder caused by a deletion of around 25 genes in the 7q11.23 region of chromosome 7 (Ewart et al., 1993). It is associated with a number of characteristic physical and medical features including an 'elfin' facial profile and a high incidence of cardiac anomalies, specifically supra-valvular aortic stenosis (see e.g., Morris, 2006). Most individuals with Williams syndrome are categorised as having mild to moderate intellectual disability but the cognitive profile is uneven with notable difficulties faced on tasks that involve visuo-spatial construction (see e.g., Farran & Jarrold, 2003). Language abilities, by contrast, are considered by most researchers to be a relative strength (e.g., Bellugi et al., 2000; Mervis et al., 2003), even if claims of 'intact' or 'normally developing' language have largely been discredited (see Karmiloff-Smith et al., 2003).

In fact, the evidence for even relative language strengths is less than clear-cut (Bates, 2004; Brock, 2007). Despite intense interest in the syntactical, morphological, and phonological competence of people with Williams syndrome, in the vast majority of studies (including those cited as evidence for linguistic strengths), their performance on linguistic tasks is no better than might be predicted on the basis of their overall level of functioning (e.g., Bellugi et al., 1988; Clahsen & Almazan, 1998; Karmiloff-Smith et al., 1997; Vicari et al., 2004; Zukowski, 2004). Individuals with Williams syndrome do appear to have relatively good grammatical and phonological skills when compared to those with Down syndrome (Bellugi et al., 1990; Jarrold et al., 1999; Vicari et al., 2002, 2004), but it is important to bear in mind that Down syndrome is associated with specific weaknesses in these areas (e.g., Laws & Bishop, 2004), so comparisons may tend to flatter those with Williams syndrome (see Temple et al., 2002).

The best evidence for linguistic strengths in Williams syndrome comes from studies using standardized tests of receptive vocabulary such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981) in which participants are presented with a series of displays, each containing four pictures, and are required to select the picture that corresponds to a single spoken. Adolescents and young adults with Williams syndrome have consistently been found to perform relatively well on such tasks, at least when compared to their overall or nonverbal abilities (e.g., Bellugi et al. 1990;

Grant et al 1997; Jarrold et al., 1999; Robinson et al., 2003; Vicari et al., 2004). Having said that, it is important to note that evidence for a vocabulary advantage comes primarily from studies comparing age-equivalent scores across tasks and from comparisons with matched groups of younger typically developing children. As discussed below, there are numerous psychometric issues pertaining to both approaches. This concern is heightened by the fact that individuals with Williams syndrome appear to perform less well on other measures of vocabulary knowledge in which they are required to provide definitions for words or provide names for pictures (see e.g., Clahsen et al., 2004; Thomas et al., 2006).

Assuming that the receptive vocabulary advantage is genuine rather than artefactual, a number of further issues remain. It is unclear, for example, whether good receptive vocabulary is specific to Williams syndrome or is instead a general consequence of developmental delay. Performance on receptive vocabulary tests is considered to be a measure of 'crystallized intelligence' or accumulated knowledge, so comparison with data from younger typically developing children who have considerably less life experience (even if they have comparable intellectual abilities) may unduly favour those with Williams syndrome (cf. Baddeley, 1993). Indeed, comparisons with individuals with Down syndrome matched on chronological age as well as measures of overall intellectual functioning have sometimes failed to find the expected Williams syndrome advantage (Klein & Mervis, 1999; Ypsilanti et al., 2005; but see Bellugi et al., 1990; Vicari et al., 2004). In addition, parental reports provide no indication that very young children with Williams syndrome have particularly extensive vocabularies (Laing et al., 2002; Singer-Harris et al., 1997; Vicari et al., 2002), suggesting the possibility that the vocabulary advantage in Williams syndrome emerges with age or developmental level (Jarrold et al., 1998).

In this paper, we attempted to address these issues by pooling and re-analysing data collected from a number of studies in which children with Williams syndrome had been tested on the British Picture Vocabulary Scale (BPVS-II; Dunn et al., 1997), the British equivalent of the Peabody Picture Vocabulary Test. Participants were also tested on the Ravens Coloured Progressive Matrices (RCPM; Raven, 1993) – a measure of nonverbal reasoning ability and fluid intelligence in which participants are required to complete geometric patterns by selecting the correct piece. Importantly, individuals with Williams syndrome have been found to perform considerably better on the RCPM than on measures specifically testing visuo-spatial skills (Farran et al., 2001, 2003), suggesting that it provides a fairer test of nonverbal ability in Williams syndrome.

As in previous studies, children with Williams syndrome were compared with typically developing children and children with Down syndrome. In addition, data were available from a large group of children with non-specific learning disabilities. The deployment of such a comparison group is often criticized on the grounds that it will likely contain children whose learning disabilities arise for a wide range of different reasons, including children who have specific genetic disorders that have not been identified (e.g., Burack et al., 2004). While this is almost certainly

true, a large group such as this enables us to explore the *combined* effect of chronological age and nonverbal ability on vocabulary knowledge, both in terms of the average level of performance and its variability. In this sense the heterogeneity within the LD group can be seen as an advantage rather than a concern.

The data were subject to a number of analyses including comparison of age-equivalent scores, comparison with matched control groups, analysis of covariance and regression-based standardization. The aim was not only to address the issues specific to vocabulary knowledge in Williams syndrome but also to demonstrate some of the psychometric issues that should be considered when analysing this kind of data and to highlight the advantages and limitations of various statistical and methodological approaches.

Participants

Different participants were included in different analyses so we only report brief descriptions of the children from whom data were available. The Williams syndrome group were 41 6- to 17-year-old children with clinically confirmed diagnoses of Williams syndrome, who had been recruited via the Williams Syndrome Foundation, UK - a support group for families of individuals with the syndrome. The Down syndrome group comprised 24 children with clinically confirmed diagnoses of Trisomy 21 (age range 9 to 17 years), who were recruited via personal contacts and schools for children with special educational needs. The typically developing group were 133 4- to 10-year-old children recruited from three mainstream primary schools as part of other studies. Finally, 122 5- to 17-year-old children with learning difficulties were recruited from a total of 6 schools for children with special educational needs. None of the children in this group had Williams syndrome, Down syndrome, or any other specific diagnosis at the time of testing.

Comparison of age-equivalent and standard scores

Most standardized assessments provide norms so that an individual's raw scores can be converted into age-equivalent scores, which show the age at which typically developing children are expected to achieve a particular raw score. In studies of Williams syndrome, age-equivalent scores for receptive vocabulary are significantly higher than for overall mental age, nonverbal reasoning ability, or visuo-spatial construction skills (Bellugi et al. 1990; Grant et al 1997; Jarrold et al., 1999; Mervis et al., 1999). Analysis of our own data is concordant with these findings. Seventeen of the children in the sample had raw scores below 15 on RCPM and it was not, therefore, possible to calculate an age-equivalent score given the norms provided. For the remaining 24 children with Williams syndrome, age-equivalent scores were considerably higher for the BPVS-II than for the RCPM, $t(23) = 3.56, p = .002$ (see Table 1 for descriptive statistics).

Table 1: Comparison of age-equivalent scores for the BPVS-II and RCPM among individuals with Williams syndrome (WS) and typically developing (TD) children

	WS N = 24	TD N = 28
Age (years; months)	12;0 (3;1)	8;6 (0;4)
RCPM (years; months)	7;5 (1;5)	8;9 (1;3)
BPVS-II (years; months)	8;3 (2;2)	8;1 (1;3)

Such findings appear consistent with the view that receptive vocabulary is a relative strength in Williams syndrome. However, the comparison of age-equivalent scores across different measures assumes comparability of standardization when, in fact, tests are often standardized on different groups of typically developing children from different countries or social backgrounds at different points in recent history. In addition, age-equivalent scores are a function of the tightness of the relationship between age and performance – if age accounts for little variation in performance then a relatively mild impairment can result in a large mental age 'delay' because the individual will be performing at the level expected of much younger children. Given that the relationship between performance and age is rarely the same across tasks, this distortion effect is unlikely to be the same either (see e.g., Bishop, 1997; Mervis & Klein-Tasman, 2004).

These concerns are illustrated by comparison of age-equivalent scores for all the typically developing children in our large sample who were aged between 8;0 years and 9;0 years (see Table 1). These children performed well within the limits of the standardization for the BPVS-II and the RCPM and yet we still find significant differences in their age-equivalent scores across the tasks, $t(27) = 2.31, p = .028$. In this case, the effect is in the opposite direction to that found in our Williams syndrome sample, suggesting that the vocabulary advantage observed in Williams syndrome is genuine (and perhaps underestimated). Nevertheless, the fact that typically developing children can show significant discrepancies between their age-equivalent scores serves to demonstrate the need for caution when interpreting similar differences among children with developmental disorders.

Comparison with matched 'controls'

An alternative approach to comparing scores across different measures is to employ a control group that are matched on one measure and then determine whether groups differ significantly on the other. For example, if groups are matched on nonverbal ability then one might ordinarily expect them also to have equivalent vocabulary knowledge. An important issue concerns the closeness with which groups are matched. Often, groups are described as being matched if there is no significant difference in their scores on the matching measure. The problem here is that conventional tests of group differences such as *t*-tests are designed to quantify the risk of falsely concluding that group means are different when in fact they are drawn from the same distribution. For the purposes of matching, we want to be confident that groups are *not* different. Mervis and Klein-Tasman (2004) have argued that groups should only be considered 'matched' if the *p*-value for group differences is greater than .60. It is worth considering, however, that the *p*-value in a *t*-test is contingent on the size of the groups involved and the heterogeneity in the sample – small groups covering a wide range of ages or abilities are much more likely to achieve the $p > .60$ criterion than larger more constrained groups.

A number of studies have compared receptive vocabulary knowledge in individuals with Williams syndrome and younger typically developing controls matched on measures of overall mental age, visuo-spatial ability, or nonverbal reasoning ability. The majority of researchers have reported superior vocabulary in children with Williams syndrome (Clahsen et al., 2004; Jarrold et al., 1999; Robinson et al., 2003; Vicari et al., 2004). A notable exception is the study by Volterra et al. (1996), who found that their children with Williams syndrome had receptive vocabulary scores that were no better than younger typically developing control children. Critically, however, the two groups were not actually matched on any measure of intelligence; instead, the nonverbal mental ages of the children with Williams syndrome were reported as being similar to the chronological ages of the typically developing children. Although a common practice in research into developmental disorders, as we have seen already in our own sample, chronological age and measured mental age are not necessarily equated in typically developing children, so the groups may not in fact have been matched at all. Our own data are consistent with the consensus in the existing literature. Table 2 shows the performance of 36 of the children with Williams syndrome for whom we were able to find individual matches on RCPM scores (to within 1 point) from the large sample of typically developing children. Obviously, there were no group differences in RCPM score, but the children with Williams syndrome performed considerably better on the BPVS-II.

Table 2: Comparison of BPVS-II scores in children with Williams syndrome (WS) and typically developing (TD) children matched on RCPM scores

	WS N = 36	TD N = 36	<i>t</i> (70)	<i>p</i>
Age (mo)	150 (35)	77 (11)	12.08	< .001
RCPM score	17.6 (5.0)	17.6 (4.9)	0.02	.981
BPVS-II score	85.2 (16.4)	70.0 (12.7)	4.39	< .001

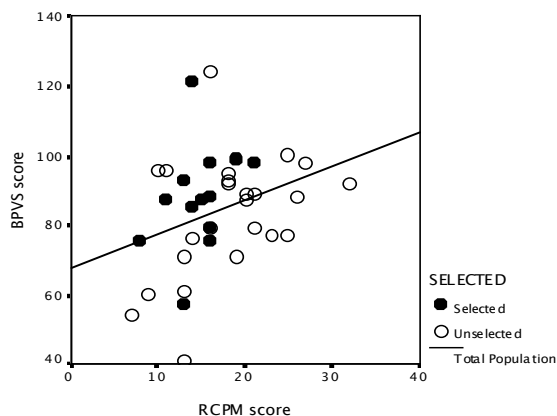
Table 3: Comparison of BPVS-II scores in children with Williams syndrome (WS) and children with Down syndrome (DS) matched on RCPM scores and chronological age

	WS N = 13	DS N = 13	t(24)	p
Age (mo)	159 (32)	157 (41)	0.10	.924
RCPM score	14.8 (3.3)	14.5 (3.2)	0.18	.857
BPVS-II score	87.9 (15.3)	54.8 (17.1)	5.20	< .001

As noted earlier, studies directly comparing vocabulary knowledge in Williams syndrome and Down syndrome have provided mixed results (Bellugi et al., 1990; Klein & Mervis, 1999; Vicari et al., 2004; Ypsilanti et al., 2005), leading to the suggestion that relatively good vocabulary knowledge may not be specific to Williams syndrome (Klein & Mervis, 1999). We were able to select 13 pairs of individuals (one from each group) who were matched to within one point on the RCPM and to within 12 months in age. Overall, groups were non-significantly different in terms of age and RCPM score but the Williams syndrome group performed considerably better on the BPVS-II (see Table 3).

Although this finding appears conclusive, the *strategic* selection of individuals to form matched groups raises a concern insofar as those selected may not necessarily be representative of the broader groups from which they are drawn. This can be seen in Figure 1, which shows the BPVS-II scores of all of the children with Williams syndrome who were in the same age range as the Down syndrome sample, plotted as a function of their RCPM scores. The majority of individuals who were selected for the comparison with Down syndrome are to be found above the regression line, indicating that we have a biased sample containing predominantly individuals with Williams syndrome who have relatively high BPVS-II scores given their RCPM scores. This sample bias occurs because we have preferentially selected individuals with Williams syndrome who have low RCPM scores (given the fact that they have Williams syndrome) in order to match them to the generally lower-functioning individuals with Down syndrome. BPVS-II and RCPM scores are not perfectly correlated, so individuals with low RCPM scores do not necessarily have low BPVS-II scores (the mean BPVS-II score of the selected subgroup 'regresses to the mean' of the overall group; cf. Lord, 1967). As a consequence, the discrepancy between BPVS-II and RCPM is, on average, greater in the subgroup than in the overall group, increasing the likelihood that we will get significant group differences. This problem affects all studies in which participants are selected on the basis of their performance on a matching measure (age is normally measured perfectly so is not affected by regression to the mean), but is particularly apparent here because the initial groups differed considerably.

Figure 1: BPVS-II scores in the Williams syndrome group as a function of RCPM score. Filled circles represent individuals selected for comparison with matched individuals with Down syndrome.



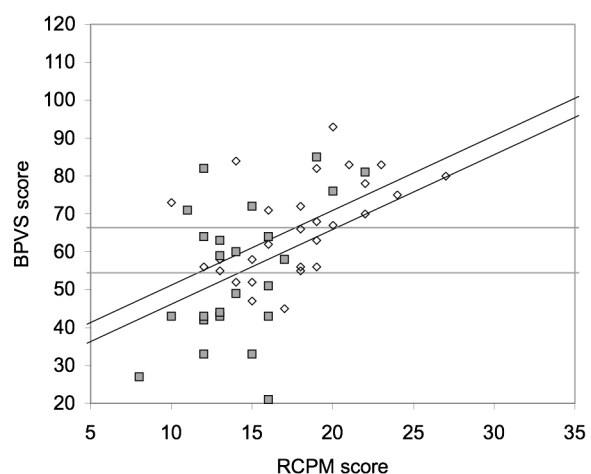
Analysis of covariance

Analysis of covariance (ANCOVA) is often treated as a solution to many of the problems associated with matching designs outlined above. Rather than excluding participants strategically to match the groups, researchers can in principle include data from all participants, thus avoiding problems of biased sampling. Moreover, there is no need to set arbitrary criteria for the acceptable degree of matching because any group differences on the matching measure are controlled for statistically. Adopting this approach, we can compare the vocabulary scores of all the children with Williams syndrome with all of our typically developing sample, treating RCPM score as a covariate. In this analysis, the effect of the covariate is highly significant, $F(1,163) = 58.75$, $p < .001$, $\eta_p^2 = .265$, but the group effect remains, $F(1,163) = 39.34$, $p < .001$, $\eta_p^2 = .194$.

Unfortunately, the application of ANCOVA relies on a number of assumptions and, if these are violated, there is potential for misinterpretation of the data (Miller & Chapman, 1991). ANCOVA is best thought of as a model-fitting exercise (e.g., Elashoff, 1969). In the ANCOVA above, we were effectively comparing two models: in the first, BPVS-II score was predicted solely by RCPM score; in the second, BPVS-II score was a joint function of RCPM score and group membership. The second, more complex model will always provide a better fit to the data, but the question is whether or not the improvement is significant.

There are two critical points to note. First, a standard ANCOVA assumes that the same relationship holds between the dependent variable and the covariate (the assumption of 'homogeneity of regression slopes'). In our example, this means that BPVS-II score is assumed to increase by the same amount in both groups for a given increase in RCPM score. It is possible to test whether this assumption holds or not by conducting an ANCOVA in which the two groups are allowed to have different regression slopes and then determining whether or not this improves the fit of the model further (Wright, 1997; see Brock & Jarrold, 2004 for a worked example). If the regression slopes are not the same, the main effect of group may be difficult to interpret: instead of looking for absolute differences between the groups, it may be more appropriate to consider the extent of group differences for different values of the covariate (see Karmiloff-Smith et al., 2004; Thomas et al., 2006).

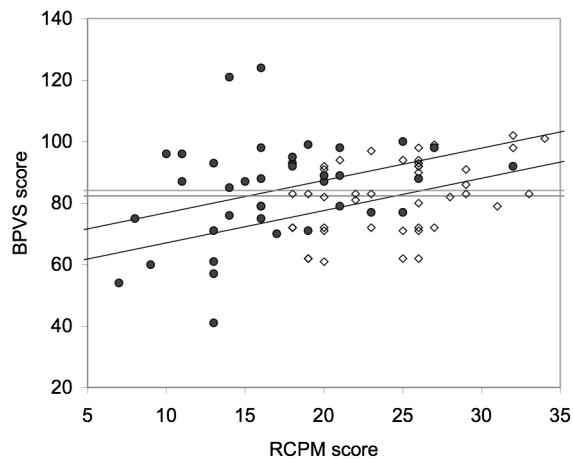
Figure 2: BPVS-II scores of children with Down syndrome (squares) and typically developing children (diamonds), plotted as a function of RCPM scores. Grey lines represent the fit of an ANOVA to the data. Black lines show the fit of an ANCOVA. In both instances, the upper line shows the model fit for typically developing children



The second key assumption of ANCOVA is that the covariate and experimental group are independent. In other words, the groups should not differ on the covariate. If this constraint is violated, then genuine group differences may disappear and, in some circumstances, spurious group differences may appear. To illustrate the first scenario, we compare data from typically developing children and children with Down syndrome (see Figure 2). Because the mean RCPM score is considerably higher in our

typically developing sample than in our Down syndrome sample, we include only typically developing children with chronological ages of 70 months or below. Even so, the typically developing children still have higher RCPM scores on average. A simple ANOVA confirms the impression that children with Down syndrome have poorer vocabulary knowledge than typically developing children, $F(1,51) = 9.68, p = .006$. However, if we add RCPM as a covariate, $F(1,50) = 13.8, p = .001$, the group difference disappears, $F(1,50) = 1.4, p = .235$. In other words, the ANCOVA shows that a model with group and RCPM score as factors is not significantly better than a simpler model in which RCPM score is the only factor. The ANCOVA indicates that it is possible that group differences in vocabulary are simply a function of lower overall intellectual capacities. An alternative interpretation, however, is that, because individuals with Down syndrome have both poor vocabulary knowledge and poor nonverbal reasoning abilities, in the overall sample (including typically developing children), BPVS-II score is strongly associated with RCPM score. Consequently, much of the individual variation that would otherwise have been attributed to group membership is instead attributed to RCPM scores.

Figure 3: BPVS-II scores of children with Williams syndrome (circles) and typically developing children (diamonds) plotted as a function of RCPM scores. Grey lines represent the fit of an ANOVA to the data. Black lines show the fit of an ANCOVA. In both instances, the upper line shows the model fit for children with Williams syndrome



Given that ANCOVA can remove significant group differences, it is often treated as being a conservative analysis; if significant effects of group are found after the covariate is added then the covariate cannot fully explain the effect. However, this intuition is false because, in some situations, ANCOVA can lead to large group differences that are negligible under any other analysis (see Evans & Anastasio, 1968). This is possible because the relationship between the dependent variable and the covariate is not necessarily the same in the first model (covariate only) and the second (covariate and group included). To illustrate, we compare the data from our Williams syndrome group with that from a subgroup of typically developing children (see Figure 3). For the sake of argument, imagine that we had only tested children aged 90 months or above. With this constraint, the children with Williams syndrome no longer perform significantly better on the BPVS-II, $F(1,78) = 0.3, p = .590$, but perform much worse on the RCPM. In an attempt to equate groups for nonverbal ability, we add RCPM score as a covariate. The effect of the covariate is significant, $F(1,77) = 11.6, p = .001$, and suddenly so too is the effect of group membership, $F(1,77) = 6.5, p = .013$. Although this result is consistent with our other analyses based on the full data set, if the data in Figure 3 were all we had to go by, we would not be justified in claiming that the children with Williams syndrome had superior vocabulary knowledge.

The fact that ANCOVA can qualitatively alter our conclusions in both directions means that it should be applied and interpreted with caution and should not be seen as the panacea for all matching ills. Certainly, ANCOVA ought not to be used to equate groups when there are large differences on a matching variable. Indeed, strictly

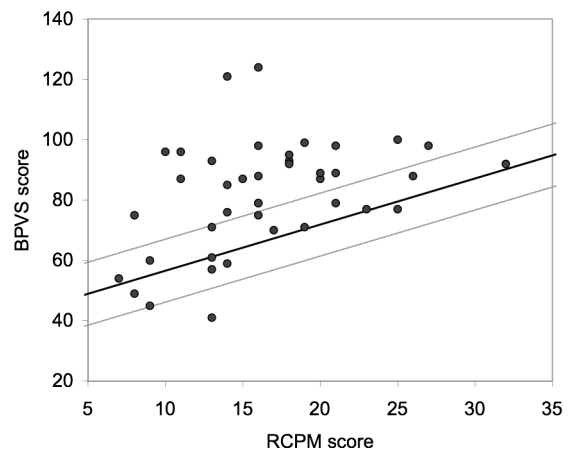
speaking, it should only be used when group assignment is random, thus allowing any group differences in the covariate to be treated as uncorrelated noise (Evans & Anastasio, 1968). This requirement is, of course, impractical in the majority of studies of developmental disorders and a case can be made for including ANCOVA in addition to more conventional group analyses in situations where groups are relatively well-matched on the covariate in the first place. ANCOVA may also be a useful tool for exploring data to evaluate possible models for describing the data (see Miller & Chapman, 1991). Any conclusions must, however, be tempered by a recognition of the limitations of ANCOVA and the assumptions that may have been violated.

Standardization using linear regression

An alternative to ANCOVA is to conduct a regression analysis to standardize the data from a comparison group (e.g., typically developing children) and then determine whether individuals in the clinical group perform better or worse than predicted by, for example, their age or performance on another measure. Although ANCOVA can be treated as a form of regression analysis, the two approaches differ in the sense that the regression conducted here is only aimed at modelling the data from the comparison group, whereas ANCOVA attempts to model data from participants in both groups simultaneously. Consequently, the standardization approach side-steps concerns with group differences in the predictor (covariate).

As a first example, we conduct simple linear regression on data from the typically developing children to allow us to predict an individual's BPVS-II score on the basis of their RCPM scores. The linear function is plotted in Figure 4, together with the observed data from the children with Williams syndrome. The residuals are calculated by subtracting the expected BPVS-II score for each child with Williams syndrome (based on the typically developing data) from the observed scores and then standardizing them by dividing by the standard error of the regression estimate. If vocabulary developed in line with nonverbal abilities then the mean standardized vocabulary score should be zero. Instead, the mean standardized score for the children with Williams syndrome is 1.46 ($SD = 1.60$), which is significantly above zero on a one-sample t -test, $t(40) = 5.84, p < .001$.

Figure 4: BPVS-II scores of children with Williams syndrome as a function of RCPM scores. Heavy regression line shows fit to data from typically developing children. Fine lines represent one standard error in the regression model.

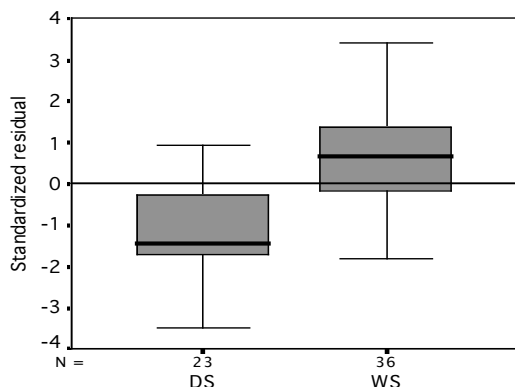


Although this kind of standardization procedure avoids some of the pitfalls associated with other analyses, it should still be used carefully. In particular, it is important to ensure that the relationship between the measures is approximately linear. Non-linearity might arise as a consequence of floor or ceiling effects on test performance, especially if individuals are sampled across a large age and ability range. This problem can often be addressed by transforming the data to create a linear relationship between measures, and then subjecting the data to linear regression as described above (for an example of this approach in a dataset where with floor and ceiling effects led to non-linearity see Jarrold, Baddeley, and Phillips, 2007)¹.

A further constraint is that the data from the clinical group should not go beyond the range of the standardization data. For instance, as discussed earlier, we may be concerned that age differences could lead to differences in vocabulary level, but we could not simply include the ages of the typically developing children in the regression model above and then use this to standardize the scores of the children with Williams syndrome – this would rely on the assumption that the relationship between BPVS-II score and age is the same at all ages. Another possibility would be to collect more data from typically developing children in the same age range as those with Williams syndrome and add this to the model. The concern here would be that age, RCPM score, and BPVS-II score might interact very differently at different age and ability levels.

A further solution is to use data from a large sample of children with other forms of learning disability who are of comparable age and comparable ability to those with Williams syndrome. Because we are not comparing groups directly, it does not matter if they are slightly different in terms of either age or ability, so long as the standardization data covers the full range of ages and RCPM scores found in the Williams syndrome group. Data from our group of children with non-specific learning difficulties were therefore used to develop a standardization for BPVS-II scores based on both age and RCPM score (both entered simultaneously into the regression analysis). Standardized BPVS-II scores for children with Williams syndrome as well as children with Down syndrome were calculated as before and are plotted in Figure 5 (excluding those individuals who were older than 195 months). Again, individuals with Williams syndrome have standardized receptive vocabulary scores that are significantly above zero, $M = 0.702$; $SD = 1.117$; $t(35) = 3.77$, $p = .001$, whereas in Down syndrome, scores are significantly below zero, $M = -1.063$; $SD = 1.104$; $t(22) = -4.62$, $p < .001$.

Figure 5: Standardized residuals for BPVS-II scores (controlling for age and RCPM scores) for children with Down syndrome (DS) and children with Williams syndrome (WS)

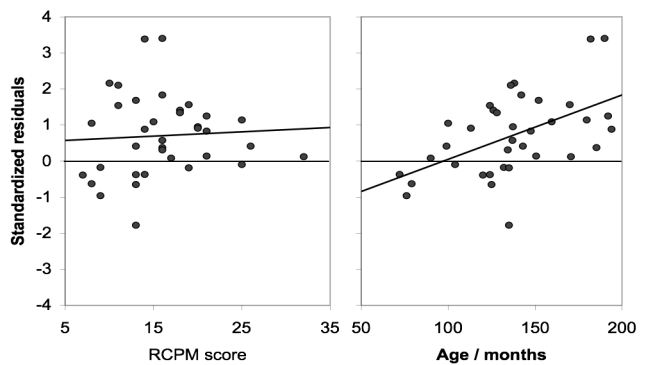


This analysis strongly indicates that receptive vocabulary is relatively good in Williams syndrome and relatively poor in Down syndrome. This is, of course, consistent with our earlier comparison of matched children with the two disorders. Importantly, however, the standardization procedure allows us to compare vocabulary knowledge in Williams and Down syndrome against a common baseline, without having to artificially select unrepresentative members of each group. This approach also gives a better indication of the within-group variation that is found in both disorders. Reporting standard deviations, standard errors, or ranges gives some indication of this, but is necessarily a function of the general heterogeneity within the group in terms of age and overall level of functioning. Figure 5 shows that there remains considerable variation in receptive vocabulary knowledge in both Williams syndrome and Down syndrome, beyond what might be expected on the basis of age and overall ability. Although group means differ, there are many individuals with Williams syndrome who have relatively poor vocabulary and many with Down syndrome who perform relatively well.

The source of this variation can be explored further by plotting the standardized residuals for the children with Williams syndrome against their RCPM score and chronological age. If the relationship between vocabulary knowledge, age and nonverbal ability is the

same in Williams syndrome as it is in the comparison group, then the residuals should be uncorrelated with either age or RCPM score (as these have been accounted for by the standardization procedure). Figure 6 shows that the residuals are uncorrelated with RCPM score, $r(36) = .060$, $p = .729$, but are positively correlated with age, $r(36) = .523$, $p = .001$. This correlation remains significant even if we exclude the two older individuals with Williams syndrome with the largest positive residuals. These findings indicate that the vocabulary advantage is driven primarily by the older children with Williams syndrome.

Figure 6: Standardized residuals for BPVS-II scores of children with Williams syndrome, plotted against RCPM score and chronological age



Discussion

In this paper, we used the example of receptive vocabulary knowledge in Williams syndrome to illustrate some of the methodological and psychometric challenges facing researchers in developmental disorders and to demonstrate the advantages and limitations of various analytical techniques. We highlighted the fact that many of the approaches adopted in current studies in this area are prone to significant biases that can seriously distort the results. For example, the use of age-equivalent scores is problematic for a number of reasons, to the extent that even typically developing children evidenced a discrepancy between their vocabulary knowledge and nonverbal intelligence. For this reason, researchers are strongly advised against simply assuming that children's mental ages are the same as their chronological ages when recruiting matched control groups. Even if groups are well-matched, it is also important to consider how representative the chosen participants are of their respective groups. For instance, in matching children with Williams syndrome to children with Down syndrome on nonverbal ability, we had to select children with Williams syndrome who had relatively low RCPM scores and, due to the phenomenon of regression to the mean, evidenced a relatively large discrepancy between their BPVS-II and RCPM scores when compared to other individuals with Williams syndrome.

We also cautioned against the use of ANCOVA as an alternative to matching, particularly if there are large group differences on the covariate. As we demonstrated with our vocabulary data, in such circumstances, ANCOVA and more conventional ANOVA can lead to radically different conclusions. Problems with ANCOVA arise primarily because the analysis attempts to model the data from the clinical group and the comparison group at the same time, making numerous assumptions along the way. Our preferred method, therefore, is to use regression techniques to model the data from the comparison group and then use this model to predict the performance of individuals in the target clinical group. In our case, we were able to show that children with Williams syndrome had better-than-expected vocabulary knowledge, whereas children with Down syndrome evidenced relative impairments. By conducting correlational analyses on the residuals, we were further able to conclude that the vocabulary advantage was strongest in older children with Williams syndrome.

The analyses conducted in this paper therefore allowed us to address conclusively the empirical issue of whether or not children with Williams syndrome really do have good vocabulary knowledge. An important aspect of our findings is that the receptive vocabulary advantage in Williams syndrome was not

restricted to comparisons with data from typically developing children, but was also observable in comparisons with children with Down syndrome and children with non-specific learning disabilities. Although it is impossible to say whether this pattern of performance is *specific* to Williams syndrome, it appears safe to conclude that relatively good vocabulary knowledge is not merely a consequence of the relatively greater age of children with developmental disorders compared with younger matched control children.

It is also important to note that, although the group differences reported here were statistically reliable, the effect sizes were not large and there was a good deal of individual variation within the groups as well as overlap between the groups. This highlights two important issues: on the one hand, it is critical that large numbers of individuals are tested before one makes generalisations to syndromes as a whole; on the other hand, it is dangerous to assume that all individuals with a particular syndrome will exhibit a similar cognitive profile. Part of this within-group variation was due to the fact that the vocabulary advantage was most apparent in the older children with Williams syndrome, as demonstrated by the positive correlation between the standardized scores and chronological age. This is consistent with the view that the vocabulary advantage increases with age (Jarrod et al., 1998), although it is important to recognise that the current data are not longitudinal.

A further unresolved issue concerns the generalizability of the current findings. Numerous researchers have argued that individuals with Williams syndrome have relatively good language abilities and the current findings are certainly consistent with this view. However, we are only comparing one specific measure of language with one specific measure of nonverbal ability and there is little evidence that language abilities other than receptive vocabulary are any better than might be expected on the basis of overall or nonverbal reasoning ability (Brock, 2007). Indeed, as noted earlier, performance on other measures of vocabulary knowledge such as naming or definitions is often found to be poorer than receptive vocabulary (e.g., Clahsen et al., 2004; Thomas et al., 2006). The reader should, therefore, be cautious about extrapolating our findings beyond the confines of the receptive vocabulary test.

A final obvious question is “why do individuals with Williams syndrome perform so well on receptive vocabulary tests?” It has been widely argued that individuals with Williams syndrome have good vocabulary knowledge specifically as a consequence of relatively preserved phonological short-term memory skills (e.g., Grant et al., 1997; Mervis et al., 1999; Vicari et al., 1996). This suggestion is based on the theory that short-term memory plays a critical role in the learning of new phonological word forms (see Baddeley et al., 1998), but there is, in fact, relatively little evidence for strengths in phonological short-term memory in Williams syndrome (e.g., Brock et al., 2004; Jarrod et al., 2004; see Brock, 2007 for a review). It is also worth pointing out that phonological short-term memory is thought to have the greatest influence during the earliest stages of vocabulary development (e.g., Gathercole, 1995), whereas the vocabulary advantage appears to be greatest in the oldest children. The receptive vocabularies of individuals with Williams syndrome remains, therefore, something of mystery and certainly warrants further consideration.

To conclude, although this paper has allowed us to illustrate a number of issues confronting any researcher who is attempting to compare the relative strengths and weaknesses of individuals with a particular developmental disorder, it is perhaps reassuring to note that all the analyses in this paper led to the same conclusion - that children with Williams syndrome have relatively good receptive vocabulary knowledge. Of course, this may not always be the case, particularly if the effect under consideration is weaker or less reliable than the vocabulary advantage in Williams syndrome appears to be. Given this, researchers should be wary of reading too much into the findings of single studies in isolation and should always consider these psychometric issues when interpreting evidence relating to uneven cognitive profiles.

References

Baddeley, A. (1993). Short-term phonological memory and long-term learning: a single-case study. *European Journal of Cognitive Psychology*, 5, 129-148.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.

Bates, E. A. (2004). Explaining and interpreting deficits in language development across clinical groups: where do we go from here? *Brain and Language*, 88, 248-53.

Bellugi, U., Bihrlle, A., Jernigan, T., Trauner, D., & Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *American Journal of Medical Genetics Supplement*, 6, 115-125.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, 12:S, 7-29.

Bellugi, U., Marks, S., Bihrlle, A., & Sabo, H. (1988). Dissociation between language and cognitive function in Williams Syndrome. In D. Bishop & K. Mogford (Eds.), *Language development in exceptional circumstances*. Edinburgh: Churchill Livingstone.

Bishop, D. V. M. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *Quarterly Journal of Experimental Psychology*, 50A, 899-923.

Brock, J. (2007). Language abilities in Williams syndrome: a critical review. *Development and Psychopathology*, 19, 97-127.

Brock, J., & Jarrod, C. (2004). Language influences on verbal short-term memory in Down syndrome: Item and order recognition. *Journal of Speech Language and Hearing Research*, 47, 1334-1346.

Brock, J., McCormack, T., & Boucher, J. (2005). Probed serial recall in Williams syndrome: Lexical influences on phonological short-term memory. *Journal of Speech Language and Hearing Research*, 48, 360-371.

Burack, J. A., Iarocci, G., Flanagan, T. D., & Bowler, D. M. (2004). On mosaics and melting pots: Conceptual considerations of comparison and matching strategies. *Journal of Autism and Developmental Disorders*, 34, 65-73.

Clahsen, H., & Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition*, 68, 167-198.

Clahsen, H., Ring, M., & Temple, C. M. (2004). Lexical and morphological skills in English-speaking children with Williams syndrome. In S. Bartke & J. Siegmüller (Eds.), *Williams syndrome across languages* (pp. 222-244). Amsterdam: John Benjamins.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test - revised* (2nd ed.). Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale* (2nd ed.). Windsor, U.K.: NFER-Nelson.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 6, 383-401.

Evans, S. H., & Anastasio, E. J. (1968). Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin*, 69, 225-234.

Ewart, A. K., Morris, C. A., Atkinson, D., Jin, W. S., Sternes, K., Spallone, P., et al. (1993). Hemizyosity at the elastin locus in a developmental disorder, Williams syndrome. *Nature Genetics*, 5, 11-16.

Farran, E. K., & Jarrod, C. (2003). Visuo-spatial cognition in Williams syndrome: Reviewing and accounting for strengths and weaknesses in performance. *Developmental Neuropsychology*, 23, 173-200.

Farran, E. K., Jarrod, C., & Gathercole, S. E. (2001). Block Design performance in the Williams syndrome phenotype: A problem with mental imagery? *Journal of Child Psychology and Psychiatry*, 42, 719-728.

Farran, E. K., Jarrod, C., & Gathercole, S. E. (2003). Divided attention, selective attention and drawing: processing preferences in Williams syndrome are dependent on the task administered. *Neuropsychologia*, 41, 676-687.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge—It all depends on the nonwords. *Memory and Cognition*, 23, 83-94.

Grant, J., Karmiloff-Smith, A., Gathercole, S. A., Paterson, S., Howlin, P., Davies, M., & Udwin, O. (1997). Phonological short-term memory and its relationship to language in Williams syndrome. *Cognitive Neuropsychiatry*, 2, 81-99.

- Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1998). Verbal and non-verbal abilities in the Williams syndrome phenotype: evidence for diverging developmental trajectories. *Journal of Child Psychology and Psychiatry*, 39, 511-524.
- Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1999). Genetically dissociated components of working memory: evidence from Down's and Williams syndrome. *Neuropsychologia*, 37, 637-651.
- Jarrold, C., Baddeley, A. D., Hewes, A. K., & Phillips, C. (2001). A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex*, 37, 423-431.
- Jarrold, C., Baddeley, A. D., & Phillips, C. (2007). Long-term memory for verbal and visual information in Down syndrome and Williams syndrome: Performance on the Doors and People Test. *Cortex*, 43, 233-247.
- Jarrold, C., & Brock, J. (2004). To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders*, 34, 81-86.
- Jarrold, C., Cowan, N., Hewes, A. K., & Riby, D. M. (2004). Speech timing and verbal short-term memory: Evidence for contrasting deficits in Down syndrome and Williams syndrome. *Journal of Memory and Language*, 51, 365-380.
- Jarrold, C., Cowan, N., Hewes, A. K., & Riby, D. M. (2004). Speech timing and verbal short-term memory: Evidence for contrasting deficits in Down syndrome and Williams syndrome. *Journal of Memory and Language*, 51, 365-380.
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends In Cognitive Sciences*, 2, 389-398.
- Karmiloff-Smith, A., Brown, J. H., Grice, S., & Paterson, S. (2003). Dethroning the myth: Cognitive dissociations and innate modularity in Williams syndrome. *Developmental Neuropsychology*, 23, 227-242.
- Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). Language and Williams syndrome: How intact is "intact"? *Child Development*, 68, 246-262.
- Karmiloff-Smith, A., & Thomas, M. (2003). What can developmental disorders tell us about the neurocomputational constraints that shape development? The case of Williams syndrome. *Development and Psychopathology*, 15, 969-990.
- Klein, B. P., & Mervis, C. B. (1999). Contrasting patterns of cognitive abilities of 9- and 10-year-olds with Williams syndrome or Down syndrome. *Developmental Neuropsychology*, 16, 177-196.
- Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G., et al. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental Science*, 5, 233-246.
- Laws, G., & Bishop, D. V. M. (2004). A comparison of language abilities in adolescents with Down syndrome and children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 46, 1324-1339.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305.
- Mervis, C. B., & Klein-Tasman, B. P. (2004). Methodological issues in group-matching designs: a levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders*, 34, 7-17.
- Mervis, C. B., Morris, C. A., Bertrand, J., & Robinson, B. F. (1999). Williams syndrome: Findings from an integrated program of research. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders: Contributions to a new framework from the cognitive neurosciences*. Cambridge, MA: MIT Press.
- Mervis, C. B., Robinson, B. F., Rowe, M. L., Becerra, A. M., & Klein-Tasman, B. P. (2003). Language abilities of individuals with Williams syndrome. *International Review of Research in Mental Retardation*, 27, 35-81.
- Morris, C. A. (2006). The dysmorphology, genetics, and natural history of Williams-Beuren syndrome. In C. A. Morris, H. M. Lenhoff, & P. P. Wang (Eds.) *Williams-Beuren syndrome: Research, evaluation, and treatment* (pp. 3-17). Baltimore, MD: John Hopkins University Press.
- Pinker, S. (1999). *Words and rules*. London: Weidenfeld & Nicolson.
- Raven, J. C. (1993). *Coloured progressive matrices*. Oxford, U.K.: Information Press Ltd.
- Robinson, B. F., Mervis, C. B., & Robinson, B. W. (2003). The roles of verbal short-term memory and working memory in the acquisition of grammar by children with Williams syndrome. *Developmental Neuropsychology*, 23, 13-31.
- Singer-Harris, N. G., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. *Developmental Neuropsychology*, 13, 345-370.
- Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2006). Speeded naming, frequency and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes*, 21, 721-759.
- Vicari, S., Bates, E., Caselli, M. C., Pasqualetti, P., Gagliardi, C., Tonucci, F., & Volterra, V. (2004). Neuropsychological profile of Italians with Williams syndrome: An example of a dissociation between language and cognition? *Journal of the International Neuropsychological Society*, 10, 862-876.
- Vicari, S., Carlesimo, G., Brizzolara, D., & Pezzini, G. (1996). Short term memory in children with Williams syndrome: A reduced contribution of lexical-semantic knowledge to word span. *Neuropsychologia*, 34, 919-925.
- Vicari, S., Caselli, M. C., Gagliardi, C., Tonucci, F., & Volterra, V. (2002). Language acquisition in special populations: a comparison between Down and Williams syndromes. *Neuropsychologia*, 40, 2461-2470.
- Volterra, V., Capirci, O., Pezzini, G., Sabbadini, L., & Vicari, S. (1996). Linguistic abilities in Italian children with Williams syndrome. *Cortex*, 32, 663-677.
- Wright, D. B. (1997). *Understanding statistics: An introduction for the social sciences*. London: Sage.
- Ypsilanti, A., Grouios, G., Alevriadou, A., & Tsapkini, K. (2005). Expressive and receptive vocabulary in children with Williams and Down syndromes. *Journal of Intellectual Disability Research*, 49, 353-364.
- Zukowski, A. (2004). Investigating knowledge of complex syntax: Insights from experimental studies of Williams syndrome. In M. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies*. Mahwah, NJ: Erlbaum.

Footnote

1. If an appropriate transformation cannot be found, an alternative approach is to use polynomial regression in which a quadratic term (the square of the predictor) or a cubic term is included in the regression. However, while this may significantly improve the fit of the model relative to the linear model, it is worth bearing in mind that psychometric functions tend to be monotonic and so rarely approximate a quadratic or cubic function. For example, if children reach ceiling by a particular age, a quadratic function would predict that scores would start going down again in older children.

Author note

Correspondence should be addressed to Dr Jon Brock, Macquarie Centre for Cognitive Sciences (MACCS), Macquarie University, Sydney, NSW 2109, Australia. Email: jbrock@maccs.mq.edu.au