Brain morphometry and IQ measurements in preterm children

E. B. Isaacs,[1] C. J. Edmonds,[1] W. K. Chong,[3] A. Lucas,[1] R. Morley[4] and D. G. Gadian[2]

[1]*MRC Childhood Nutrition Research Centre and* [2]*Radiology and Physics Unit, Institute of Child Health, University College London, London WC1N 1EH,* [3]*Department of Radiology, Great Ormond Street Hospital for Children NHS Trust, London, UK and* [4]*University of Melbourne Department of Paediatrics and Murdoch Childrens Research Institute, Royal Children's Hospital, Victoria, Australia*

Correspondence to: Dr E. B. Isaacs, MRC Childhood Nutrition Research Centre, Institute of Child Health, University College London, 30 Guilford Street, London WC1N 1EH, UK
E-mail: E.Isaacs@ich.ucl.ac.uk

Summary

Although IQ is thought to remain relatively stable in the normal population, a decline in IQ has been noted in children born preterm. It is not clear, however, to what extent the inclusion of children with clear neurological damage has influenced these findings. We examined IQ scores obtained in childhood and then again in adolescence from a group of children born at 30 weeks gestation or less who had been classified as neurologically normal at 7.5–8 years. They showed a significant decline in mean IQ scores over time. MRI scans obtained from a subset of children at adolescence were read as normal in ~50% of cases and, in the others, there were no consistent relationships between radiological abnormalities and IQ results. Such children can, however, have relatively subtle brain abnormalities that are not seen on conventional MRI, and we hypothesized that these would be related to declines in IQ. Voxel-based morphometry (VBM) analyses of the MRI scans revealed that absolute IQ scores were related to areas in both the parietal and temporal lobes. The analyses also showed that frontal and temporal lobe regions were associated with the decline in VIQ, while occipital and temporal lobe regions (including the hippocampi) were associated with the decline in PIQ. Hippocampal volume measurements were consistent with the VBM findings. We concluded that preterm children are at risk of declining IQ over time even if they have not suffered obvious neurological damage and that the decline is associated with specific neural regions. Whether this is true of children born at >30 weeks gestation and what other factors predispose to this decline have yet to be determined.

Introduction

Although it is assumed that measured intelligence remains fairly stable after the age of 5 years in normal individuals (Zigler *et al.*, 1984), there have been relatively few empirical demonstrations of this effect. High correlations have been observed between IQ scores at different time points in adulthood [e.g. $r = 0.79$ between 19 and 50 years (Owens, 1966); $r = 0.73$ between 30 and 41 years (Kangas and Bradway, 1971)], while correlations of a similar size have been observed across childhood in cohorts of non-learning-disabled children [e.g. $r = 0.72$ between 4 and 13 years (Sameroff *et al.*, 1993); $r = 0.78$ between 9 and 15 years (Humphreys, 1989)]. Deary *et al.* (2000) found a correlation of 0.63 in the scores of a test of general mental ability, the Moray House Test, across the whole life span from first administration at the age of 11 years to re-test at the age of 77 years.

These studies describe the degree to which the position or rank of an individual's score is maintained within a group over time. This 'relative stability' (Thompson and Molly, 1993) between two time points will be high when most

scores remain the same, or when their magnitude changes to the same extent and in the same direction. The term 'stability', however, can also refer to how consistent the scores are in terms of their absolute value rather than rank. Because the correlation coefficient is independent of the means of the distributions, high relative stability may exist along with marked changes in the absolute value of the IQ scores (McCall *et al*., 1973). Studies examining the 'absolute stability' of IQ scores have produced a variety of outcomes, although McCall *et al*. (1973) point out that IQ tests, when repeated, almost always show an increase in scores with age. Deary *et al*. (2000) reported that participants scored significantly higher in adulthood compared with childhood, although the elderly may show declining scores (Deary *et al*., 1999). Kangas and Bradway (1971) reported increases when scores were obtained at 4, 14, 30 and 41 years. The well-known Flynn effect (Flynn, 1984, 1987) describes a general upward trend in the mean IQ score over time in the general population; the re-standardization of IQ tests is undertaken partly to correct for this tendency. The finding of increased scores over time, while typical, is not universal. Mortensen and Kleven's (1993) sample tested at 50 and 70 years old showed small reductions in IQ, as did Bauman's (1991) follow-up of 8-year-old children at the age of 11 years. Finally, Canivez and Watkins (1998) reported that three indices of IQ remained virtually unchanged in children between 9 and 12 years. Nevertheless, the results of the above studies support the general conclusion that IQ scores, particularly in a young population, tend to increase over time.

There are indications in the literature, however, that certain atypical populations may not follow the usual pattern. Banich *et al*. (1990) reported that children with congenital hemiplegia showed decreasing IQ scores after the age of 6 years. They speculated that these children were not able to gain cognitive skills at the same rate as children whose brains were intact. If early damage limits the ability of the brain to undergo normal developmental changes, then disparities in function between these children and age-matched controls would increase over time. Interestingly, the decrease in IQ was not observed in children whose hemiplegia was acquired due to a documented incident occurring after birth and that was not, therefore, congenital in nature.

Children born preterm constitute another population subject to injury in the pre/perinatal period. In some cases, this takes the form of frank lesions that can be related to deficits in function, both motor and cognitive. In other cases, however, neurological examination is normal and visual inspection of MRI scans does not reveal consistent abnormalities that correlate with functional deficits. Recent reports have indicated, however, that it is possible to demonstrate subtle neural anomalies associated with specific deficits in some such cases (Isaacs *et al*., 2000, 2001, 2003). Studies have shown that children born preterm or with very low birth weight (<1500 g) have lower IQ scores than age-matched control children (e.g. Klein *et al*., 1989; Saigal *et al*., 1990, 2000; Wolke and Meyer, 1999). As well as being lower, meta-analysis indicates that the IQ scores of preterm children decrease over time, diverging from those of full-term children after the age of 2 years (Aylward *et al*., 1989).

Studies of preterm children often include those with obvious neural insults such as intraventricular haemorrhage and periventricular leukomalacia. It could be that the decline in IQ scores is largely attributable to the inclusion of these children in the samples. O'Brien *et al*. (2000), for example, showed a decline of ∼9 points in mean full-scale IQ (FSIQ) scores between the ages of 8 and 15 years, but the sample included children with neurological impairments. In the absence of obvious neurological impairment, is a decrease in IQ still observed? Since even those children who appear normal on examination have been shown to have subtle cortical abnormalities (Isaacs *et al*., 2000, 2001, 2003), we tested the hypothesis that, as a group, preterm children who were apparently neurologically normal would demonstrate a decline in IQ in the period between 7 years of age and adolescence.

Declining IQ scores may be associated with alterations in neural structure, either long-standing abnormalities or changes that occur later in childhood. Very few studies have investigated these relationships and none has been conducted in young subjects without obvious neurological impairment. Garde *et al*. (2000) found that hyperintensities in white matter explained a small part of the variance in age-related decline in IQ in 80-year-old subjects. Mulhern *et al*. (2001) reported relationships between volume loss in normal-appearing white matter and decline in IQ in children who had received cranial radiotherapy for treatment of medullablastoma, but there were no significant relationships between cognition and normal-appearing grey matter in this clinical sample. Studies relating absolute IQ scores, rather than changes, to structure have been more numerous, although surprisingly little is known about this relationship. The neural regions most consistently identified as important contributors to variance in FSIQ scores are bilateral areas of grey matter in the frontal (Reiss *et al*., 1996; Flashman *et al*., 1998; Thompson *et al*., 2001; Wilke *et al*., 2003) and temporal lobes (Andreasen *et al*., 1993; Peterson *et al*., 2000). Relationships have also been reported between IQ and subcortical grey matter (Reiss *et al*., 1996) and the hippocampus (Andreasen *et al*., 1993; Abernethy *et al*., 2002). The second aim of this study, therefore, was to determine whether changes in IQ scores in preterm children were associated with altered morphology in the brain. To this end, we collected MRI scans that were then entered into voxel-based morphometry (VBM) analyses. VBM analyses can reveal subtle structural differences in neural architecture that may not be detectable by visual inspection of MRI scans (Wright *et al*., 1995). The study was designed to test the hypothesis that preterm children without obvious neurological impairment would demonstrate decreasing IQ scores over time and that this decline would be associated

with abnormalities in specific neural areas such as those identified in previous research.

Methods
Subjects

Eighty-two children (47 males and 35 females) took part in the study; they were members of a cohort of preterm infants born between 1982 and 1985 (Lucas *et al.*, 1992) who had participated in a series of follow-up studies. All children in the present study had been born at 30 weeks gestation or less (mean = 28.5 weeks; range: 26–30). They were first seen for IQ assessment in childhood (mean = 7 years 6 months; range: 7 years 4 months–7 years 10 months) and we included in this study only those classified as neurologically normal, defined as an absence of neuromotor or neurosensory impairment, based on the results of a paediatric neurological examination carried out at that time. At adolescence, mean age at assessment was 15 years 3 months (range: 12 years 5 months–16 years 11 months).

IQ data at two time points were available for all 82 children. MRI was carried out only at adolescence; two children did not attend for the MRI session. Scans were available for 80 children, but the first 12 of these were obtained on a scanner that subsequently was replaced. It is desirable in VBM analyses that all scans be collected on the same system, so these 12 scans were not used in subsequent analyses. Three scans were discarded for technical reasons, leaving a total of 65 scans available for analysis.

Procedure

At the childhood assessment, data were collected during one testing session carried out at either the child's school or home. At adolescence, two testing sessions were conducted, one at Great Ormond Street Hospital and a second at the child's choice of home or school. All children and parents gave informed, written consent, and the study was approved by the local hospital and regional ethics committees (The Dunn Nutrition Unit, The Great Ormond Street/Institute of Child Health, Norwich District, South Sheffield Research, East Suffolk Local Research and Cambridge Local Research Ethics Committees).

Childhood intelligence test
Wechsler Intelligence Scale for Children-Revised (WISC-R)

A short form of the WISC-R (Wechsler, 1974) was administered, following standard procedures (Sattler, 1992); estimates of verbal IQ (VIQ), based on the Similarities, Arithmetic and Vocabulary subtests, and performance IQ (PIQ), based on the Block Design and Object Assembly subtests, were calculated. IQ scores generated by the WISC-R have a mean of 100 and an SD of 15.

Adolescence intelligence test
Wechsler Intelligence Scale for Children-3rd Edition (WISC-III)

The WISC-III (Wechsler, 1992) was given in full, following standard procedures, and VIQ and PIQ were calculated. The population mean for this test is 100 and the SD is 15. Pro-rated VIQ and PIQ scores, based on the subtests administered at the childhood assessment, were also calculated for use in analyses.

Image acquisition

MRI studies were performed at adolescence using a 1.5 T Siemens Vision system. Investigations included: (i) magnetization-prepared rapid acquisition gradient echo (MPRAGE 3-D) (Mugler and Brookeman, 1990) volume acquisition with repetition time of 10 ms; echo time, 4 ms; inversion time, 200 ms; flip angle, 12°; matrix size, 256 × 256; field of view, 250 mm, partition thickness, 1.25 mm; 128 sagittal partitions in the third dimension, and acquisition time, 8.3 min; and (ii) coronal and axial turbo spin-echo T2-weighted scans with typical repetition times of 3500–4600 ms; echo time, 90–96 ms; matrix size 196 × 512; field of view 158 × 210 mm; slice thickness 5 mm; and acquisition time 3.4–4.3 min for each orientation.

Image analysis

The scans were first inspected visually by an experienced paediatric neuroradiologist, blind to group membership and all cognitive data. The 3D MRI data sets were then analysed using VBM (Ashburner and Friston, 2000), a method of analysing structural scans that compares regional grey or white matter signal intensities on a voxel-by-voxel basis while controlling for total tissue volume. Voxels showing statistically significant differences in grey or white matter are displayed on an output map (statistical parametric map; SPM). The method is particularly appropriate when the abnormalities are of very early origin, as in preterm birth, because these are likely to be at the level of neural organization with resulting subtle morphometric differences, not always apparent on conventional neuroradiological assessment of MRI scans. The data sets were processed using SPM99 software (Wellcome Department of Imaging Neuroscience; http://www.fil.ion.ucl.ac.uk/spm; Friston *et al.*, 1995), running in Matlab5 on a SUN workstation. They were first spatially normalized to a template constructed from a collection of 20 data sets from normal children aged between 8 and 17 years (themselves normalized to the template in SPM-T1.img). The normalized scans were then segmented into grey matter, white matter, CSF and scalp images. Each voxel was classified as belonging to one of the four categories, based on both its signal intensity and its location in the brain. The segmented grey and white matter images were smoothed, using a 12 mm full width half-maximum isotropic kernel, and then entered into statistical analyses. Because of the wide range of age at testing at adolescence, we first correlated age at test in months with grey and white matter. Since these analyses produced some significant results, we used age as a covariate in all subsequent VBM analyses. We then correlated both absolute IQ scores and IQ decline scores with grey and white matter intensities. Finally, a two-sample *t* test design was used to compare groups showing different degrees of IQ decline in order to identify regions of greater or lesser grey and white matter density. Because developmental anomalies are frequently bilateral, the scans both for the correlations and for the *t* tests were normalized to a symmetrical template and analysed using a conjunction analysis that searches explicitly for the presence of symmetrical bilateral abnormalities (Salmond *et al.*, 2000).

In addition to the VBM analyses, hippocampal volumes were also measured, blinded with respect to IQ scores and group membership. For these measurements, the 3D MPRAGE data sets were reformatted into 1 mm thick contiguous slices in a tilted coronal plane perpendicular to the long axis of the hippocampus. Cross-sectional areas were measured for every slice along the entire length of the hippocampi. The volumes were calculated by summing the cross-sectional areas and multiplying by the distance between the slices

(i.e. 1 mm). Intracranial volumes were measured from the unreformatted sagittal 3D MRI data sets. The hippocampal volumes were then corrected for intracranial volume as described by Van Paesschen et al. (1997), and they are presented here in this corrected form.

MRI analysis groups

In order to select groups for statistical analyses, each subject was assigned VIQ and PIQ change scores, defined as (IQ score at childhood – IQ score at adolescence); the more positive the score, the bigger the decline in IQ that had taken place. Change scores indicating a decline were then assigned percentile values and two groups were formed by selecting children above the 75th percentile (large decline group) and below the 25th (small decline group). For VIQ, large decline group, $n = 15$, >20 point drop; small decline group, $n = 15$, <6 point drop. For PIQ, large decline group, $n = 14$, >23 point drop; small decline group, $n = 13$, <8 point drop. For the correlational analysis of decline scores, all children showing a decrease (VIQ, $n = 54$; PIQ, $n = 49$ out of 65 children) were included. For these analyses of change scores, we excluded children who had increased their scores, as increases and declines might be underpinned by different, and hence confounding, neural changes.

The absolute IQ score analysis included those who showed no change or increases as well as decreases ($n = 65$), i.e. the full range of ability.

Results
Group characteristics
Total study group

Data concerning the perinatal status of the study group are presented in Table 1 along with the same data for the total cohort. Although perinatal data were available for the 1532 children recruited to the original study, the total included children who died subsequently or did not proceed to cognitive follow-up for various reasons. It seemed that the best reference group for the present study would be the children who took part in the first comprehensive cognitive follow-up at 18 months (1056 children) and it is these data that appear in Table 1. In the present study group, 15.9% were small for gestational age, 25.6% had had maternal steroids and 13.4% were asphyxiated.

The VBM groups

Perinatal data for the four VBM groups, VIQ and PIQ small decline and large decline, are given in Table 2. None of the small decline versus large decline group t test comparisons for the VIQ groups was significant and only that for Apgar score at 5 min was significant for the PIQ groups ($P < 0.021$). The percentages for the occurrence of other baseline characteristics in the four groups are given in Table 3. We tested the significance of the difference between these proportions, but none was significant.

Analyses of IQ scores
Relative stability

IQ scores obtained at the two time points, childhood and adolescence, correlated significantly, with the size of the coefficients similar to those reported in the literature. The correlation between VIQ scores at the two time points

Table 1 *Means (SDs) of the present study group and original cohort for perinatal variables and social class*

	Birthweight (g)	Gestational age (weeks)	Apgar 1 min	Apgar 5 min	Ventilation (days)	Days in 30% O$_2$	Social class
Original cohort ($n = 1056$)	1371 (301.2)	30.3 (2.9)	5.8 (2.3)	8.2 (3.6)	3.9 (7.5)	9.6 (16.8)	3.6 (1.7)
Present study group ($n = 82$)	1175 (250.4)	28.5 (1.2)	5.3 (2.4)	7.6 (1.9)	(6.4) (9.1)	16.1 (20.0)	3.4 (1.7)

Table 2 *Means (SDs) of the four VBM groups for perinatal variables and social class, and P values for group comparisons*

	Birthweight (g)	Gestational age (weeks)	Apgar 1 min	Apgar 5 min	Ventilation (days)	Days in 30% O$_2$	Social class
VIQ small decline ($n = 15$)	1179 (175.5)	28.6 (1.1)	5.7 (2.3)	7.6 (2.3)	7.3 (11.1)	16.4 (20.8)	3.1 (90.3)
VIQ large decline ($n = 15$)	1083 (231.0)	28.1 (1.5)	4.6 (2.9)	7.4 (2.3)	6.0 (8.5)	14.6 (20.8)	3.4 (0.6)
P value for t test between VIQ groups	0.21	0.29	0.27	0.80	0.71	0.81	0.69
PIQ small decline ($n = 13$)	1192 (242.1)	28.6 (1.2)	6.5 (2.6)	8.7 (1.5)	1.6 (2.8)	7.7 (13.9)	3.4 (1.6)
PIQ large decline ($n = 14$)	1273 (272.8)	28.7 (1.3)	4.8 (2.1)	7.0 (1.8)	5.7 (10.8)	11.1 (19.4)	3.2 (1.9)
P value for t test between PIQ groups	0.43	0.84	0.08	0.20	0.20	0.62	0.60

(r = 0.674, P < 0.001) was somewhat higher than the corresponding correlation for PIQ scores (r = 0.498, P < 0.001).

Absolute stability

Table 4 presents mean IQ scores, SDs and ranges, obtained at childhood and adolescence; the mean scores at both time points were within the average range of 90–109, defined by Wechsler (1992). Paired t tests revealed that both VIQ and PIQ mean scores decreased significantly between the two time points, VIQ scores by ∼9 points, [VIQ $t(81)$ = 6.95, P < 0.001], and PIQ scores by ∼12 points [PIQ $t(81)$ = 7.26, P < 0.001].

We considered two factors that might have accounted for the observed decreases. The WISC-R had been replaced by the WISC-III in the interval between assessments. Could different forms of the test have caused a decline in mean scores? With restandardization, Wechsler (1992) reported that VIQ and PIQ scores obtained using the WISC-R were ∼2 and 7 points higher than the corresponding scores obtained using the WISC-III. In order to compare scores obtained using the two different measures, we added 2 points to each child's WISC-III VIQ score and 7 points to the WISC-III PIQ score, as suggested by Kaufman (2001). These adjusted mean scores are also reported in Table 4. Repeating the above analyses on

these adjusted scores showed that the differences over time remained: in VIQ, $t(81)$ = 5.322, P < 0.001; and in PIQ, $t(81)$ = 2.970, P < 0.004.

A second consideration was the use of two different forms of the IQ test, a short form in childhood and a full form at adolescence. We therefore calculated pro-rated IQ scores at adolescence using the same subtests administered in childhood (listed in the Methods; means reported in Table 4). A short-form PIQ could not be calculated for five children (one male and four females) because they were not given the Object Assembly subtest. These short-form scores correlated well, as would be expected, with full-form scores for both VIQ (r = 0.958, P < 0.001) and PIQ (r = 0.873, P < 0.001). When we compared the short forms obtained at the two ages, a significant decrease existed for both: VIQ, $t(81)$ = 6.511, P < 0.001; and PIQ, $t(81)$ = 6.595, P < 0.001.

One final adjustment to the IQ scores at adolescence was the application of the same corrections made to the full-form scores to the calculated short-form scores (i.e. VIQ short form + 2 and PIQ short form + 7). With these adjustments, significant reductions in both VIQ and PIQ scores remained [VIQ, $t(81)$ = 6.414, P < 0.001; PIQ, $t(81)$ = 5.526, P < 0.001].

In summary, the decline in mean IQ score demonstrated by the preterm children could not be fully accounted for by restandardization of the WISC, nor the use of short/long forms. Since the analyses of IQ score changes produced the same results whichever form of IQ scores were used, the unadjusted scores obtained at adolescence were used in all the VBM studies.

These reductions in mean VIQ and PIQ scores were reflected by high frequencies of individual children showing decreases in IQ scores over time, indicating that the reductions could not be attributed to a few children with unusually large declines. We defined change in an individual child as a difference in scores between the two time points, in either direction, that was larger than the measurement error of VIQ (3.6 points) or PIQ (4.7) points (Wechsler, 1992). The percentages of children thus classified as having an increase or decrease in IQ scores are reported in Table 5. In every case, the proportion of children showing decreased scores was markedly greater than the proportion showing increased scores. Measurement error alone would be expected to produce roughly equal numbers of children falling above and below the error band. Similarly, regression towards the mean would result in as many children with low initial scores increasing their scores as those with high initial scores

Table 3 *Percentages of occurrence of certain baseline characteristics in the four VBM groups, and P values for group comparisons*

	% with asphyxia	% maternal steroids	% SGA
VIQ small decline (n = 15)	20.0	40.0	20.0
VIQ large decline (n = 15)	13.3	26.7	13.3
P value for VIQ group comparisons	0.58	0.46	0.58
PIQ small decline (n = 13)	7.7	30.8	23.0
PIQ large decline (n = 14)	21.4	14.3	14.3
P value for PIQ group comparisons	0.22	0.08	0.42

SGA = small for gestational age.

Table 4 *Means (SDs and ranges) for VIQ and PIQ scores at 7 years and at adolescence, with adjustments for possible measurement error*

	Age 7 years	Adolescence full form	Adolescence full form adjusted[*]	Adolescence short form	Adolescence short form adjusted[*]
VIQ	105.5 (13.8) (73–154)	96.9 (13.7) (73–136)	98.9 (13.8) (75–138)	97.8 (13.8) (74–133)	99.8 (13.8) (76–135)
PIQ	107.3 (15.5) (71–146)	95.5 (13.8) (71–135)	102.5 (13.8) (78–142)	95.7 (18.4) (55–152)	102.7 (18.4) (62–159)

[*]IQ scores adjusted to take into account the transition from WISC-R to WISC-III IQ tests (see text for details).

Table 5 *Percentages of children showing increases and decreases in VIQ and PIQ between 7 years and adolescence, with adjustments for possible measurement error*

		Adolescence full form	Adolescence full form adjusted[*]	Adolescence short form	Adolescence short form adjusted[*]
VIQ	% children with decrease	62.2	56.1	59.8	51.2
	% children with increase	11.0	15.9	14.6	20.7
PIQ	% children with decrease	68.3	50.0	67.1	52.4
	% children with increase	14.6	30.5	17.1	31.7

[*]IQ scores adjusted to take into account the transition from WISC-R to WISC-III IQ tests (see text for details).

Table 6 *Pearson r values (P values) for the relationships between VIQ and PIQ change scores and perinatal variables*

	Birthweight (g)	Gestation (weeks)	Apgar 1 min	Apgar 5 min	Ventilation (days)	Days in 30% O_2	Social class
VIQ change	0.084 (0.45)	0.088 (0.43)	0.074 (0.52)	0.082 (0.49)	−0.073 (0.46)	−0.082 (0.47)	0.057 (0.61)
PIQ change	0.160 (0.15)	0.173 (0.12)	0.072 (0.53)	0.052 (0.67)	−0.093 (0.41)	−0.105 (0.35)	0.011 (0.92)

decreasing them, so neither of these explanations can account fully for our observations.

We also compared the scaled scores for the IQ subtests common to assessment at both time points, using paired *t* tests. Mean scores on every subtest except Vocabulary decreased significantly [Similarities, $t(81) = 7.25$, $P < 0.001$; Arithmetic, $t(81) = 2.88$, $P < 0.005$; Block Design, $t(81) = 7.02$, $P < 0.001$; Object Assembly, $t(81) = 3.53$, $P < 0.001$].

IQ change and perinatal factors

We correlated the VIQ and PIQ change scores with perinatal factors to see if any of these were related to the differences in IQ over time. Although none of these correlations approached significance, they are given in Table 6 for information.

Initial IQ in the VBM groups

Table 7 presents mean VIQ and PIQ scores obtained for the four VBM groups at the first assessment to address the question of whether the observed declines in IQ scores are due simply to large decreases in those who score high initially. The fact that this regression towards the mean is unlikely because of the disproportionate number of declines versus increases has been noted above, but these data reinforce this point. Mean IQ scores do not differ between the small decline and large decline groups, in both the VIQ and PIQ comparisons. They are significantly different for initial PIQ scores in the same analyses, but the difference is in opposite directions for the VIQ and PIQ decrease groups.

These results are not those that would be expected from regression towards the mean.

The hypothesis that IQ declines in this population over time would suggest that there would be a positive correlation between size of the decline and age at test in adolescence. This was true for VIQ ($r = 0.296$, $P < 0.007$), but not for PIQ ($r = 0.035$, $P < 0.758$).

Analyses of MRI scans at adolescence
Visual inspection

Details of the results of visual inspection of the scans are given in Table 8; the decline score data are divided into quartiles, with the top and bottom representing the children included in the VBM groups. It is apparent from Table 8 that normal scans and those with abnormalities are approximately evenly distributed between the two groups and cannot, therefore, be meaningfully related to cognitive changes. There is no consistent relationship between presence/absence of abnormality and IQ.

Voxel-based morphometry

A VBM analysis carries out voxel-by-voxel comparisons across the whole brain, meaning that a very large number of *t* tests is inherent in the method. Significance values corrected for these multiple comparisons are used when there is no *a priori* hypothesis about the locus of grey/white matter differences, while uncorrected levels of significance may be used, by convention, when such hypotheses exist. We have noted in the Introduction that correlations with IQ might be predicted to occur in the frontal and temporal lobes and in the hippocampus. Since

Table 7 *The mean (SD) VIQ and PIQ scores obtained at childhood for the VBM groups and the outcome of t test comparisons*

	VIQ drop			PIQ drop		
	Small decline	Large decline	*P* value	Small decline	Large decline	*P* value
VIQ in childhood	108.0 (11.6)	111.9 (7.7)	NS	112.3 (9.2)	111.3 (10.5)	NS
PIQ in childhood	117.8 (12.8)	97.9 (17.3)	<0.002	108.4 (12.1)	120.3 (12.4)	<0.022

Table 8 *Results of the visual inspection of MRI scans for each quartile of VIQ and PIQ decline*

	VIQ	PIQ
<25th percentile (VBM small decline)	$n = 15$: 6, normal; 8, small corpus callosum; 1, immature myelination pattern; 1, peritrigonal atrophy	$n = 15$: 8, normal; 4, small corpus callosum; 1, immature myelination; 1, small hippocampi; 1, peritrigonal atrophy
25th to 50th percentile	$n = 14$: 10, normal; 3, small corpus callosum; 1, porencephaly; 1, mild PVL	$n = 12$: 9, normal; 1, small corpus callosum; 1, immature myelination; 1, mild PVL
51st to 75th percentile	$n = 10$: 8, normal; 1, small corpus callosum; 1, immature myelination	$n = 8$: 3, normal; 2, small corpus callosum; 1, small hippocampi; 1, immature myelination; 1, watershed abnormalities
(>75th percentile) (VBM large decline)	$n = 15$: 7, normal; 2, small corpus callosum; 2, small hippocampi; 2, immature myelination pattern; 1, perintrigonal atrophy; 1, mild PVL	$n = 14$: 9, normal; 3, small corpus callosum; 1, porencephaly; 1, mild PVL

Some children have more than one abnormality and some are included in more than one group, since VIQ and PIQ are presented separately.

the frontal and temporal lobes encompass a very large proportion of the brain, we do not consider these regions specific enough to constitute an adequate hypothesis in VBM terms and, although we will report all findings in these areas, we give most weight to those regions that reach a corrected significance of $P < 0.05$. Where no hypothesis existed, we report only those reaching this level of significance. Talaraich x, y, z coordinates are reported for the peak voxel identified in each cluster (P values given below are corrected for multiple comparisons unless otherwise stated). As pointed out in the Methods, the results refer to the presence of symmetrical bilateral abnormalities.

Age at testing

The wide range in age at testing at adolescence was noted earlier. As a first step, therefore, we correlated age at test with grey and white matter to see if there were any significant relationships. There were no significant positive or negative correlations in the white matter analysis. For grey matter, there was one significant positive correlation in the region of the thalamus/putamen (± 21, -26, 10: $P < 0.01$). In view of this finding and because several grey and white matter analyses came close to the $P < 0.05$ level, we used age at test as a covariate in all the following analyses.

Absolute VIQ and PIQ scores

VBM correlational analyses were carried out to see if there were any significant relationships between VIQ and PIQ scores obtained at adolescence and grey and white matter density, with age at test used as a covariate. There was a negative correlation between VIQ and grey matter in the vicinity of the angular gyrus and intraparietal sulcus in the parietal lobe (± 40, -70, 30; $P < 0.01$; see Fig. 1), and a positive correlation between VIQ and a similar region of white matter (± 39, -69, 28; $P < 0.01$). Interestingly, grey matter in the same region (± 40, -74, 28; $P < 0.05$) also showed a negative correlation with PIQ, thus implicating it in both aspects of IQ function. A further significant negative correlation was found between PIQ and grey matter in a region including the fusiform gyrus in the temporal lobe (± 44, -50, -10; $P < 0.05$).

VIQ change scores

Correlation analyses demonstrated a significant positive correlation between the magnitude of VIQ decline and white matter in a frontal lobe region underlying the medial/superior frontal gyri (± 18, 58, 3; $P < 0.03$) and a negative correlation in the temporal lobe (near the anterior transverse temporal sulcus and gyrus (± 46, -22, 14; $P < 0.02$). Consistent with
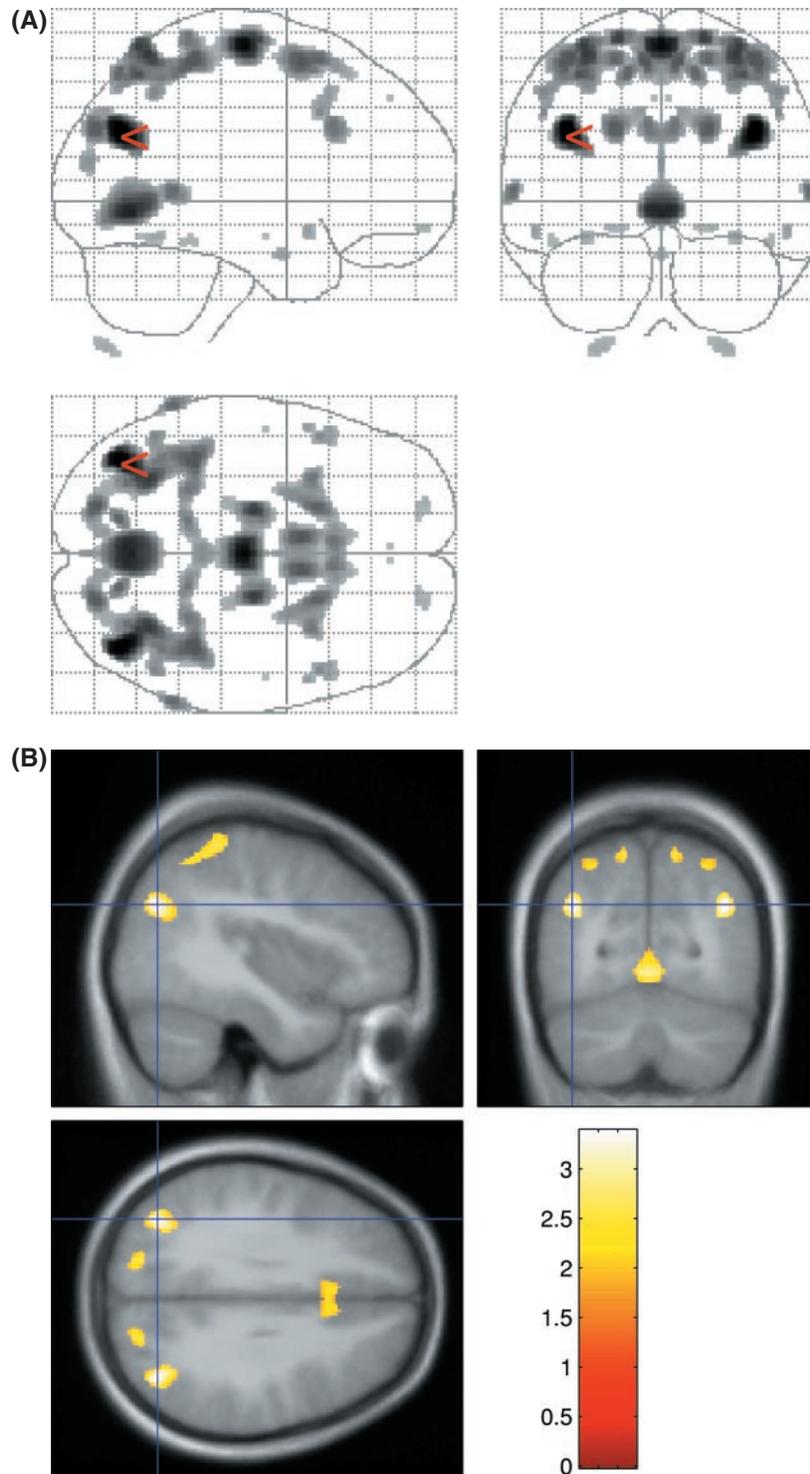
Fig. 1 Statistical parametric maps showing the regions where there was a significant negative correlation between VIQ and grey matter density: (**A**) glass brain representation; (**B**) the superimposition of Z-scores on the mean anatomical image is shown in colour for planes through the most significant parietal lobe voxels. Left is left in accordance with neurological convention. A threshold of $P < 0.001$, uncorrected, was chosen for display.

the first of these results, the *t* test analysis showed that the large decline group had significantly more white matter than the small decline group in the same frontal region ($\pm 22, 62, 3$; $P < 0.05$). Although the grey matter analyses revealed no

significant relationships with changes in VIQ at a $P < 0.05$ level, it is interesting to note that the large decline group had less grey matter in the same region as the white matter increase ($\pm 18, 58, 3$) at $P < 0.11$. Consistent with this, the
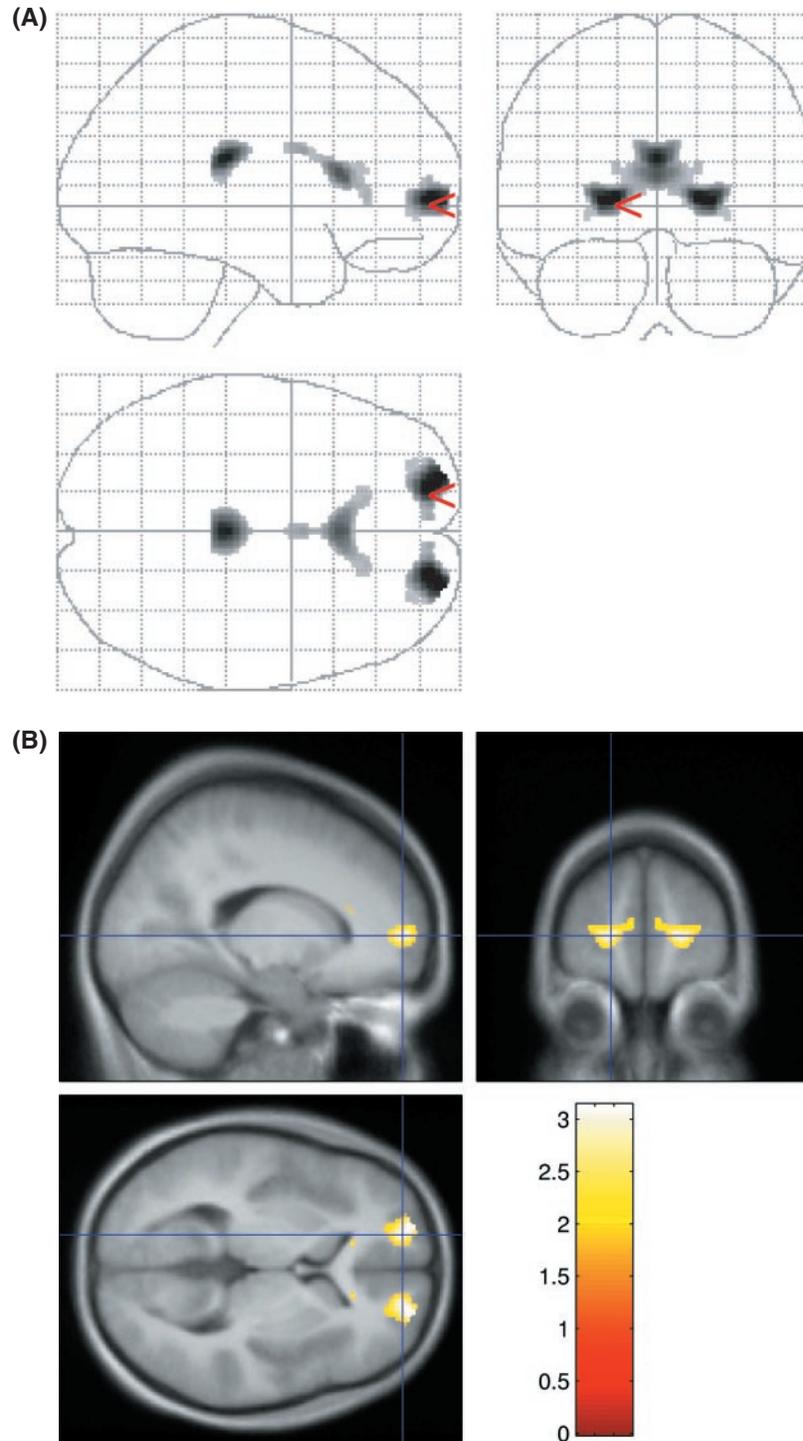
Fig. 2 Statistical parametric maps showing the regions where there was a significant positive correlation between the VIQ decline and white matter density: (**A**) glass brain representation; (**B**) the superimposition of Z-scores on the mean anatomical image is shown in colour for planes through the most frontal lobe significant voxels. Left is left in accordance with neurological convention. A threshold of $P < 0.001$, uncorrected, was chosen for display.

grey matter correlation, although producing no significant results at a corrected level, indicated a negative relationship at $\pm16, 57, 4$ ($P < 0.28$), reciprocal to the significant white matter finding described above. This frontal lobe region is shown in Fig. 2.

PIQ change scores

Different areas were associated with changes in PIQ. There was a significant negative correlation between the PIQ decrease scores and grey matter density in the region of the hippocampi ($\pm38, -21, -12$; $P < 0.03$; see Fig. 3).
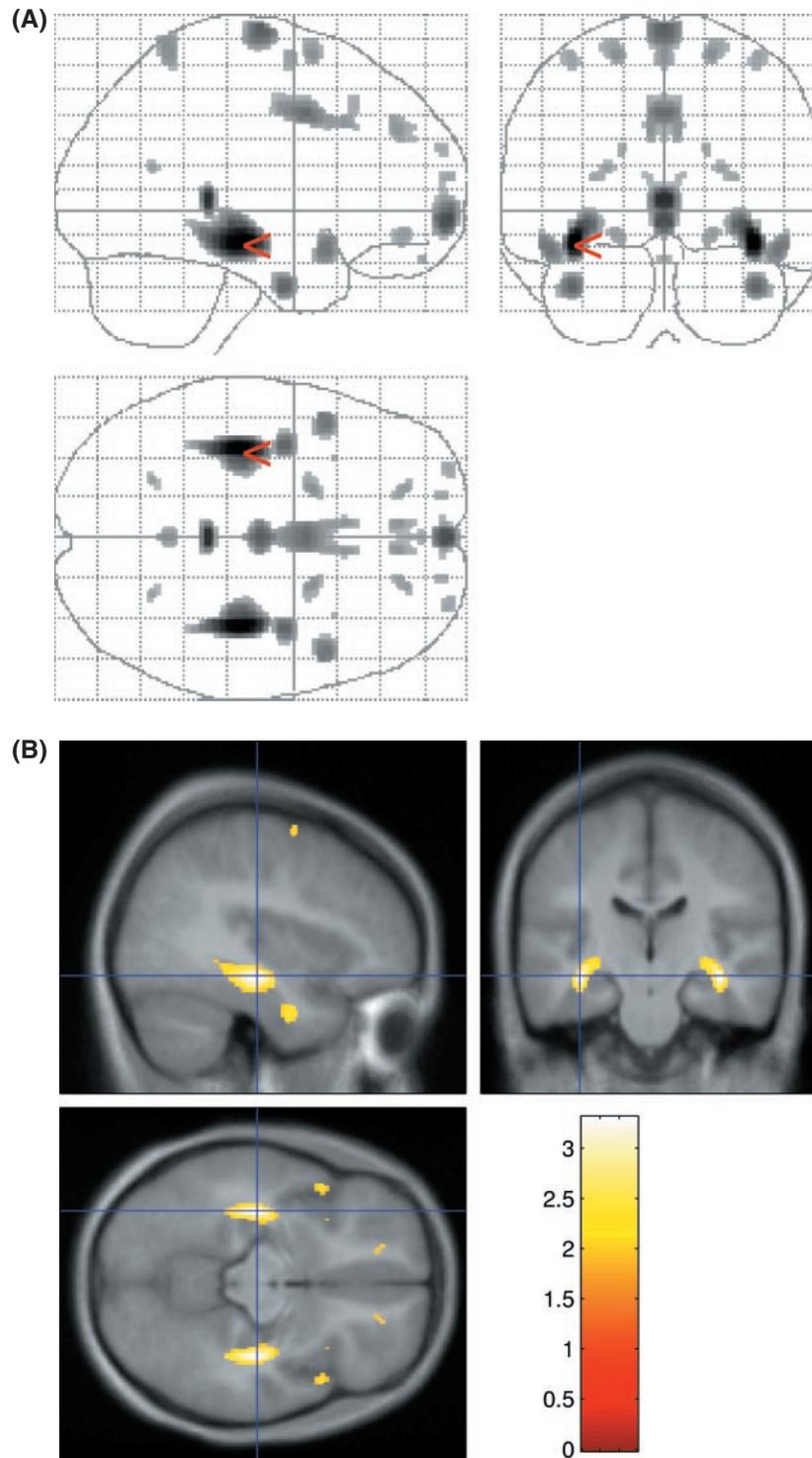
Fig. 3 Statistical parametric maps showing the regions where there was a significant negative correlation between the PIQ decline and grey matter density: (**A**) glass brain representation; (**B**) the superimposition of Z-scores on the mean anatomical image is shown in colour for planes through the most significant voxels in the hippocampal region. Left is left in accordance with neurological convention. A threshold of $P < 0.001$, uncorrected, was chosen for display.

The *t* test analysis indicated that the small decline group had significantly more grey matter in these regions than did the large decline group (±36, -20, -10; $P < 0.15$ corrected, $P < 0.001$ uncorrected). The hippocampal region was also identified as significant in the white matter correlation, where a positive relationship was shown between PIQ decrease scores and a cluster with peak voxel at ±38, -26, -14 ($P < 0.02$). The large decline group also had

more white matter than the small decline group in the occipital lobe (white matter underlying the striate cortex; 26, −102, 0; $P < 0.002$) and in the temporal lobe (underlying the middle temporal gyrus; ±50, −8, −34; $P < 0.03$). In relation to this latter finding, the large decline group showed less grey matter in the same region (±48, −6, 34), although this was not significant at a corrected level ($P < 0.28$).

The scans that were entered into these analyses had been smoothed with a 12 mm kernel which is optimal for identifying cortical differences. However, in view of this hippocampal finding, we repeated the t test analysis with the scans smoothed to 4 mm, as this degree of smoothing might be preferable given the small cross-sectional dimensions of the hippocampus. Again, there was a highly significant difference in grey matter intensity between groups only in the hippocampi (±18, −20, −15; $P < 0.04$), with the small decline group having more grey matter than the large decline group. The differing (more anterior) location obtained with this degree of smoothing presumably is a reflection of the precise pixel-wise distribution of abnormalities in the two hemispheres and of the way in which these abnormalities, after smoothing, fall into exactly homologous brain regions.

Hippocampal volume measurements were consistent with the VBM findings. The mean volumes for the large decline group were 3397 ± 298 and 3496 ± 307 mm^3 for the left and right hippocampi, respectively, and for the small decline group the corresponding values were 3538 ± 232 and 3699 ± 280 mm^3. The differences between the groups approached or reached significance in a one-sided t test ($P < 0.10$ and $P < 0.05$ for the left and right hippocampi), in concordance with the VBM findings, and suggest a mean hippocampal volume reduction of ∼4–5% in the large decline group compared with the small decline group.

Discussion

The first important finding was that, as a group, the children we studied, born preterm at a gestational age of 30 weeks or less, showed a decline in both VIQ and PIQ scores over the period between 7 years of age and adolescence. Since all children in the study were considered to be neurologically normal, based on examination at 7 years, the decline cannot be attributed to the inclusion of children with neurological disabilities. Nevertheless, in a majority of the children, absolute IQ scores decreased over time in a manner that would not be expected in a normal population. We considered whether the change in IQ test over time and the use of short forms might have caused these declines, but the declines remained once these were taken into account, so while they may have contributed to the magnitude of the decrease in scores, they were not enough in themselves to explain it totally. A comparison of the frequencies of declines and increases in IQ scores also made it unlikely that regression towards the mean or measurement error could fully explain the results. Further research will be needed to determine whether these effects

are also observed in preterm children born later than 30 weeks gestation.

The decline in scores would not have been apparent if only relative stability had been assessed. The correlations between IQ scores at the two time points were in line with those reported by Deary *et al.* (2000) for a variety of studies over a wide age range, concealing the fact that a large number of children obtained lowered scores at adolescence. We found that the relative stability of VIQ scores was somewhat higher than that of PIQ scores ($r = 0.674$ versus 0.498), as Deary *et al.* (2000) also report.

The study by Banich *et al.* (1990), cited above, suggests an interesting interpretation of the present results. Both the congenital hemiplegics in that study and the preterm children in the present study were drawn from populations where any neural damage had been sustained prenatally or perinatally. The congenital hemiplegics showed decreasing IQ scores with increasing age, apparent from the age of 6 years. On the other hand, children with acquired hemiplegia, caused by lesions that were similar but sustained postnatally, did not show a decrease in IQ scores over time. Two explanations are offered. One is that there is some sparing of function after injury initially, since damaged areas may not contribute to performance early in life. As the relevant cortical areas mature, their impaired ability to contribute to task execution results in the appearance of deficits. An alternative explanation is that early damage attenuates the ability to maintain a normal rate of development throughout childhood, particularly as the complexity of the cognitive demands increases. Banich *et al.* (1990) point out that these are not mutually exclusive explanations and both may have a role to play. It may be that the situation with the preterm brain is analogous. It is interesting to note that the only IQ subtest score not to decline significantly in the preterm children was that for Vocabulary. This particular subtest is often used as the best predictor of premorbid intellectual status in adults with brain injury because it is relatively impervious to the effects of neural dysfunction (Yates, 1954).

If IQ declines over time, then we might expect to see a positive relationship between age at test and the size of the decline from some earlier measurement. This proved to be true for VIQ in the present study but not for PIQ. Since large declines in PIQ scores were apparent between childhood and adolescence, the implication is that most of the PIQ decline had occurred before the youngest age at test at adolescence, i.e. before 12 years 5 months. This suggests that VIQ and PIQ declines follow different trajectories, an interesting subject for further research.

The MRI analyses were designed to address the question of whether there were demonstrable relationships between brain structure and several aspects of IQ. We first looked at absolute IQ scores to determine whether any areas of grey and white matter varied systematically with variation in IQ score, both verbal and performance. Higher scores for both IQ scores were associated with less grey matter, and more white matter, in a region of the parietal lobe, at a highly significant level. IQ

has been associated with the parietal lobe in previous reports. Peterson *et al.* (2000), for example, found significant relationships between bilateral parieto-occipital volumes and both VIQ and PIQ in a cohort of 8 year olds born prematurely. Flashman *et al.* (1998) also reported associations between the parietal lobe and PIQ scores. For PIQ only, the same pattern of less grey matter associated with higher scores was found in the fusiform gyrus area of the temporal lobes. The association between the temporal lobes and IQ has been well documented and, more specifically, there is a striking consistency between our finding and a report by Blanton *et al.* (2003) that normal children (age range: 6–16 years) showed an inverse relationship between cortical grey matter thickness and FSIQ scores in the left and right fusiform gyri.

When we considered change in IQ scores, different neural regions were identified. Both correlation and *t* tests showed that as the decline in VIQ scores became larger, more white matter, and less grey matter, was found in certain regions of the frontal lobe; the implication is that more grey matter in these regions is advantageous. More VIQ decline was also associated with less white matter in the temporal lobe. Sowell *et al.* (2003) have reported that the temporal cortices have the most protracted course of maturation of all neural regions, with myelination continuing into adulthood. Since the period up to adolescence is one of active myelination, it may be that these findings relate to how successfully this process is proceeding, with less white matter in these regions being suboptimal. PIQ decline scores were associated with significant differences in three regions. *t* tests showed that the large decline group had more white matter in areas of the occipital and temporal lobes, again with the implication that more grey matter in these areas leads to better outcome. Although we had not anticipated the occipital lobe finding, it was highly significant and perhaps is not surprising in view of the fact that the main input modality for the PIQ subtests is visual. The third region, the hippocampi, was identified as significant in both grey (negative) and white (positive) correlations, supported by the *t* test analysis that showed more grey matter in the small decline group. An association between absolute IQ scores and the hippocampi has been reported before in both adults and children, although change in IQ scores has not been studied. Andreasen *et al.* (1993) noted that PIQ was significantly related to the left hippocampus, but not the right, in adults, while VIQ showed a bilateral association. Abernethy *et al.* (2002) found that the volume of the left hippocampus was reduced in adolescents of low IQ (<85). The hippocampus, in view of its role in learning and memory, could play a part in maintaining a rate of information acquisition sufficient to ensure that IQ does not fall over time. More specifically, given the importance of the hippocampus for spatial processing, it may be that the role of the hippocampus in PIQ change is concerned with the spatial aspects of the PIQ subtests.

Although we have noted a relationship, we do not have the information available to determine causality. A comparison of scans obtained in childhood with those at adolescence would have allowed us to use VBM to determine whether structural changes had occurred over the same period as the IQ decrease, but these were not available. In their absence, we do not know if the structural differences are long standing, acting as a cause of the IQ decline, or whether they have occurred later in childhood or even as a by-product of IQ change itself. Neither do we know what caused the decline to occur in particular children and not in others. Although the ranges were restricted, we correlated perinatal factors with the VIQ and PIQ change scores but found no significant relationships. We are now conducting a study in which the ranges for these variables will be much wider, allowing us in the future to correlate these factors with IQ change but also with MRI data.

It seems that having a normal neurological examination does not preclude the possibility that a child born prematurely will show a decline in general cognitive level over time. The presence of subtle morphometric differences indicates that some insult to the brain may have taken place, even in the absence of frank lesions, with subsequent cognitive consequences. It is important now to extend the gestational age range to determine whether the decline in IQ over time also occurs in children born later than 30 weeks and whether the same differences in neural architecture accompany any changes. Meanwhile, the findings of this study contribute to our understanding of the relationship between brain and cognitive behaviour.

References

Abernethy LJ, Palaniappan M, Cooke RWI. Quantitative magnetic resonance imaging of the brain in survivors of very low birth weight. Arch Dis Child 2002; 87: 279–83.

Andreasen NC, Flaum M, Swayze V, O'Leary DS, Alliger R, Cohen G, et al. Intelligence and brain structure in normal individuals. Am J Psychiatry 1993; 150: 130–4.

Ashburner J, Friston K. Voxel-based morphometry—the methods. Neuroimage 2000; 11: 805–21.

Aylward GP, Pfeiffer SI, Wright A, Verhulst SJ. Outcome studies of low birth weight infants published in the last decade: a metanalysis. J Pediatr 1989; 115: 515–20.

Banich MT, Levine SC, Kim H, Huttenlocher, P. The effects of developmental factors on IQ in hemiplegic children. Neuropsychologia 1990; 28: 35–47.

Blanton R, Levitt J, Peterson J, Sadoun T, To D, Thompson P, et al. IQ correlations with regional gray matter decreases in normal children. Neuroimage 2003; 2, Suppl 1: s554.

Bauman E. Determinants of WISC-R subtest stability in children with learning difficulties. J Clin Psychol 1991; 47: 430–5.

Canivez GL, Watkins MW. Long-term stability of the Wechsler Intelligence Scale for Children-Third Edition. Psychol Assess 1998; 10: 285–91.

Deary IJ, MacLennan WJ, Starr JM. Is age kinder to the initially more able? Differential ageing of a verbal ability in the Healthy Old People in Edinburgh study. Psychol Rev 1999; 26: 357–75.

Deary IJ, Whalley LJ, Lemmon H, Crawford JR, Starr JM. The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. Intelligence 2000; 28: 49–55.

Flashman LA, Andreasen NC, Flaum M, Swayze VW II. Intelligence and regional brain volumes in normal controls. Intelligence 1998; 25: 149–60.

Flynn JR. The mean IQ of Americans: massive gains 1932 to 1978. Psychol Bull 1984; 95: 29–51.

Flynn JR. Massive IQ gains in 14 nations: what IQ tests really measure. Psychol Bull 1987; 101: 192–212.

Friston KJ, Holmes AP, Worsley K, Poline J-B, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 1995; 2: 223–35.

Garde E, Mortensen EL, Krabbe K, Rostrup E, Larsson HBW. Relation between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: a longitudinal study. Lancet 2000; 356: 628–34.

Humphreys LG. Intelligence: three kinds of instability and their consequences for policy. In: Linn RL, editor. Intelligence. Urbana: University of Illinois Press; 1989; 193–218.

Isaacs EB, Lucas A, Chong WK, Wood SJ, Johnson CL, Marshall C, et al. Hippocampal volume and everyday memory in children of very low birth weight. Pediatr Res 2000; 47: 713–20.

Isaacs EB, Edmonds CJ, Lucas A, Gadian DG. Calculation difficulties in children of very low birthweight: a neural correlate. Brain 2001; 124: 1701–7

Isaacs EB, Edmonds CJ, Chong WK, Lucas A, Gadian DG. Cortical anomalies associated with visuospatial processing deficits. Ann Neurol 2003; 53: 768–73.

Kangas J, Bradway K. Intelligence at middle age: a thirty-eight-year follow-up. Dev Psychol 1971; 5: 333–7.

Kaufman AS. WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. Intelligence 2001; 29: 131–67.

Klein NK, Hack M, Breslau N. Children who were very low birthweight: development and academic achievement at nine years of age. J Dev Behav Pediatr 1989; 10: 32–7.

Lucas A, Morley R, Cole TJ, Lister G, Leeson-Payne C. Breast milk and subsequent intelligence quotient in children born preterm. Lancet 1992; 339: 261–4.

McCall RB, Appelbaum MI, Hogarty PS. Developmental changes in mental performance. Monogr Soc Res Child Dev 1973; 38: 1–83.

Mortensen EL, Kleven, M. A WAIS longitudinal study of cognitive development during the life span from ages 50 to 70. Dev Neuropsychol 1993; 9: 115–30.

Mugler JP 3rd, Brookeman JR. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magn Reson Med 1990; 15: 152–7.

Mulhern RK, Palmer SL, Reddick WE, Glass JO, Kun LE, Taylor J, et al. Risks of young age for selected neurocognitive deficits in medulloblastoma are associated with white matter loss. J Clin Oncol 2001; 19: 472–9.

O'Brien F, Roth S, Stewart A, Rifkin L, Rushe T, Townsend J, et al. Neurodevelopmental progress into adolescence of very preterm infants.

Presented at the Annual Meeting of the European Society Paediatric Research, Rhodes; 2000.

Owens WA. Age and mental abilities: a second adult follow-up. J Educ Psychol 1966; 57: 311–25.

Peterson BS, Vohr B, Staib LH, Cannistraci CJ, Dolberg A, Schneider KC, et al. Regional brain volume abnormalities and long-term cognitive outcome in preterm infants. J Am Med Assoc 2000; 284: 1939–47.

Reiss AL, Abrams MT, Singer HS, Ross JL, Denckla MB. Brain development, gender and IQ in children. Brain 1996; 119: 1763–74.

Saigal S, Szatmari P, Rosenbaum P, Campbell D, King S. Cognitive abilities and school performance of extremely low birth weight children and matched term control children at age 8 years: a regional study. J Pediatr 1990; 117: 409–16.

Saigal S, Hoult LA, Streiner DL, Stoskopf BL, Rosenbaum PL. School difficulties at adolescence in a regional cohort of children who were extremely low birth weight. Pediatrics 2000; 105: 325–31.

Salmond CH, Ashburner J, Vargha-Khadem F, Gadian DG, Friston KJ. Detecting bilateral abnormalities with voxel-based morhometry. Hum Brain Mapp 2000; 11: 223–32.

Sameroff AJ, Seifer R, Baldwin A, Baldwin C. Stability of intelligence from preschool to adolescence: the influence of social and family risk factors. Child Dev 1993; 64: 80–97.

Sattler JM. Assessment of children. 3rd edn. San Diego: Jerome M. Sattler Publishers; 1992.

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. Nature Neurosci 2003; 6: 309–15.

Thompson AP, Molly K. The stability of WAIS-R IQ for 16-year-old students retested after 3 and 8 months. J Clin Psychol 1993; 49: 891–8.

Thompson PM, Cannon T D, Narr KL, van Erp T, Poutanen V-L, Huttenen M, et al. Genetic influences on brain structure. Nature Neurosci 2001; 4: 1253–8.

Van Paesschen W, Connelly A, King MD, Jackson GD, Duncan JS. The spectrum of hippocampal sclerosis: a quantitative magnetic resonance imaging study. Ann Neurol 1997; 41: 41–51.

Wechsler D. Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation; 1974.

Wechsler D. Wechsler Intelligence Scale for Children. 3rd edn. London: Psychological Corporation; 1992.

Wilke M, Sohn J-H, Byars AW, Holland SK. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. Neuroimage 2003; 20: 2002–15.

Wolke D, Meyer R. Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: the Bavarian Longitudinal Study. Dev Med Child Neurol 1999; 41: 94–109.

Wright IC, McGuire PK, Poline JB, Travere JM, Murray RM, Frith CD, et al. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. Neuroimage 1995; 2: 244–52.

Yates AJ. The validity of some psychological tests of brain damage. Psychol Bull 1954; 51: 359–79.

Zigler E, Balla D, Hodapp R. On the definition and classification of mental retardation. Am J Ment Defic 1984; 89: 215–30.