CORRECTIONS & AMENDMENTS

Addendum: Verbal and non-verbal intelligence changes in the teenage brain

Sue Ramsden, Fiona M. Richardson, Goulven Josse, Michael S. C. Thomas, Caroline Ellis, Clare Shakeshaft, Mohamed L. Seghier & Cathy J. Price

We measured[1] changes in intelligence quotient (IQ) between time 1 and time 2 in teenage subjects and searched their brains for regions where changes in IQ predicted changes in grey matter density (GMD). We found highly significant effects in two localized brain regions, after correcting for multiple comparisons across the whole brain. This provided an unbiased inference that longitudinal changes in IQ were meaningful rather than attributable to measurement error. In *post hoc* analyses, we quantified the (standardized) effect sizes[2] by reporting that 20% of the variance in verbal IQ (VIQ) at time 2 and 13% of the variance in performance IQ (PIQ) at time 2 could be explained by changes in GMD at the most significant voxel in the regions identified by the whole brain analyses. These (in-sample) effect sizes pertain to the sample studied and should not be confused with out-of-sample predictions[3]: that is, IQ predictability given new or independent subjects. Out-of-sample predictions finesse the inherent sampling bias of (in-sample) effect sizes—known in neuroimaging as 'the non-independence problem'.

Here we report out-of-sample estimates of effect size using a split-half procedure (and take the opportunity to compare VIQ and PIQ effects by including both in the same model). This involved splitting our sample into two groups (group 1 with $n = 17$ and group 2 with $n = 16$). One group was used to select voxels in which IQ changes predict GMD changes and the other group was used to predict IQ change from GMD change in these voxels. We found that GMD changes in group 2 (in voxels selected using group 1) predicted 16% of time-2 VIQ and 11% of time-2 PIQ. Conversely, GMD changes in group 1 (in voxels selected using group 2) predicted 16% of time-2 VIQ and 15% of time-2 PIQ (see Fig. 1 for details). These out-of-sample predictions are consistent with our original effect sizes. However, the split-half procedure is one of several procedures we could have used: an alternative approach—that minimizes type II (false negative) errors during voxel selection—is based on 'leave one out' procedures[4] and provides unbiased out-of-sample predictions of IQ change from GMD for each subject, in voxels identified in the other subjects. We will report this analysis elsewhere.

1. Ramsden, S. *et al.* Verbal and non-verbal intelligence changes in the teenage brain. *Nature* **479**, 113–116 (2011).
2. Wilkinson, L. APA task force on statistical inference statistical methods in psychology journals: guidelines and explanations. *Am. Psychol.* **54**, 594–604 (1999).
3. Leeb, H. Evaluation and selection of models for out-of-sample prediction when the sample size is small relative to the complexity of the data-generating process. *Bernoulli* **14**, 661–690 (2008).
4. Martens, H. A. & Dardenne, P. Validation and verification of regression in small data sets. *Chemom. Intell. Lab. Syst.* **44**, 99–121 (1998).
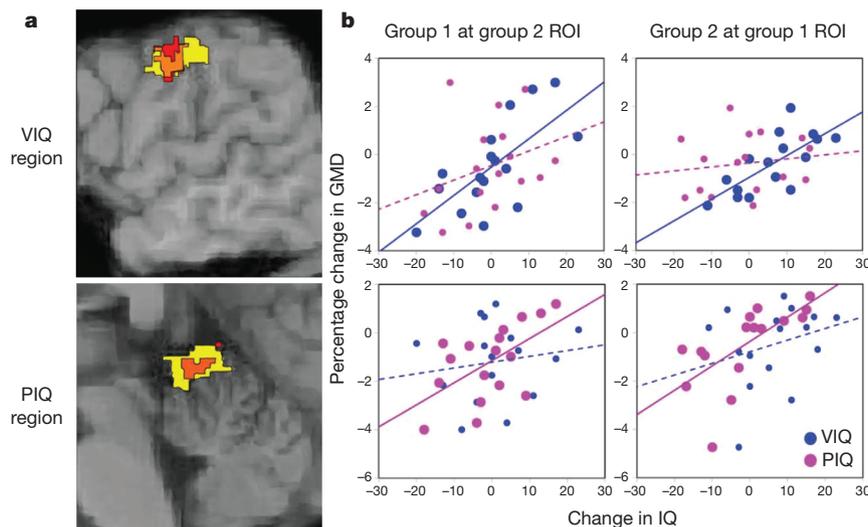

Figure 1 | Results of the split-half analysis. a, Brain images showing regions where IQ change predicted GMD change in group 1 (red) and group 2 (yellow)—in the left motor/premotor cortex for VIQ and in the anterior cerebellum for PIQ. Orange indicates an overlap of red and yellow. The criteria for selecting these voxels requires that 100 contiguous voxels (or more) survived an uncorrected threshold of $P < 0.01$ (VIQ) or $P < 0.05$ (PIQ) and a difference between VIQ and PIQ at $P < 0.05$ (uncorrected). Using these criteria, the only overlapping voxels, in a whole-brain analysis, are shown in orange above. No other (non-overlapping) effects survived these criteria in the slices shown. **b**, Plots of GMD change against IQ change for group 1 averaged across all voxels in the group 2 region of interest (yellow and orange areas in the left-hand panel) and for group 2 averaged across all voxels in the group 1 region of interest (red and orange areas in the left-hand panel). The solid line is the significant regression slope and the dashed line is the non-significant regression slope. The *P* values for the difference in regression slopes for VIQ and PIQ pooled over both groups (that is, after pooling the unbiased data points from the group 1 and group 2 regressions above) were $P = 0.009$ ($t = 2.5$; one-tailed) in the VIQ region (above) and $P = 0.04$ ($t = 1.9$, one-tailed) in the PIQ region (below). The analyses used to identify voxels (left) and to quantify the data (right) were identical to those in the original paper[1] except that each step included only half the data (and we combined VIQ and PIQ in the same regression model).