

# Interaction of homocysteine and conventional predisposing factors on risk of ischaemic stroke in young people: consistency in phenotype-disease analysis and genotype-disease analysis

A Pezzini, M Grassi, E Del Zotto, D Assanelli, S Archetti, R Negrini, L Caimi, A Padovani



See Editorial Commentary, p 1103

*J Neurol Neurosurg Psychiatry* 2006;**77**:1150–1156. doi: 10.1136/jnnp.2005.076083

See end of article for authors' affiliations

Correspondence to:  
Alessandro Pezzini,  
Clinica Neurologica,  
Università degli Studi di  
Brescia, Ple Spedali Civili,  
1, 25100 Brescia, Italy;  
ale\_pezzini@hotmail.com

Received 16 July 2005  
Revised version received  
8 March 2006  
Accepted 7 April 2006  
Published Online First  
19 April 2006

**Background and objectives:** Whether the association between mild hyperhomocysteinaemia and ischaemic stroke is the consequence of a predisposing genetic background or is due to the confounding influence of established predisposing factors remains to be determined.

**Methods:** Plasma total homocysteine (tHcy) concentration and the distribution of the C677T genotypes of the methylenetetrahydrofolate reductase gene (*MTHFR*) were compared in 174 consecutive patients with stroke aged <45 years and 155 age and sex-matched controls. The effect of conventional risk factors on the relationship between phenotype-disease and genotype-disease was analysed by two-way and three-way interaction analysis and by the classification and regression trees (CART) model.

**Results:** tHcy concentrations were markedly higher in patients with ischaemic stroke (median 11.9  $\mu\text{mol/l}$ , range 2.0–94.0) than in controls (median 9.8  $\mu\text{mol/l}$ , range 4.7–49.6). An increased risk was also associated with the TT677 genotype (odds ratio (OR) 1.98; 95% confidence interval (CI) 1.04 to 3.78) and with the T allele (1.40; 95% 1.03 to 1.92) of the *MTHFR* gene. A differential effect of Hcy levels on risk of stroke was observed according to the distribution of environmental-behavioural risk factors, with a stronger influence in the subcategory of people with hypertension and smokers (OR 24.8; 95% CI 3.15 to 196). A comparable environmental-dependent TT677 *MTHFR* genotype-stroke association was observed in the genotype-disease analysis.

**Conclusions:** A consistency of phenotype-disease analysis and genotype-disease analysis is indicated by analysing specific subcategories of patients, defined by the distribution of established risk factors. The assumption that the Hcy-stroke relationship is unlikely due to a reverse-causality bias is indirectly supported by our data.

Epidemiological evidence indicates that mild hyperhomocysteinaemia is independently associated with an increased risk of cerebral ischaemia.<sup>1–3</sup> The magnitude of the influence on stroke risk, however, seems to be modest in comparison to that of classic risk factors, such as arterial hypertension and cigarette smoking. A positive correlation between plasma total homocysteine (tHcy) levels and many of these factors has been shown repeatedly in previous reports.<sup>4–5</sup> Hcy concentration is also higher in people with pre-existing atherosclerosis than in those without.<sup>5</sup> On the basis of these observations, the possibility that the Hcy-stroke relationship may be subject to confounding, and that hyperhomocysteinaemia could be the consequence of the biological effect of these factors, rather than a cause of the disease, cannot be theoretically ruled out. Such a hypothesis is also supported by the fact that the association between the TT677 genotype in the coding region of the gene for methylenetetrahydrofolate reductase (*MTHFR*), the strongest genetic determinant of mild hyperhomocysteinaemia, and stroke risk has not been definitively proved.<sup>5</sup> In the largest meta-analysis to date of studies examining the association between *MTHFR* and stroke, Casas and colleagues<sup>1</sup> found, among people who are homozygous for the T allele, an increase in the risk of stroke close to that predicted by raised tHcy levels, further supporting the hypothesis of a causal relationship between Hcy and cerebral ischaemia. Whether the coexistence of established predisposing factors may have

confounded the results remains to be determined. The analysis of younger age groups provides a unique opportunity in this regard. Genetic factors may be expressed more at a young age, because a family history of stroke is more common<sup>6</sup> and physiological-behavioural risk factors are expressed less and, when present, act for a shorter time.<sup>7</sup> Hence, genetic effects are often more prominent in the young, as cumulative environmental factors have not had the time to substantially modify phenotype. For these reasons, the confounding effect of classic risk factors and that of established vascular diseases on the relationship between Hcy and stroke is likely to be modest at a young age.

We therefore carried out the present analysis of our hospital-based registry of young patients with stroke, aged 18–45 years, to investigate the contribution of Hcy and the C677T *MTHFR* polymorphisms on the risk of cerebral ischaemia and their role in predicting stroke risk on the basis of conventional predisposing factors, by using specific models to detect interactive effects.

## PARTICIPANTS AND METHODS

### Study participants

This study is part of an ongoing research programme aimed at evaluating the influence of gene-environment interactions

**Abbreviations:** CART, classification and regression trees; *MTHFR*, methylenetetrahydrofolate reductase gene; tHcy, total homocysteine

in the development of ischaemic cerebrovascular disease in young adults. Study design, diagnostic and laboratory procedures are Described in detail previously.<sup>8,9</sup> Briefly, consecutive, unrelated patients, aged <45 years, admitted to the Department of Neurology, University of Brescia, Brescia, Italy with their first-ever acute ischaemic stroke, were entered into the group of cases. On completion of the diagnostic investigation, cases were subtyped into four major aetiological categories, according to a classification based on the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria, accommodated and validated for the aetiology of stroke in the young:<sup>10</sup>

- Atherosclerotic vasculopathy: cerebral infarction caused by large-vessel atherosclerotic vasculopathy or small-vessel disease
- Cervical artery dissection
- Cardiac or transcerebral embolism: also including cases with a cardioembolic source in combination with a proved thrombophilic disorder
- Other: cerebral infarction that did not meet the criteria for one of the categories outlined above.

Participants from the staff of our hospital with no known history of vascular disease, matched to the patients by sex and age in 3-year bands, served as controls. Both patients and controls were white people, and were from the same geographical location and social status. For the purpose of this study, demographic data and history of hypertension, diabetes mellitus, cigarette smoking and hypercholesterolaemia were obtained from each participant.<sup>8</sup> Hypertension was considered to be present if systolic blood pressure was >160 mm Hg and diastolic pressure >95 mm Hg in two separate measurements after the acute phase, or if the participant was receiving treatment for hypertension before recruitment. The diagnosis of diabetes mellitus was established according to World Health Organization criteria.<sup>11</sup> Cigarette smokers were categorised as current smokers or non-smokers (the non-smokers included former smokers who had quit smoking at least 6 months before the study). Hypercholesterolaemia was considered present if serum cholesterol concentrations were >5.69 mmol/l or if the participant was under treatment with cholesterol-lowering drugs. Participants with known malignant diseases, hypothyroidism or chronic renal failure, or taking anti-convulsants, multivitamins, methotrexate or nitrous oxide were excluded from the final biochemical analysis, because of the potential influence of these variables on plasma tHcy concentrations. Unwillingness to participate was a further criteria for exclusion. The study was designed and carried out observing the ethical principles established by the local Institutional Guidelines on Clinical Investigation. All participants provided written informed consent.

### Biochemical and genetic parameters

For the determination of biochemical markers, venous blood samples were taken in the early morning (before 07:00) after overnight fasting. For patients with stroke, tHcy levels were measured at a follow-up evaluation >2 months after the acute event. tHcy levels were measured photometrically after separation with a reversed-phase column. The interassay and intra-assay coefficients of variation for this assay were <7%. Hyperhomocysteinaemia was defined as fasting plasma tHcy concentrations >12 µmol/l.<sup>12</sup> Folate and vitamin B<sub>12</sub> concentrations were determined in heparinised plasma by routine hospital assays. By using a standard DNA extraction method, genomic DNA was isolated from samples and kept frozen at -20°C whole blood anticoagulated with EDTA.

*MTHFR* genotypes were determined according to the method of Frosst and coworkers,<sup>13</sup> with PCR amplification and restriction digestion with *HinfI* to distinguish mutant from wild-type allele.

### Statistical analyses

Data were analysed with the SPSS V.11.1 software. Results are presented as frequencies (percentages) and mean (SD). Odds ratio (OR) and mean differences (MD) were used to compare frequencies and means, respectively, between patients and controls. A value of  $p < 0.05$  on a two-sided test was considered to be significant. As plasma tHcy concentrations are positively skewed, tHcy data were log transformed. Therefore, the mean values presented in this report are log means or geometric means in original scale. Significance testing was carried out by computing ORs and mean differences with 95% confidence intervals (CIs).

To explore the hypothesis that plasma tHcy concentrations and the *C677T MTHFR* genotypes may have a differential influence on stroke risk according to the presence or absence of traditional risk factors, we tabulated the groups of patients and controls into subcategories (yes or no on the basis of the status of exposed or non-exposed to conventional predisposing conditions) and compared the ORs for stroke according to the Hcy status (hyperhomocysteinaemic *v* non-hyperhomocysteinaemic) in each subcategory. The same cross-tabulation was carried out to test the effect of the *C677T MTHFR* genotypes (*TT* and *CT*, respectively, *v* *CC*). Diabetes mellitus was not entered into the explorative analysis, owing to the low frequency of this condition in this series.

The two-way and three-way interaction hypotheses were ascertained by the likelihood ratio testing<sup>14</sup> and the ORs (and 95% CIs) obtained by logistic regression modelling of case status on the indicator variables (1, exposed and 0, non-exposed) of predictor subcategories. For example, in the 2×2×2×2 table of smoke×hypertension×hyperhomocysteinaemia×case status, we consider 7 (=2<sup>3</sup>-1) indicator variables of the possible combinations of smoke×hypertension×hyperhomocysteinaemia, with non-smokers, non-hypertensive and non-hyperhomocysteinaemic as the reference category.

Finally, we processed the classification and regression tree (CART) models<sup>15</sup> for the identification of subgroups at risk from six potential predictor variables (smoke, hypertension, hypercholesterolaemia, vitamin B<sub>12</sub>, folate, hyperhomocysteinaemia or *TT MTHFR* genotype). CART constructs a tree by using a sequence of binary splits to stratify data into risk subgroups on which to discriminate the outcome of interest—that is, patients or controls). At each subdivision, CART selects the single variable that produces two subgroups of the greatest purity. A subgroup is pure when it includes only patients or only controls, and is impure when both patients and controls are represented according to their proportion in the total dataset. The strength of CART with respect to a traditional parametric technique, such as logistic regression, is its ability to detect high-level interactions among the risk factors. It forms subgroups, or Terminal nodes, that have homogeneous effects with respect to the outcome. The commercial SPAD Data Mining software ([http://www.decision.com/SPAD\\_Presentation.html](http://www.decision.com/SPAD_Presentation.html)) was used in these analyses. To set up these models, the Gini Index was used as a splitting criterion, with a minimum Terminal node size of 10. A 10-fold cross-validation was then applied to prune the tree to its final size. The ORs (and 95% CIs) for the final classification tree were computed by logistic regression modelling of case status on the indicator variables (1, yes and 0, no) of the Terminal nodes identified in the classification tree, with the last Terminal node as the reference category.

## RESULTS

Of the 329 participants, 174 were consecutive patients with ischaemic stroke and 155 were controls. A cardioembolic aetiology was presumed in 54 (31.0%) patients, cervical artery dissection in 28 (16.1%) and atherosclerotic vasculopathy in 40 (23.0%; large-vessel atherosclerosis in 24 and small-vessel disease in 16, respectively), whereas 52 (29.9%) patients met the criteria for other aetiological categories. Table 1 shows the baseline characteristics of the study group. As expected, the prevalence of both hypertension and cigarette smoking was considerably higher in patients than in controls.

### Homocysteine levels

Twelve participants met either of the exclusion criteria reported earlier (3 had hypothyroidism, 1 had chronic renal failure, 2 were taking anticonvulsants and 6 were taking multivitamins). tHcy levels were not available in 22 participants because of technical reasons. Consequently, plasma tHcy measurements from a total of 161 patients and 134 controls were entered in the final analysis.

The fasting geometric mean plasma tHcy concentrations were higher in patients (11.9  $\mu\text{mol/l}$ , range 2.0–94.0) than in controls (9.8  $\mu\text{mol/l}$ , range 4.7–49.6). As expected, higher tHcy levels were observed in *TT677 MTHFR* carriers than in *CT* carriers and in *CC* carriers (not shown). A positive relationship between tHcy concentration and the risk of cerebral ischaemia was observed, with a mean difference of the log-transformed tHcy concentration of +0.20 (95% CI 0.08 to 0.31) between patients and controls. Similar results were obtained comparing the prevalence of participants whose plasma tHcy levels were above the cut-off value in each group (patients 76/161 *v* controls 36/134; OR 2.34; 95% CI 1.49 to 3.98). We found no noticeable differences in folate levels and vitamin B<sub>12</sub> levels between the two groups. We also examined the association between tHcy levels and established risk factors for stroke. Plasma tHcy concentrations increased markedly with age and serum cholesterol levels, and were positively associated with male sex, hypertension and smoking, both in patients and controls (not shown).

### Distribution of C677T MTHFR genotypes

Distribution of genotypes did not differ from that predicted by the Hardy–Weinberg equilibrium in patients, controls and the whole population (allelic correlations: 0.005, –0.002 and 0.009, respectively). A significant increase in stroke risk was associated with the *TT677 MTHFR* genotype when compared with the *CC* genotype (OR 1.98; 95% CI 1.04 to 3.78). The frequency of the *T* allele was also significantly higher in the group of patients (OR 1.40; 95% CI 1.03 to 1.92).

### Two-way and three-way interaction analyses

The effect of hyperhomocysteinaemia (and *TT677 MTHFR* genotype) on stroke risk turned out to be independent of hypercholesterolaemia, vitamin B<sub>12</sub> or folate (data not shown), whereas interactions were observed with smoke and hypertension.

Hyperhomocysteinaemia (in non-hypertensive patients and non-smokers), hypertension (in non-homocysteinaemic patients and non-smokers), and smokers (in non-hypertensive and non-homocysteinaemic patients) showed a statistically significant increase in stroke risk of about 2.5-fold (OR 2.45; 95% CI 1.24 to 4.85), about six-fold (OR 6.21; 95% CI 1.23 to 31.4), and about three-fold (OR 2.96; 95% CI 1.52 to 5.76), respectively (table 2).

The stroke risk related to hyperhomocysteinaemia in the subcategory of non-smokers and non-hypertensive patients decreased to 0.5-fold in the subcategory of hypertensive patients and non-smokers, and to 1.5-fold in the subcategory of non-hypertensive patients and smokers, although the discrepancies in ORs were not statistically significant (likelihood ratio testing (LRT) = 2.113, *df* = 1, *p* = 0.146; and LRT = 0.686, *df* = 1, *p* = 0.407, respectively). By contrast, a significant discrepancy in ORs was observed in hypertensive patients and smokers, with an increase in stroke risk from about 2.5-fold to 19-fold (LRT = 6.048, *df* = 1, *p* = 0.014).

The same findings were obtained when the hypotheses of two-way and three-way interactions among smokers, hypertensive patients and *TT677 MTHFR* genotype were tested (LRT = 1.481, *df* = 1, *p* = 0.224 for *MTHFR*–hypertensive patients; LRT = 2.497, *df* = 1, *p* = 0.114 for *MTHFR*–smokers,

**Table 1** Characteristics of the study group according to conventional risk factors, homocysteine levels, genotype distribution and allele frequencies

| Characteristics                          | Patients, n = 174 |               | Controls, n = 155     |                |
|--|-------------------|---------------|-----------------------|----------------|
|  | n (%)             | n (%)         | Crude OR              | 95% CI         |
| Sex (male)                               | 89 (51.1)         | 83 (53.5)     | 1.01                  | 0.71 to 1.70   |
| Current smokers                          | 85 (48.9)         | 40 (25.8)     | 2.78                  | 1.74 to 4.43   |
| Hypertension                             | 30 (17.2)         | 10 (6.5)      | 3.02                  | 1.42 to 6.41   |
| Diabetes mellitus                        | 6 (3.5)           | 5 (3.2)       | 1.07                  | 0.32 to 3.58   |
| Hypercholesterolaemia                    | 48 (27.6)         | 30 (19.3)     | 1.59                  | 0.94 to 2.67   |
| Hyperhomocysteinaemia*                   | 76 (47.2)         | 36 (26.9)     | 2.34                  | 1.49 to 3.98   |
| <i>C677T MTHFR</i> genotype distribution |                   |               |                       |                |
| <i>CC</i>                                | 51 (29.3)         | 60 (38.7)     | 1                     |                |
| <i>CT</i>                                | 86 (49.4)         | 73 (47.1)     | 1.39                  | 0.85 to 2.25   |
| <i>TT</i>                                | 37 (21.3)         | 22 (14.2)     | 1.98                  | 1.04 to 3.78   |
| <i>C677T MTHFR</i> allele frequency      |                   |               |                       |                |
| <i>C</i>                                 | 188 (54.0)        | 193 (62.3)    | 1                     |                |
| <i>T</i>                                 | 160 (46.0)        | 117 (37.7)    | 1.40                  | 1.03 to 1.92   |
|  | Mean (SD)         | Mean (SD)     | Crude mean difference | 95% CI         |
| Age (years)                              | 35.0 (7.5)        | 34.7 (5.9)    | +0.43                 | –1.04 to +1.90 |
| Homocysteine (log $\mu\text{mol/l}$ )*   | 2.48 (0.5)        | 2.28 (0.4)    | +0.20                 | +0.08 to +0.31 |
| Vitamin B <sub>12</sub> (pmol/l)*        | 395.3 (123.5)     | 393.4 (115.8) | +1.88                 | –25.7 to +29.5 |
| Folate (nmol/l)*                         | 6.47 (2.9)        | 6.61 (2.2)    | –0.14                 | –0.74 to +0.45 |

MTHFR, methylenetetrahydrofolate reductase gene.

\*In 161 patients and 134 controls.

**Table 2** Frequency distributions, ORs (95% CIs) and likelihood ratio testing (LRT) of plasma total homocysteine levels × hypertension × smoke × case status

|            | tHcy concentration (µmol/l) | Patients (n = 161) | Controls (n = 134) | Specific OR | 95% CI | Conditional OR | p Value (LRT) |           |
|------------|-----------------------------|--------------------|--------------------|-------------|--------|----------------|---------------|-----------|
| Non-smoker | Non-hypertensive            | <12                | 40                 | 71          | 1      | Reference      |               |           |
|            |                             | ≥12                | 29                 | 21          | 2.45   | 1.24 to 4.85   | 2.45          | Reference |
|            | Hypertensive                | <12                | 7                  | 2           | 6.21   | 1.23 to 31.4   |               |           |
|            |                             | ≥12                | 5                  | 3           | 2.96   | 0.67 to 13.0   | 0.48          | 0.146     |
| Smoker     | Non-hypertensive            | <12                | 35                 | 21          | 2.96   | 1.52 to 5.76   |               |           |
|            |                             | ≥12                | 28                 | 11          | 4.52   | 2.03 to 10.0   | 1.53          | 0.407     |
|            | Hypertensive                | <12                | 3                  | 4           | 1.33   | 0.28 to 6.25   |               |           |
|            |                             | ≥12                | 14                 | 1           | 24.8   | 3.15 to 196    | 18.7          | 0.014     |

and LRT = 4.273, df = 1, p = 0.039 for *MTHFR*–hypertensive–smokers) with comparable ORs (table 3) .

**CART**

Figures 1 and 2 show the selected CART models from six potential predictor variables (smoking, hypertension, hypercholesterolaemia, vitamin B<sub>12</sub>, folate, hyperhomocysteinaemia or *TT MTHFR* genotype).

The number and percentage of patients and controls are shown for each node. Four Terminal nodes were selected. These were a combination of three categorical variables: smoking, hypertension and plasma tHcy concentrations (fig 1) and smoking, hypertension and *MTHFR* genotypes (fig 2). The first split was according to the smoking status (Node 1). The subcategory of non-smokers was further split according to a positive–negative history of hypertension (Node 2). Finally, the subcategory of non-smokers and non-hypertensive patients was split, according to plasma tHcy concentrations ( ≤ 12 µmol/l v > 12 µmol/l; Node 3; fig 1) and the *MTHFR* status (*TT* v non-*TT*; Node 3; fig 2), respectively.

Participants included in the Terminal node 4—that is, non-smokers and non-hypertensive with tHcy levels ≤ 12 µmol/l (fig 1), and non-smokers and non-hypertensive patients with non-*TT MTHFR* genotype (fig 2), respectively—served as the reference group. Table 4 shows the ORs and 95% CI for Terminal nodes 1, 2 and 3 as obtained by logistic regression models, both unadjusted and adjusted for sex, age, and excluded predictors (hypercholesterolaemia, vitamin B<sub>12</sub> and folate). On the basis of the CART decision rule, the interactions among hyperhomocysteinaemia, hypertension and smoking were reduced to three hierarchical comparisons: (1) smokers v non-smokers (OR 3.84, 95% CI 2.22 to 6.65) in all the participants; (2) hypertensive v non-hypertensive

among non-smokers (OR 4.26, 95% CI 1.40 to 13.0); and (3) hyperhomocysteinaemic v non-hyperhomocysteinaemic (OR 2.45, 95% CI 1.24 to 4.85) in non-smokers and non-hypertensive patients. Similar results were observed for the *MTHFR* genotype or after adjustment for covariates (table 4). Overall, these findings are in line with those of the two-way and three-way interaction analyses.

**DISCUSSION**

The results of our study confirmed that raised plasma tHcy levels are associated with increased risk for ischaemic stroke as a whole. In this regard, our findings are consistent with those from the few previous studies investigating the role of this molecule as a predisposing condition for cerebral ischaemia among young adults.<sup>16–21</sup> The magnitude of the effect on disease risk was also similar to that observed in the largest meta-analysis to date, including 564 people aged <40 years.<sup>1</sup> As opposed to some of the previous reports, we found that the *TT677* genotype of the *MTHFR* gene is related to an increased risk of developing cerebral ischaemia, a finding that indirectly supports the prevailing hypothesis of a causal link between Hcy and stroke. Furthermore, our data provide evidence that the association between both Hcy and the *TT677* genotype of the *MTHFR* gene and ischaemic stroke is dependent on the coexistence of lifestyle, environmental and vascular risk factors such as blood pressure and smoking history, and that there is a similar effect of phenotype and genotype in different subcategories at risk.

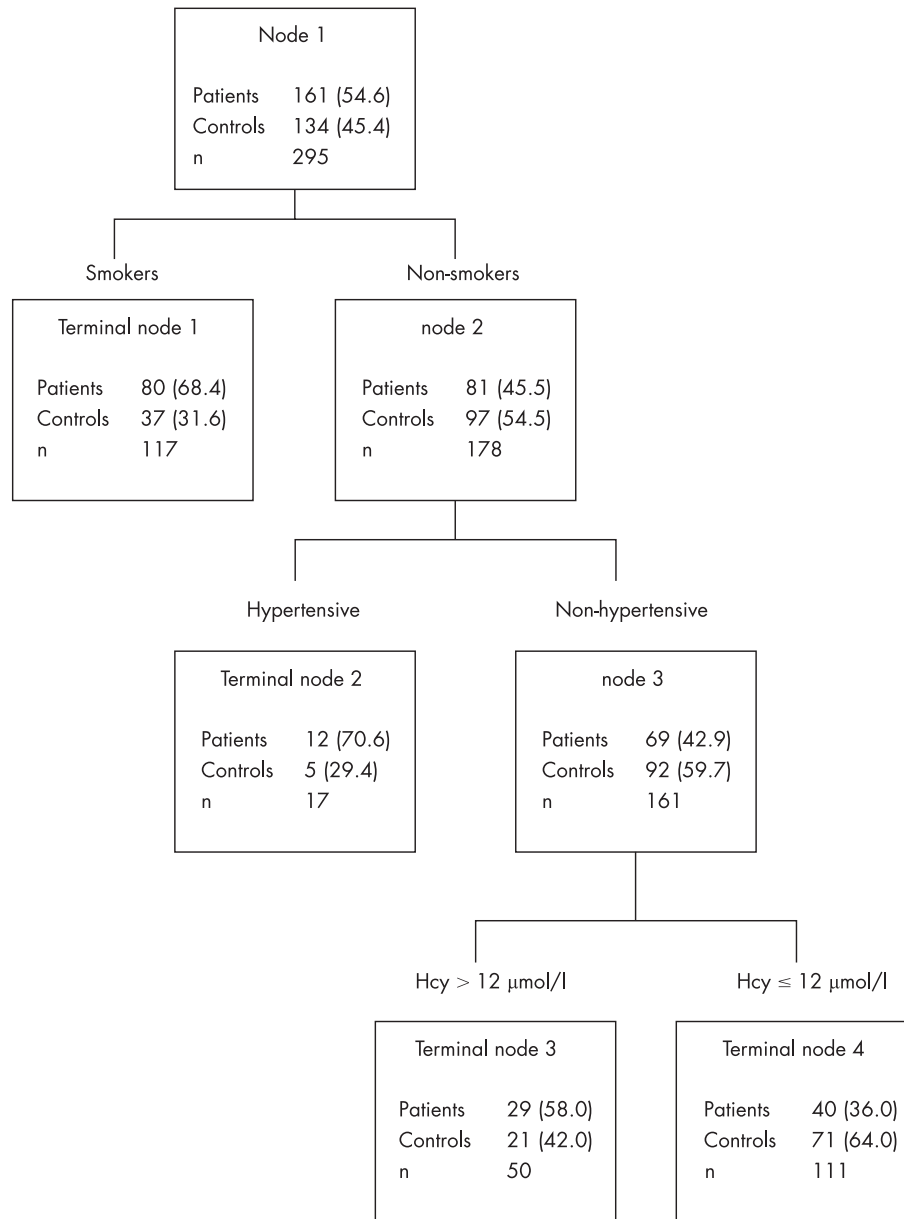
The confounding effect of environmental–behavioural variables has often been advocated in the relationship of homocysteine with cerebral ischaemia as a possible explanation for the reported inconsistency between the positive findings of phenotype-disease analyses and the negative findings of genotype-disease analyses. Actually, the influence

**Table 3** Frequency distributions, ORs (95% CIs) and likelihood ratio testing (LRT) of *C677T MTHFR* genotype × hypertension × smoke × case status

|            | Genotype         | Patients (n = 174)     | Controls (n = 159) | Specific ORs | 95% CI | Conditional ORs | p Value (LRT) |           |
|------------|------------------|------------------------|--------------------|--------------|--------|-----------------|---------------|-----------|
| Non-smoker | Non-hypertensive | <i>CC</i> or <i>CT</i> | 56                 | 96           | 1      | Reference       |               |           |
|            |                  | <i>TT</i>              | 20                 | 14           | 2.45   | 1.15 to 5.23    | 2.45          | Reference |
|            | Hypertensive     | <i>CC</i> or <i>CT</i> | 12                 | 4            | 5.14   | 1.58 to 16.7    |               |           |
|            |                  | <i>TT</i>              | 1                  | 1            | 1.71   | 0.11 to 28.0    | 0.33          | 0.224     |
| Smoker     | Non-hypertensive | <i>CC</i> or <i>CT</i> | 56                 | 28           | 3.43   | 1.96 to 6.01    |               |           |
|            |                  | <i>TT</i>              | 12                 | 7            | 2.94   | 1.09 to 7.90    | 0.86          | 0.114     |
|            | Hypertensive     | <i>CC</i> or <i>CT</i> | 12                 | 5            | 4.11   | 1.38 to 12.3    |               |           |
|            |                  | <i>TT</i>              | 5                  | 0            | 18.6   | 1.02 to 347     | 4.54          | 0.039     |

*MTHFR*, methylenetetrahydrofolate reductase gene.



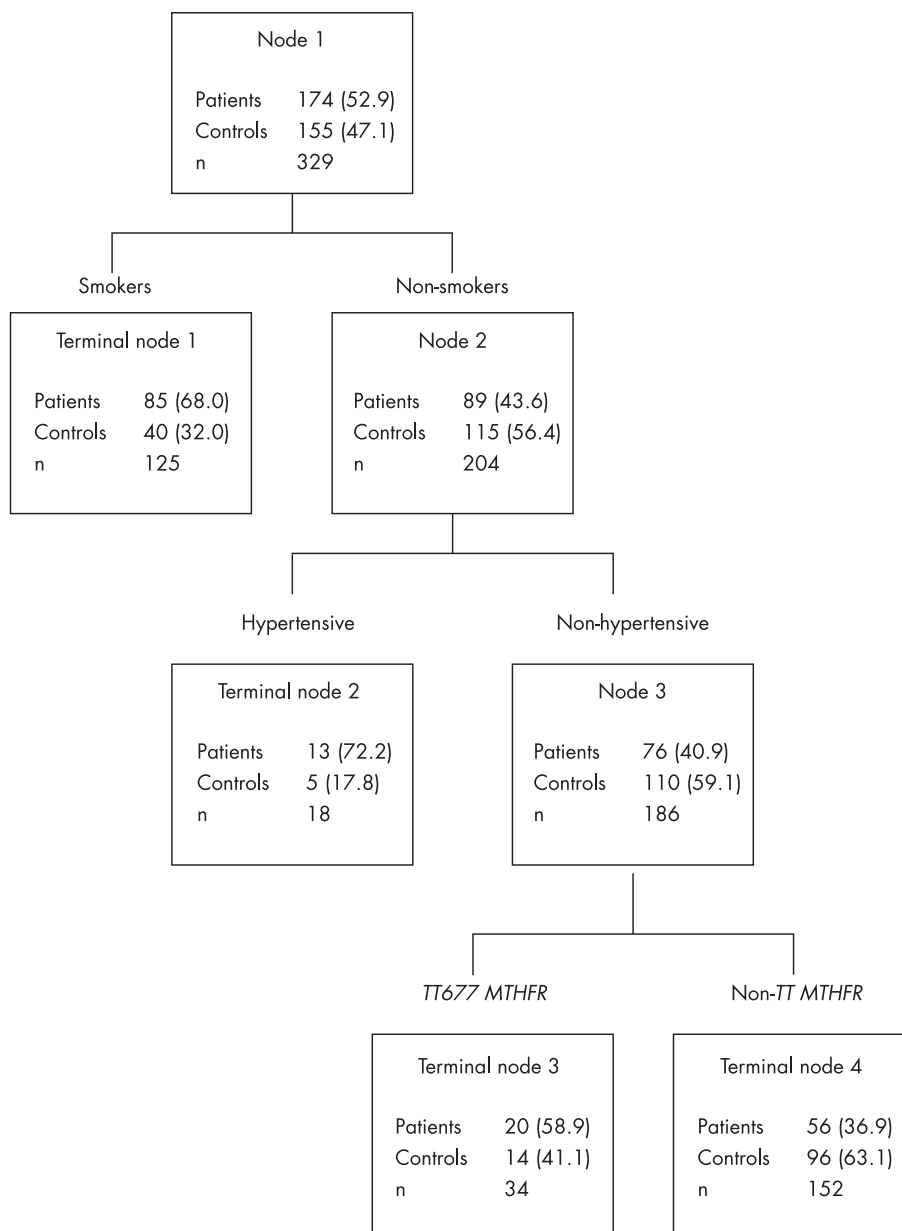


**Figure 1** Classification and regression tree model including plasma homocysteine (Hcy) levels. Values in parentheses are percentages.

of mild hyperhomocysteinaemia on stroke risk may be masked by the presence of other predictors, such as hypertension or smoking, thus leading to inconsistent associations. Although this may be true for a biochemical variable such as homocysteine, it should be noted that genotype is a fixed characteristic and is free from the confounding effect by other determinants of Hcy concentration, as shown by previous studies on coronary heart disease.<sup>22–23</sup> At least theoretically we cannot rule out the possibility that the *MTHFR* genotypes may affect behavioural–environmental factors by being in linkage disequilibrium with polymorphic variants that predispose people with a given genotype to unhealthy behaviours (such as smoking), or abnormal physiological risk factors for stroke (such as high blood pressure), and confound the finding. The potential influence of these interactions is noteworthy and should be taken into account when interpreting our findings. As opposed to the hypothesis of confounded results, however, we consider the hypothesis of a differential effect of

homocysteine on stroke risk according to the coexistence of other predisposing conditions more likely. Furthermore, and more importantly, consistency in the results from the phenotype–disease analysis and the genotype–disease analysis of our data indirectly suggests that the relationship between Hcy and stroke is not likely to be due to a reverse-causality. In fact, on the basis of the paradigm of mendelian randomisation,<sup>24</sup> if Hcy is a causal factor, the *TT677 MTHFR* genotype should be related to disease risk to the extent predicted by its influence on homocysteine levels. When confirmed, the practical implication would be that the intervention of increasing folate intake at a population level could lead to a decrease in risk for ischaemic stroke.

The results of our study apply to young people and may not be generalised to other age groups at present. Similar findings were, however, observed in several previous studies on older people, which showed that the association of homocysteine with carotid atherosclerosis and stroke risk may be dependent on coexistent risk factors.<sup>5–25</sup> These



**Figure 2** Classification and regression tree model including the *TT677 MTHFR* genotype of the methylenetetrahydrofolate (*MTHFR*) gene. Values in parentheses are percentages.

**Table 4** Crude and adjusted ORs (95% CIs) of the classification and regression tree (CART) Terminal node groups, on the basis of smoking, hypertension and homocysteine (or the *TT677 MTHFR* genotype)

| CART models  | Crude OR | 95% CI       | Adjusted OR* | 95% CI       |
|--|----------|--------------|--------------|--------------|
| Terminal nodes in fig 1  |          |              |              |              |
| Node 1: smokers  | 3.84     | 2.22 to 6.65 | 4.93         | 2.71 to 8.99 |
| Node 2: non-smokers and hypertensive                                       | 4.26     | 1.40 to 13.0 | 4.66         | 1.47 to 14.8 |
| Node 3: non-smokers, non-hypertensive and hyperhomocysteinaemic            | 2.45     | 1.24 to 4.85 | 2.94         | 1.43 to 6.06 |
| Node 4: non-smokers, non-hypertensive and non-hyperhomocysteinaemic        | 1        |              | 1            |              |
| Terminal nodes in fig 2  |          |              |              |              |
| Node 1: smokers  | 3.64     | 2.21 to 6.01 | 4.12         | 2.35 to 7.26 |
| Node 2: non-smokers and hypertensive                                       | 4.46     | 1.51 to 13.6 | 3.87         | 1.24 to 12.1 |
| Node 3: non-smokers, non-hypertensive and <i>TT677 MTHFR</i> genotype      | 2.44     | 1.15 to 5.23 | 2.44         | 1.10 to 5.46 |
| Node 4: non-smokers, non-hypertensive and non- <i>TT677 MTHFR</i> genotype | 1        |              | 1            |              |

\*After adjustment for sex, age and excluded predictors (hypercholesterolaemia, vitamin B<sub>12</sub> and folate).

observations further support the hypothesis of a differential genotype expression, according to the individual pattern of environmental and behavioural predisposing conditions.

The use of a powerful statistical approach, based on two different models (logistic regression and CART) to test for interactive effects, represents a strength of our study, as the same three-way interaction was signed by both the models. In this regard, CART offers advantages in that it may show up interactions among predictor variables that do not show strong main effects and it may help in the interpretation of high-level interactive effects that are identified by logistic regression.<sup>26</sup> As such, the CART method is particularly useful in analyses of gene–environment interactions of complex diseases.

A potential limitation may be theoretically inherent in the relatively small size of our sample. The estimated sample size for an association study, however, seems to be age dependent, according to the results of a recent analysis on the feasibility of genetic approaches to stroke in humans.<sup>27</sup> The younger the age of the study group, the lower the sample size required for a case–control study. On the basis of these estimates, we consider the size of our sample to be adequate to test the primary hypothesis under investigation. Our decision to analyse the effects of phenotype and genotype in specific subgroups (each including a limited number of participants) defined by the presence of traditional risk factors makes some of our findings statistically unstable. In this regard, confirmation is necessary and this can be reliably obtained in the setting of a prospective multicentre trial on young patients with stroke.

We also analysed the effects of genetic–environmental–behavioural exposures without taking into consideration the differential role of such interactions in each specific pathogenic subtype of stroke. Stratification by subtypes, however, reduces the statistical power and dilutes the estimate of effect towards the null. Finally, we cannot rule out the possibility that potential drawbacks of observational studies, including population stratification and pleiotropic effects of gene polymorphisms on more than one biological system, among others,<sup>28</sup> may be operant in our analysis.

In conclusion, our results support the assumption of a causal relationship among the *MTHFR* genotype, tHcy and stroke. In young adults aged <45 years, the increase in stroke risk associated with the *TT MTHFR* genotype-raised levels of tHcy seems to be influenced by the effect of conventional risk factors. While waiting for conclusive evidence from a powered randomised trial of multivitamin supplementation,<sup>29</sup> the practical implication of these findings may be that measuring tHcy levels (and screening for *MTHFR* genotype) is warranted in this specific group, as this can help identifying people at higher risk and may eventually result in individualised and effective long-term vitamin B treatment for prevention of stroke.

## ACKNOWLEDGEMENTS

We thank Dr Mauro Magoni and Dr Angelo Costa, Clinica Neurologica, Università degli Studi di Brescia, Brescia, Italy for their help in the recruitment of patients. We also thank all the people who participated in the study.

## Authors' affiliations

**A Pezzini, E Del Zotto, A Padovani**, Clinica Neurologica, Università degli Studi di Brescia, Brescia, Italy

**M Grassi**, Dipartimento di Scienze Sanitarie Applicate, Sezione di Statistica Medica ed Epidemiologia, Università degli Studi di Pavia, Pavia, Italy

**D Assanelli**, Dipartimento di Scienze Mediche e Chirurgiche, Università degli Studi di Brescia

**S Archetti**, III Laboratorio di Analisi, Biotecnologie, Università degli Studi di Brescia

**R Negrini**, Laboratorio di Microbiologia, Università degli Studi di Brescia

**L Caimi**, Dipartimento di Biochimica, Università degli Studi di Brescia

Competing interests: None declared.

## REFERENCES

- Casas PJ, Bautista LE, Smeeth L, et al. Homocysteine and stroke: evidence on a causal link from mendelian randomization. *Lancet* 2005;**365**:224–32.
- Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* 2002;**325**:1202–8.
- Kelly PJ, Rosand J, Kistler JP, et al. Homocysteine, MTHFR 677C>T polymorphism, and risk of ischemic stroke: results of a meta-analysis. *Neurology* 2002;**59**:529–36.
- Nygard O, Vollset SE, Refsum H, et al. Total plasma homocysteine and cardiovascular risk profile. The Hordaland Homocysteine Study. *JAMA* 1995;**274**:1526–33.
- Ueland PM, Refsum H, Beresford SAA, et al. The controversy over homocysteine and cardiovascular risk. *Am J Clin Nutr* 2000;**72**:324–32.
- Flossman E, Schulz UGR, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* 2004;**35**:212–27.
- Hassan A, Markus HS. Genetics and ischaemic stroke. *Brain* 2000;**123**:1784–812.
- Pezzini A, Del Zotto E, Archetti S, et al. Plasma homocysteine concentration, C677T *MTHFR* genotype and 844ins68bp *CBS* genotype in young adults with spontaneous cervical artery dissection and atherothrombotic stroke. *Stroke* 2002;**33**:664–9.
- Pezzini A, Del Zotto E, Magoni M, et al. Inherited thrombophilic disorders in young adults with ischemic stroke and patent foramen ovale. *Stroke* 2003;**34**:28–33.
- Johnson CJ, Kittner SJ, McCarter RJ, et al. Interrater reliability of an etiologic classification of ischemic stroke. *Stroke* 1995;**26**:46–51.
- WHO Study Group on Diabetes Mellitus. *World Health Organization technical report, series 727*. Geneva: World Health Organization, 1985.
- Ubbink JB, Delport R. Reference ranges for homocysteine concentrations. In: Robinson K, ed. *Homocysteine and vascular disease*. Dordrecht: Kluwer Academic, 2000:41–57.
- Frosst P, Blom HJ, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature Genetics* 1995;**10**:111–3.
- Agresti A. *Categorical data analysis*. 2nd edn. New Jersey: Wiley, 2002.
- Breiman L, Friedman JH, Olshen RA, et al. *Classification and regression trees*. Belmont, CA: Wadsworth, 1984.
- Soriente L, Coppola A, Madonna P, et al. Homozygous C677T mutation of the 5,10 methylenetetrahydrofolate reductase gene and hyperhomocysteinemia in Italian patients with a history of early-onset ischemic stroke. *Stroke* 1998;**29**:869–71.
- Kittner SJ, Giles WH, Macko RF, et al. Homocysteine and risk of cerebral infarction in a biracial population. The stroke prevention in young women. *Stroke* 1999;**30**:1554–60.
- Kristensen B, Malm J, Carlberg B, et al. Epidemiology and etiology of ischemic stroke in young adults aged 18 to 44 years in Northern Sweden. *Stroke* 1997;**28**:1702–9.
- Beccia M, Mele MC, Ferrari M, et al. Young stroke and basal and post-methionine load homocysteine and cysteine levels 1 year after the acute event: do plasma folates make the difference? *Eur J Neurol* 2004;**11**:269–75.
- Choon-Kiat TN, Venketasubramanian N, Saw SM, et al. Hyperhomocysteinemia and risk of ischemic stroke among young Asian adults. *Stroke* 2002;**33**:1956–62.
- Sobol AB, Bald E, Loba J. Fractions of total plasma homocysteine in patients with ischemic stroke before the age of 55 years. *Angiology* 2005;**56**:201–9.
- Klerk M, Verhoef P, Clarke R, et al. MTHFR 677C>T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA* 2002;**288**:2023–31.
- Russo GT, Friso S, Jacques PF, et al. Age and gender affect the relation between methylenetetrahydrofolate reductase C677T genotype and fasting plasma homocysteine concentrations in the Framingham Offspring Study Cohort. *J Nutr* 2003;**133**:3416–21.
- Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
- Brattstrom L, Wilken DE. Homocysteine and cardiovascular disease: cause or effect? *Am J Clin Nutr* 2000;**72**:315–23.
- Cook NR, Zee RYL, Ridker PM. Tree and spline based association analysis of gene-gene interaction models for ischemic stroke. *Stat Med* 2004;**23**:1439–53.
- Hassan A, Sham PC, Markus HS. Planning genetic studies in human stroke. Sample size estimates based on family history data. *Neurology* 2002;**58**:1483–8.
- Hassan A, Markus HS. Polygenic ischemic stroke including new genetic and statistical approaches. In: Markus HS, ed. *Stroke genetics*. Oxford, UK: Blackwell Science, 2003:165–95.
- The VITATOPS Trial Study Group. The VITATOPS (Vitamin To Prevent Stroke) trial: rational and design of an international, large, simple, randomised trial of homocysteine-lowering multivitamin therapy in patients with recent transient ischemic attack or stroke. *Cerebrovasc Dis* 2002;**13**:120–6.