Iron deficiency and educational achievement in Thailand

Ernesto Pollitt, Phongjan Hathirat, Nittaya J Kotchabhakdi, Lavon Missell, and Aree Valyasevi

ABSTRACT This double-blind clinical trial was conducted in Thailand to assess the impact of iron treatment on the IQ and educational attainment of 1358 9–11-yr-old children. The children were classified into one of three groups: iron replete, iron depleted, and iron-deficient anemic. The Raven Progressive Matrices was used to measure IQ. A Thai language and a math test were administered to assess school attainment. A 50-mg/d tablet of ferrous sulphate was given for 2 wk and a 100 mg/d tablet, for 14 wk. An anthelmintic drug was given on the day of the blood test before treatment and 3 mo after the intervention started. There is evidence of a positive association between iron status and IQ and a language school achievement test but there is no support for the internal validity of the hypothesis that this association is causal. Am J Clin Nutr 1989;50:687–97.

KEY WORDS Iron deficiency, schooling, cognition, Thailand

Introduction

It has been consistently demonstrated that the scores on the Bayley Scales of Mental and Motor Development among iron-deficient anemic (IDA) infants are lower than those among iron-replete infants (1–4). In some studies significant increments in scale scores were observed among IDA infants after iron treatment (3–5); however in others (1–2), these improvements were not observed. Problems of construct validity notwithstanding (6, 7), the large magnitude of the increments in the developmental-scale scores that have been observed in the IDA infants treated with iron conjointly with the robustness of the studies that reported such findings suggest that IDA has adverse effects on adaptive behaviors. One critical issue is whether these effects are dependent on the developmental stage of the organism and may be observed only during periods when brain growth reaches its peak of growth velocity. Such effects might not be observed among older children who are more resilient to stress.

Results from a double-blind experimental study in Indonesia showed that ideniciency (ID) affected the school behavior of 9–11-yr-old IDA children in grades three to five (6). They obtained lower scores in a standardized educational achievement test than did iron-replete children. After 12 wk of iron treatment the anemic children improved ($p < 0.05$) their test scores whereas the anemic children who received a placebo failed to show such an improvement. However, the performance of the anemic children treated with iron was not large enough to catch up with the iron-replete children.

These findings on school children strengthen the previous conclusion, based on studies of infants, of a causal relationship between iron and behavior and indicate that this effect is not restricted to a period of high developmental vulnerability such as infancy. Learning in school is apparently at risk under conditions of ID.

The internal and external validity (7) of the conclusions to be derived from the study in Semarang (6) would be bolstered if the findings were replicated in a different ecological setting, population, and cultural milieu. It would also solidify the inferences made from the infancy data that ID has a deleterious developmental effect. An additional justification for the use of school achievement measures was the need to assess whether ID does indeed affect the process of learning in formal schooling, which is critical to the productivity of an individual in modern society.

This is a double-blind clinical trial conducted in 16 elementary schools in Chon Buri, Thailand, to assess the impact of iron treatment on the intelligence quotient and educational attainment of 9–11-yr-old children. The experimental interventions (iron and placebo) were implemented before the iron status of the subjects was determined. The protocol for this study was approved by the committees for the use of humans as subjects of the Uni-
versity of California, Davis, and of the Mahidol University, Bangkok.

**Methods**

**Site of the study and sample**

Chon Buri province, on the east coast of the eastern region of Thailand, ~85 km from Bangkok, is an agricultural, rice-producing area. The per capita gross provincial product in 1980 was of 37 000 baht, or ~$1480 US (8).

Sixteen elementary schools in five umphurs, or geographic sectors, of Chon Buri were selected on the basis of the following criteria: 1) nonmalarial area; 2) ≥ 150 children per school; and 3) access to the main roads. The schools are located 26–61 km from the city of Chon Buri. Two thousand two hundred and sixty-eight children, ~23% of the total number of children enrolled in grades three to five in the entire province according to 1980 statistics (8), were screened. Criteria for case inclusion were 1) absence of AE Bart or hemoglobin (Hb) H disease; 2) absence of abnormal Hb E (hemoglobinopathy) and 3) aged 108–144 mo.

**Measurements**

**Physical.** Skin infections, upper respiratory infections, urinary tract infections, ear infections, and gingivitis due to dental caries were treated and recorded but these conditions were not criteria for exclusion.

**Hematology.** Fifteen milliliters of venous blood from the arm was drawn from each child by disposable plastic needles and syringes and placed into the EDTA tube for blood cell count and determination of thalassemia and/or hemoglobinopathies. A drop of blood was placed and sealed with vaseline in a colored slide stained with 1% Brilliant Cresyl Blue to demonstrate inclusion bodies for α-thalassemia determination. Abnormal Hb and thalassemia were determined, in every case, by Hb electrophoresis by use of a cellulose acetate strip (9). Diagnosis of β-thalassemia was confirmed by micromoum chromatography with DEAE-Sephadex a-50 (Pharmacia Fine Chemicals AB, Uppsala, Sweden) (10). Any other Hb abnormalities were confirmed with starch gel electrophoresis (11). α-Thalassemia trait was determined by the positive inclusion bodies and percent of the Hb Barts which was calculated by solid-phase, two-site immunoradiometric assay (T Sanguansrsri, L Makornkoswkyoon, S Yaemniam, personal communication, 1987).

Hb, hematocrit, red-blood-cell and white-blood-cell counts and mean cell volume were done by blood-cell counter (Cell-Dyn 400® hematology analyzers, Sequaia- Turner Co, Mountain View, CA). A pediatric hematologist (PH) examined blood-cell morphology. Serum iron and total iron binding capacity were measured according to the specific recommendations from the International Committee for Standardization in Hematology (12). Transferrin saturation was calculated by the ratio of serum iron to total iron binding capacity, × 100. Serum ferritin was determined by double-antibody radioimmuno-

noassay with a kit (13). Free erythrocyte protoporphyrin was determined by the micromethod (14).

**Intelligence function and educational achievement.** IQ, Thai language, and math tests were administered in group form in the classrooms after the children were informed of the nature and objectives of the study. Two field researchers and a school teacher checked the IDs, distributed the test booklets and answer sheets, supervised the subjects throughout the test, and collected the completed forms.

The Raven Colored Progressive Matrices (RCPM) (15), forms A, Ab, and B, were used to measure general intellectual function (IQ). The educational achievement tests for Thai language and mathematics (math) were modified from the Thai government’s Ministry of Education instruments used in the public school system. Each test was divided into two parallel forms, based on the odd (O) and even (E) items of the original test forms. Half of the sample took the O form before treatment (T1) and the E form after treatment (T2) while the remaining half took the tests in the converse sequence (E, O). Children taking the same forms were not placed next to each other. A split-half reliability assessment conducted in a pilot study with 191 children showed that forms O and E yield scores similar to each other. The split-half correlations for the Thai language and math tests ranged from 0.92 to 0.97.

**Socioeconomic status.** The parents responded to a questionnaire on graduated indices of social structure such as parental education, family income and size, land owned, quality of housing, and availability of electrical appliances. School teachers reviewed the questionnaires for completeness and helped illiterate parents to complete the questionnaires.

**Experimental intervention**

The iron treatment consisted of a 50 mg/d tablet of ferrous sulphate during the first 2 wk (2 mg elemental Fe·kg⁻¹·d⁻¹) and a 100-mg/d tablet (4 mg·kg⁻¹·d⁻¹) during the remaining 14 wk. The placebo was a tablet with sweet cassava powder and had a color and appearance similar to those of the iron tablet. The tablets were distributed daily by the room teacher, who was supervised weekly by a field director. The teacher did not know whether the tablet given to a child contained iron or cassava.

Two 200-mg tablets of albendazole, a benzimidazole (Smith, Kline and French, London) anthelmintic drug (methyl-5 propylthio-1-H benzimidazol-2-yl carbamate), were given to every student enrolled in the study on the day of the blood test at T1 and 3 mo after the day the intervention started (16, 17).

**Data analysis**

The following criteria were used for case inclusion in the data analysis: 1) availability of T1 and T2 measurements for IQ, educational achievement, and Hb; 2) not more than one missing value at T1 or T2 for serum ferritin, transferrin saturation, and free erythrocyte protopor-
phonyrin; j) T1 and T2 Thai language and math test scores derived from different forms (E, O); 4) an IQ delta (T2 – T1) < 2 SD below the mean of the respective distribution of delta scores; 5) ingestion of tablets to have resulted in a dosage of ≥ 3 mg elemental Fe · kg⁻¹ · d⁻¹ to those treated with iron, or its equivalent (for those treated with a placebo). One thousand seven hundred and seventy-five children satisfied these criteria.

The specific objectives of the data analyses were to assess whether the iron intervention had an effect on four hematological variables (Hb, serum ferritin, transferrin saturation, and free erythrocyte protoporphyrin), on IQ, and on two educational variables (Thai language and math, as a function of the iron status of the children. Children were classified into one of three groups according to their body iron status: IDA, iron depleted, and iron replete. IDA was defined by a Hb < 120 g/L plus two out of three of the following criteria: serum ferritin < 10 µg/L, transferrin saturation < 16%, and free erythrocyte protoporphyrin > 700 µg/L RBC. Iron depleted was defined by the same criteria except for a Hb ≥ 120 g/L. Iron replete was also identified by a Hb ≥ 120 g/L plus two of the following three criteria: serum ferritin ≥ 10 µg/L, transferrin saturation ≥ 16%, and free erythrocyte protoporphyrin ≤ 700 µg/L RBC. One thousand three hundred and fifty-eight children were classified into one of the three iron groups; 417 could not be classified and were excluded from the data analysis. There were 101, 47, and 1210 children in the IDA, iron depleted, and iron-replete groups, respectively. These numbers were determined by the prevalence of the respective states of body iron within the population. As noted, the assignment to the two different forms of intervention was done before the determinations of body iron status were established. Figure 1 presents a schematic of the study population selection process and of the experimental design.

The statistical model chosen to achieve the objectives identified above was an unbalanced five-way analysis of variance (ANOVA)—grade by school by iron status by treatment by time—with repeated measures on one factor (18). The following considerations prompted the choice of model. The sample size with each body iron storage category reflected the prevalence of a given state and was not manipulated by continued sampling to achieve equal numbers in the categories. Thus, the design is unbalanced. The need for a multiway ANOVA approach was indicated by a possible difference in the responses to treatment as a function of school and grade. Gender was included as a sixth factor in the analysis of the hematological variables. Thus, the statistical model partitions the variability due to those factors and allows for a cleaner assessment of the response to treatment within iron-status groups. The repeated-measure methodology was introduced as a blocking factor at an individual level. Variability among subjects between T1 and T2 due to individual differences is removed from the error variance. The repeated-measure design also allows for the inclusion into the model of a covariance between T1 and T2.

This paper is not concerned with the reasons for, or significance of, the possible effect of school, grade, and gender on the hematology or the psychoeducational variables. Their inclusion in the present analysis merely reduces the size of the error variance and thereby increases our capacity to detect treatment differences. Therefore, little attention or discussion is given to their relationship with any of the outcome variables.

A differential response to the experimental intervention should be seen in a significant interaction of the treatment (iron or placebo), iron status (IDA, depleted, or replete) and time (T1 or T2).

Results

Hematology

Table 1 presents the T1 and T2 adjusted mean values of the body iron indicators for each of the three iron groups, classified by the type of intervention received. The initial and postintervention values were adjusted for the effects of school, grade, and gender.

As foreseen, the between-group analyses in the repeated-measure assessment yielded statistically significant (< 0.001) main effects (across T1 and T2) of iron status and treatment on the four iron-status variables (Table 2). There were also a statistically significant (< 0.01) main effect for school on each of these variables and main effects for grade and sex (< 0.01) for Hb and transferrin saturation, respectively.

Paired comparisons yielded no iron-status or treatment differences in Hb at T2. Although the other three iron-status indices differed between the IDA and iron-replete groups at T2 (< 0.01), mean values for Hb, serum ferritin, transferrin saturation, and free erythrocyte protoporphyrin for the IDA children were all within the normal range as determined by the criteria used for the classification of iron status.

The within-group analysis showed that the change-over-time variable (T2 – T1) was highly significant (< 0.00001) for all four iron indicators (Table 2). However, each interacted significantly with treatment (< 0.0001). Except for transferrin saturation, the interactions between time, treatment, and iron status were also statistically significant (< 0.01). In the case of transferrin saturation this three-way interaction had an F of 2.40 (< 0.0911).

Paired comparisons within groups and between times (T2 – T1) showed that the Hb increment of the ID children treated with either iron (change of 29.5 ± 0.27 g/L, x ± SEM) or placebo (14.6 ± 0.14 g/L) was statistically significant (< 0.0001). Likewise, the small changes in the iron-depleted (5.3 ± 0.15 g/L) and iron-replete (1.5 ± 0.03 g/L) children treated with iron were statistically significant (< 0.0001). The fact that this latter small delta reached a high statistical level of significance is explained by the large sample size (n = 1210) and the relative small standard error of the measure (0.3). The iron-replete children treated with a placebo had a negative delta (−1.5 ± 0.03 g/L), which was also statistically sig-
significant \((p < 0.001)\). The change among the iron-depleted children on placebo \((-0.2)\) was not significant.

The paired comparison for the other iron indices yielded similar results to those of Hb; the administration of iron consistently resulted in statistically significant changes in the expected direction. The changes observed in those that received a placebo (serum ferritin, transferrin saturation, and free erythrocyte protoporphyrin) were also statistically significant \((p < 0.001)\) and in the direction of replenishing iron stores. The only result that

**TABLE I**

Adjusted* means of hemoglobin (Hb), serum ferritin (SF), transferrin saturation (TS), and free erythrocyte protoporphyrin (FEP) for iron-deficient anemic, iron-depleted, and iron-replete subjects before (T1) and after (T2) experimental intervention†

<table>
<thead>
<tr>
<th>Group</th>
<th>Hb</th>
<th>SF</th>
<th>TS</th>
<th>FEP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Iron-deficient anemic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>99.00±1.02</td>
<td>113.00±1.09</td>
<td>6.50±3.40</td>
<td>16.30±5.21</td>
</tr>
<tr>
<td>Fe depleted</td>
<td>124.00±1.69</td>
<td>125.00±1.80</td>
<td>15.90±5.59</td>
<td>32.90±8.57</td>
</tr>
<tr>
<td>Placebo</td>
<td>125.00±1.45</td>
<td>130.00±1.55</td>
<td>14.90±4.81</td>
<td>18.40±7.37</td>
</tr>
<tr>
<td>Fe replete</td>
<td>129.00±0.31</td>
<td>127.00±0.33</td>
<td>39.70±1.04</td>
<td>43.80±1.60</td>
</tr>
<tr>
<td>Placebo</td>
<td>128.00±0.31</td>
<td>130.00±0.34</td>
<td>38.80±1.05</td>
<td>107.70±1.62</td>
</tr>
</tbody>
</table>

* Adjusted for school, grade, and sex.
† \(\bar{x} \pm \text{SEM.}\)
deviated from the expected pattern was the non-significant change of serum ferritin in the iron-depleted children treated with a placebo.

**IQ and educational achievement**

The correlation between T1 and T2 RCPM scores ranged from 0.58 to 0.78 across schools and grades. For Thai language and for math these correlations ranged from 0.51 to 0.73 and from 0.56 to 0.78, respectively. Partial test-retest correlations of IQ for the total sample (1775) controlling for school, grade, and treatment was 0.70. The same partial correlations for Thai language and math were 0.58 and 0.61, respectively.

Figures 2 and 3 present the T1 and T2 values, respectively, for IQ, Thai language, and math scores for each of the three iron groups broken down by the iron-placebo intervention. Table 3 presents in summary form the results from the ANOVA with repeated measures for the T1 and T2 data.

The between-group analyses yielded significant main effects for school and grade. Invariably, the mean scores of the children in the third grade were the highest and those of the fifth graders were the lowest. The main effect of iron status (across T1 and T2) was statistically significant for IQ and Thai language. The F test for the math score had a p of 0.07. Except for the comparatively high IQs at T1 and T2 of the iron-depleted children treated with a placebo, the scores of the iron-replete children were consistently higher than those of the iron-depleted and IDA children.

Multiple comparisons were calculated using the least-square means of the general linear model procedure to contrast groups in the three psychoeducational variables. For this purpose the average scores ([T1 + T2]/2) of the IQ, Thai language, and math scores were used. The IQ mean score of the IDA children (90.77 ± 0.96) was smaller and significantly different (p = 0.0008) from that of the iron-replete children (94.16 ± 0.29). The mean Thai language score of the iron-replete (58.97 ± 0.45) children was different from that of the anemic (55.92 ± 1.49) (p < 0.05) and the iron-depleted (51.76 ± 2.18) children (p < 0.01). None of the differences in the math scores reached the conventional level of statistical significance (p < 0.05); the p value for the difference between anemic (51.03 ± 1.48) and normal children (53.74 ± 0.45) was 0.08.

The within group analyses yielded a powerful effect of time (T2 - T1) for the three outcome variables; on the average, the children in the three groups improved their scores from T1 to T2. Contrary to expectations based on the study in Indonesia (6), the interactive terms of time by treatment or time by iron status by treatment were not statistically significant. There were also no significant interactions of time and treatment with either school or grade. Thus, there was no evidence that iron treatment had an effect on IQ or on either of the two educational achievement measures. Likewise, there was no evidence that the magnitude of the delta (T2 - T1) in any of the three outcome measures was associated with treatment.

**Analysis of covariates**

An absence of effects from the iron intervention on any of the three outcome measures places in question the
FIG 2. Pretreatment results, by iron and treatment status. Top: mean intelligence quotient (from Raven's progressive matrices). Middle: mean percentage correct on Thai language measure. Bottom: mean percentage correct on math measure.
### TABLE 3
Summary of between and within analysis from ANOVA with repeated measures: IQ and achievement scores

<table>
<thead>
<tr>
<th>Source</th>
<th>Raven Colored Progressive Matrices</th>
<th>Thai</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Between groups (model 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>15</td>
<td>2.57</td>
<td>0.0008</td>
</tr>
<tr>
<td>Iron status (Fe)</td>
<td>2</td>
<td>6.23</td>
<td>0.0020</td>
</tr>
<tr>
<td>Grade</td>
<td>2</td>
<td>146.8</td>
<td>0.0000</td>
</tr>
<tr>
<td>Treatment (Rx)</td>
<td>1</td>
<td>0.21</td>
<td>NS</td>
</tr>
<tr>
<td>Error</td>
<td>1337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within groups (model 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (T2 – T1)</td>
<td>1</td>
<td>50.11</td>
<td>0.0001</td>
</tr>
<tr>
<td>Time × Fe</td>
<td>2</td>
<td>0.73</td>
<td>NS</td>
</tr>
<tr>
<td>Time × school</td>
<td>15</td>
<td>1.25</td>
<td>NS</td>
</tr>
<tr>
<td>Time × grade</td>
<td>2</td>
<td>1.59</td>
<td>NS</td>
</tr>
<tr>
<td>Time × R</td>
<td>1</td>
<td>2.16</td>
<td>NS</td>
</tr>
<tr>
<td>Time × Fe × school</td>
<td>24</td>
<td>1.71</td>
<td>0.017</td>
</tr>
<tr>
<td>Time × Fe × grade</td>
<td>4</td>
<td>0.62</td>
<td>NS</td>
</tr>
<tr>
<td>Time × R × grade</td>
<td>2</td>
<td>1.74</td>
<td>NS</td>
</tr>
<tr>
<td>Time × school × grade</td>
<td>30</td>
<td>1.69</td>
<td>0.0116</td>
</tr>
<tr>
<td>Time × school × R</td>
<td>15</td>
<td>1.24</td>
<td>NS</td>
</tr>
<tr>
<td>Time × grade × R</td>
<td>2</td>
<td>1.36</td>
<td>NS</td>
</tr>
<tr>
<td>Error</td>
<td>1260</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

validity of an assumption that the statistical associations between iron status on the one hand and IQ and Thai language scores on the other are indeed causal. Conceivably, iron-replete children are taller, heavier, and are members of families with higher income and education than the iron-depleted and IDA children. To test such a hypothesis analyses of covariance (ANCOVA) were done in which the effects of anthropometric and socioeconomic variables were controlled for.

**Anthropometry.** One-way ANOVAs were calculated for the between-group comparisons of T1 mean values of height, weight, and head circumference. Height-for-age discriminated between groups at a statistically significant level ($p < 0.001$); the IDA children were smaller than the other two groups. Accordingly, height z-score (i.e., deviation from the median using the anthropometric reference standards from the United States National Center for Health Statistics) was selected as a covariate for the reassessment of the relationship between body iron status and IQ, Thai language, and math test scores.

**Socioeconomic status.** Father's and mother's education, family income, family size, and the child's ordinal position did not discriminate between the three groups of children. Likewise, there were no differences between groups in the number of families who owned the house where they lived. However, chi-square tests for the presence or absence of radio, refrigerator, and television in the three groups of children were statistically significant. Consistently, the iron-replete children were more likely to own these electrical appliances than the other two groups of children. Therefore, the presence of each of these three electrical appliances were added for each subject and the score was used as an indicator of wealth.

An ANCOVA was calculated for IQ and Thai language scores including z-height and the socioeconomic indicator as covariates. The outcome measures were the average of the T1 and T2 scores. The effects of the two covariates were large and statistically significant in the three ANCOVAs. Once these effects were removed the main effect of iron status on IQ and Thai language was still statistically significant ($p < 0.01$). Treatment, treatment by time, and treatment by iron status by time were still nonsignificant.

### Discussion

The following discussion is organized according to the four validity criteria discussed by Cook and Campbell (7) in reference to experimental field studies. These criteria have already been discussed in connection with studies on the impact of ID on behavioral development (19).

**Statistical-conclusion validity**

There were statistically reliable associations between body iron status and IQ and a language school achievement test among third to fifth graders in Chon Buri province. The validity of this observation is strengthened by 1) the large sample size used, 2) the test-retest stability of the IQ and language tests, 3) the strength of the diagnostic criteria precluding sample heterogeneity (7, 19), 4) the power of the statistical model used, and 5) the results of the ANCOVA. This statistical association, however, is not sufficient to infer a causal relationship between iron and performance.

A new finding among school children is that the chil-
Children who were iron-depleted without anemia obtained statistically significantly lower scores in Thai language than did the iron-replete children. The differences between these two groups in IQ and math were also in the same direction. The IQ and math tests were also lower in the former group, albeit not significantly. A small number of studies have also reported cognitive and developmental differences between iron-depleted and iron-replete infants and preschool children (19). Two reasons why this difference was not detected in other studies might be small sample sizes and the small magnitude of the variation. In any event the present findings suggest that there is a need for finer investigations of the behavioral consequences of ID without anemia.

Internal validity

The large and significant effects of time in the within-group analysis, in concert with a lack of significant interactions between time and treatment, and between time, treatment, and iron status indicate that the changes (T2 – T1) observed in the psychoeducational variables (Thai and math) were not different for those treated with iron as compared with those treated with a placebo. Thus, the study provides no support for an assumption of causality. However, before inferring that iron treatment has no effect on school performance, after 16 wk of treatment the following alternative explanations should be considered.

It was initially conjectured that all groups in the study would improve their performance somewhat in the achievement tests from T1 to T2 because in the 16-wk interim period the children were likely to learn information relevant to questions included in the math and Thai language tests. The data supported this first assumption. It was also conjectured that those IDA children who received placebos would not learn as fast as children in the other groups. However, there is no evidence in the study that any one of the subgroups in the study differed in their rate of learning (T2 – T1) from any of the others; thus, this second conjecture was not supported. An explanation for this finding, which is counter-intuitive, is that the albendazole treatment provided enough health and/or nutritional benefits to the IDA children who received a placebo to allow them to keep a similar pace of learning as that of the other groups.

A nutritional effect of albendazole is inferred from the hematological data, which show a statistically significant positive change (T2 – T1) in the iron-status indicators of the IDA children who received a placebo. For example, in this group the pre- to posttreatment change in Hb was +14.5 g/L. This change is not as high as that of the IDA children treated with iron (29.5 g/L) but it still is highly significant (p < 0.0001; n = 41).

A second hypothesis is that school attainment as measured by language or math achievement tests is not likely to manifest the beneficial effects of treatment with iron over a 16-wk period. School learning is a complex dynamic process influenced by a series of factors such as aptitude, time on task, perseverance, and quality of instruction (20–22), among others. ID, particularly if it is chronic, may affect rate of learning through specific effects on one or many of these factors. This condition, in turn, will impose limitations on the knowledge and skills acquired over time. A change in iron status may immediately correct the selected attentional ailment but it will not remedy at the same time the deficits in the acquisition of information.

The two alternative hypotheses posited are plausible and suggest that the absence of a group without any treatment and the relatively short interim time between T1 and T2 worked against the rejection of the null hypothesis. By the same token, it must be concluded that the statistical associations demonstrated are not sufficient evidence to infer causality.

A related issue of internal validity is the role of the socioeconomic-status (SES) variable. The findings from the ANCOVA that the between-group differences in IQ and Thai language prevail after the effects of SES are separated do not rule out the possibility that SES mediates the relationships observed between iron status and IQ and Thai language. The SES variable used (ie, number of electrical appliances) is a crude measure of wealth and is not likely to capture the different aspects of the social environment that might account for significant portions of the variance in iron status and in the two psychoeducational variables in question. Accordingly, we cannot rule out the possibility that a more rigorous control of the social and economic environment might result in nonsignificant correlations between iron status and IQ and between iron status and Thai language.

Construct validity

The comparatively lower test performance of the IDA children in this study is in agreement with the results from studies showing comparatively low developmental-scale scores among IDA infants (1–5). However, there are no theoretical justifications or empirical support for an assumption that the developmental constructs assessed by infant scales are the same or similar to those assessed by the Raven Progressive Matrices or the educational tests used in this study. In fact, the converse is true. Scores from early developmental assessments are not correlated with IQ or scores from achievement tests in middle childhood (23). This lack of predictive validity in the infant tests is uncontroversial evidence that there are no commonalities between the constructs assessed by the early developmental scales and IQ or educational achievement measures (see ref 6 for an excellent discussion of these issues). Therefore, the concurrence in the direction of the findings in studies of infants and school children is not explained by a sweeping generalization that ID affects general cognitive function. The reasons why ID relates to infant developmental scores and performance in achievement tests may be better explained by the possible effects of iron status on state variables, such as arousal, or by a specific effect on specific cognitive processes such as attention. Attentional deficits can interfere with performance in tests of very different natures. Further research is needed, however, before these assumptions can be accepted.
External validity

Can the results from this study be generalized to other populations? Clearly, a lack of internal validity precludes any claims of generalizability (7). However, the similarities of the correlational findings between this and the Indonesian study (24) cannot be discarded. They suggest that there is an educational disadvantage associated to IDA independent of ecological setting, population, and cultural milieu.

An important difference between studies is that in Thailand, but not in Indonesia, the Raven Progressive Matrices IQ test discriminated between groups at a statistically significant level. On the basis of the available information it has not been possible to identify the reasons for such differences.

References


Comments

Frank Oski

Dr Pollitt and his co-workers provide us with a well-designed study of a double-blind clinical trial conducted in third- to fifth-grade children in schools in Thailand that demonstrates that iron-replete children had significantly higher scores for IQ and the Thai language than did iron-depleted and iron-deficient anemic children. Iron treatment for a period of 14 wk could not correct the difference.

In the design of the study one can question the categorization criteria employed. A free erythrocyte protoporphyrin value of 1000 μg/L red blood cells (RBCs) is commonly regarded as the upper limit of normal whereas in this study a value of 700 μg/L RBCs was chosen. The use of a nonfasting serum transferrin saturation of 16% may also be questioned as a reliable means of classifying children with respect to iron status.

Of particular importance in drawing conclusions from this study are answers to the questions, why are these children anemic and how long have they been anemic? Are they a homogeneous group? The fact that test scores

---

1 From the Children's Medical and Surgical Center, Johns Hopkins University, Baltimore, MD.
also correlated with anthropometry and socioeconomic status raises the question, could coexistent nutritional deficiencies be responsible for the differences observed and for the failure of iron therapy to produce changes in the test scores? This study, and similar studies in which interventions fail, must include more comprehensive nutritional assessments before conclusions can be drawn with complete confidence.

Comments

Moussa BH Youdim

Dr Pollitt and his co-workers have conducted a well-planned, double-blind clinical trial to assess the impact of iron treatment on IQ and educational attainment of > 1000 children aged 9–11 y. They should be commended for the sheer number of the subjects. I have three sets of questions to ask about this quite straightforward paper.

1) Why were 9–11-y-old children chosen and were these the ones who showed the greatest preponderance to iron deficiency? Would the results have been different with another age group? Could age have some influence on IQ development?

2) I also don’t understand the iron dosage and duration of treatment. Why was it 2 mg·kg⁻¹·d⁻¹ for 2 wk followed for 14 wk of 4 mg·kg⁻¹·d⁻¹? Can these be considered the optimal regimes for iron-depleted or iron-deficient groups?

3) The data clearly demonstrate a statistically positive association between iron status and IQ and a language school achievement test. However, there is no evidence that these effects are caused by iron because there is no significant interaction between time and treatment and between treatment and iron status. The data (results), therefore, implicate other factors, such as social and economic status, that could explain the differences. Although the authors excluded that possibility, it is worth asking whether the establishment of SES was objective enough. Other variables, such as more sensitive, motivated, emotional behavior, were not examined or discussed.

Response to comments by Moussa BH Youdim

Ernesto Pollitt

The criteria for the age of the children were the following: 1) pre-pubescent to avoid the possible intervention effects of sexual maturation on cognitive development and understand clearly the objectives of the study.

2) Old enough to handle the stress of venipuncture and understand clearly the objectives of the study.

We do not have any data to assess whether the putative effects of ID on 9–11-y-old children are also observed in school-age children of different ages.

The first 2 wk of the treatment with 2 mg·kg⁻¹·d⁻¹ were intended to serve as an adaptation period, followed by a therapeutic dosage during a period of time that was considered long enough to allow for repletion of iron stores.

As we discussed in the paper, it is conceivable that other factors besides iron lie behind the pre- and post-treatment differences in IQ and educational test performance found between groups. In the absence of treatment effects there is no way we can infer that such differences are mediated solely by differences in iron status.