

Infancy: mental and motor development^{1,2}

Tomás Walter

ABSTRACT In a prospective cohort study of 196 infants from birth to age 15 mo, the relationship of iron status to psychomotor development, the effect of a short-term trial of oral iron or placebo, and the effect of longer-term oral iron therapy was assessed. Development was assessed with the Bayley Scale of Infant Development in anemic, nonanemic, iron-deficient, and control children. Anemic infants had significantly lower indices than did control or nonanemic, iron-deficient infants. Control infants and nonanemic, iron-deficient infants performed comparably. No difference between the effect of oral administration of iron or placebo was noted after 10 d or 3 mo of iron therapy. A hemoglobin concentration of < 105 g/L and anemia duration > 3 mo were correlated with significantly lower motor and mental scores, suggesting that when iron deficiency progresses to anemia, adverse influences in the performance of developmental tests appear and persist, despite iron therapy. *Am J Clin Nutr* 1989;50:655-66.

KEY WORDS Iron deficiency, anemia, psychomotor development, infancy

Introduction

For over a decade, and heralded by the pioneering work of Oski and Honig, there have been more than a dozen studies addressing the cognitive effects of iron deficiency (ID) during infancy. The inherent difficulties of identifying intervening variables in the complex field of mental development, coupled in some cases with suboptimal design, have worked against making significant headway. However, the two most recent studies, one performed in Costa Rica by Dr Betsy Lozoff (1) and our own study in Santiago, Chile, are perhaps the least contaminated with the uncertainties that have precluded firm conclusions arising from previous work.

These two studies began almost simultaneously in 1982-1983 and arose as the natural evolution of Lozoff's previous study in Guatemala and our own early study in Chile. The striking similarities in the results of both protocols gives great strength to the conclusions of both Lozoff's and our work. It is remarkable that these similarities were possible despite the fact that there are important differences in the study design and that the research was carried out by two independent investigators in two distant regions.

This study was initiated in 1983, performed in concert with a field trial of fortified infant foods (2). Briefly, a cohort of healthy, full-term infants from a community clinic in the city of Santiago were included into a food-fortification study at age 3 mo and followed to age 12 mo with monthly clinic checkups and weekly house calls by

a nurse. Complete anthropometric, nutritional, morbidity, and socioeconomic data were collected. Those infants spontaneously weaned by 3 mo received either an iron-fortified cow milk or the nonfortified milk normally provided at no cost by the clinic. Those infants who were being breast-fed at 3 mo were assigned to one of two groups. The first received a heme-iron-fortified cereal in addition to normal nonmilk foods and the second received no iron fortification. Because the assignment to fortified food products was random and fortification turned out eventually to be the most potent determinant of the infant's iron status, all other intervening factors were essentially offset by this design.

Approximately 100 infants were entered in each of the four groups, with a 20% attrition during the 9 mo of follow-up due mainly to migration. At ages 9 and 12 mo venipunctures for a full hematologic assessment were performed. At 12 mo parents were invited to participate in the psychologic study and, after informed consent was obtained, in accordance with the pertinent institutional review committee on ethical standards of research on human subjects, the tests were begun 7-10 d after the 12-mo checkup.

Since the first report 10 y ago (3) it has become clear that ID has adverse effects on infant psychomotor devel-

¹ From the Unidad Hematología, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Santiago.

² Address reprint requests to T Walter, Unidad Hematología, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Casilla 15138, Santiago 11, Chile.

TABLE 1
Classification criteria of iron status*

Group	Criteria
Anemic	Hb < 110 g/L and two or more abnormal biochemical measures
Control	Hb > 110 g/L MCV > 70 fL Fe/TIBC > 10% FEP < 1000 ZPP/L SF > 10 µg/L
NAID	Response to iron therapy Hb < 100 g/L Hb > 110 g/L, but one or more abnormal biochemical measures or response to iron therapy Hb > 100 g/L

* MCV, mean corpuscular volume; TIBC, total iron-binding capacity; FEP, erythrocyte protoporphyrin; ZPP, zinc protoporphyrin; SF, serum ferritin.

opment (4). In most studies the Bayley's Scale of Infant Development (BSID) (5) has been the test used and even though this tool has drawbacks, such as its global nature and poor IQ predictability (6), to be discussed below, it is well known, reproducible, reliable, and accepted as a good measure of behavior in infancy. During this stage the rapid progression of maturational events is appropriately addressed by this test (7). The almost universal use of this measure allows for useful comparisons, disadvantages notwithstanding.

Our aims in this study were to provide answers to the following questions:

- 1) How severe must ID be to affect behavior?
- 2) What is the effect of duration of ID?
- 3) What is the effect of short-term iron therapy (before correction of anemia)?
- 4) What is the reversibility of the changes after a long-term trial (enough to revert anemia)?
- 5) What specific areas of mental or motor processes are most affected?
- 6) Is there an association between developmental deficits and behavior patterns?

The iron-status assessment at 12 mo permitted a preliminary classification based on hemoglobin (Hb), mean corpuscular volume (MCV), and erythrocyte protoporphyrin (FEP), which were readily available, and 7-10 d after this venipuncture the first BSID test was performed. Within each preliminary iron-status group, infants were randomly assigned to FeSO₄ drops or placebo for 10 d, when BSID tests were again given. At that point all infants received FeSO₄ at 3-5 mg·kg⁻¹·d⁻¹ in two divided doses for 75 d, at the end of which BSID tests and iron-status assessments were repeated for a third time. All of the BSID tests were administered by the same psychologist who was unaware of the iron status of the child and the therapy assignment, facts also unknown to the mother.

The BSID test is a well-known and accepted tool for evaluating psychomotor development from ages 3 to 30

TABLE 2
Subclassification of NAID

Grade 1 (iron depleted)	All measures normal except for SF < 10 µg/L
Grade 2 (nonresponder)	One to four abnormal values but with a therapeutic response Hb < 100 g/L
Grade 3 (responder)	Zero to four abnormal values and/or a therapeutic response Hb > 100 g/L

mo; it consists of a mental and a motor scale. The mental scale measures functions related to the basic foundation for cognition such as language acquisition and abstract thinking. The motor, or psychomotor, scale relates to gross motor abilities such as coordination, body balance, and walking. Both are age-adjusted to give an index of mental and psychomotor development (MDI and PDI, respectively) very much like an IQ, with a mean of 100 and a SD of 16. A third scale, the Infant Behavior Record (IBR), is based on a clinical observation and does not yield an index.

The criteria used for classification into three groups (anemic; control; and nonanemic, iron deficient [NAID]) were intended to strictly define anemic and control subjects with state-of-the-art measures, including the therapeutic response considered the "gold standard" (Table 1). This methodology excluded the majority of infants with intermediate iron status, who fell into the broadly heterogeneous NAID group. The NAID infants were reclassified into grades of severity depending on response to the therapeutic trial and measures of iron status (Table 2).

Hematologic data

Complete hematologic evaluation was performed in 189 infants at age 9 mo (when no BSID tests were done) and in 196 infants at ages 12 and 15 mo. The final classification of the infants and mean values for hematologic measures demonstrate that because of the stringent criteria for classification, only 30 control and 39 anemic subjects were identified (Table 3) whereas 127 infants fell

TABLE 3
Iron status at 12 mo*

	Hb	Fe/TIBC	FEP	SF†
	g/L	%	mg ZPP/L RBC	µg/L
Anemic (n = 39)	100 ± 9	6.8 ± 2.9	1950 ± 1030	5.4 (3-9.8)
NAID (n = 127)	121 ± 7	12.2 ± 0.7	1080 ± 330	11.9 (6-24)
Control (n = 30)	127 ± 8	16.7 ± 6.3	780 ± 130	19.8 (12-34)

* $\bar{x} \pm SD$. See Table 1 for definitions of terms.

† Geometric mean (range of 1 SD).

TABLE 4
Subclassification of NAID*

	Hb	Fe/TIBC	FEP	SF†
	g/L	%	mg ZPP/L RBC	μg/L
Grade 1 (iron depleted)	129 ± 6	14.8 ± 3.6	810 ± 12	7.5 (6.7-8.2)
Grade 2 (nonresponder)	123 ± 7	12.3 ± 0.7	1070 ± 30	15 (7.7-27)
Grade 3 (responder)	117 ± 5	11.7 ± 0.5	1150 ± 38	9 (4,1-20)

* $\bar{x} \pm SD$. See Table 1 for definitions of terms.

† Geometric mean (range of 1 SD).

into the NAID group where the criteria were successful in segregating degrees of severity of ID noted in the progression of iron-status values (Table 4).

Development scores

A total of 576 BSID tests were administered and the results were distributed in a normal fashion. Mean MDI was 102 ± 9 and mean PDI, 98.1 ± 11. There was a significant difference between MDI and PDI in the population as a whole. Both distributions were, however, symmetric (Fig 1). This population is biased by design with the selection of healthy infants from a closely followed

cohort. We purposely excluded confounding influences in host and environment by longitudinally following a cohort of infants in an optimal state of health, with the exception of iron nutrition, and this iron status was determined randomly by fortified-food assignment. The proximity of the scores to the US norm and the symmetry of the BSID distribution are likely consequences of this design, eliminating confounding low performances usually present in disadvantaged populations.

Effect of iron status on developmental scores

It is clear that a decrease in Hb is necessary to affect mental and psychomotor development scores. The performance of the NAID infants as a whole was indistinguishable from the control infants' (Fig 2). None of the NAID subclassifications was successful in showing differences in performance (Fig 3). Even group 3, ie, infants who were nonanemic and responded to the therapeutic trial—who could be technically defined as anemic—showed any tendency towards lower scores.

The MDI-PDI difference is present across all iron-status levels. However, this difference between scores is less pronounced in the control group, where mental and motor performances appear to be better balanced. As ID progresses, motor performances are more affected than mental indices—as seen by the divergence of these mean

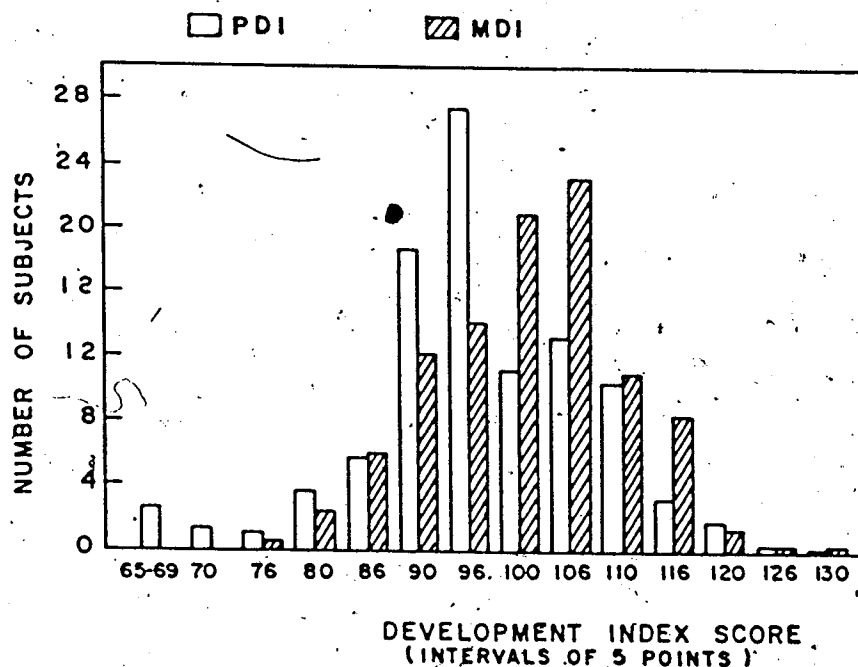


FIG 1. Histogram of Mental Development Index and Psychomotor Development Index scores on 576 Bayley's Scale of Infant Development tests measured in 196 infants at ages 12 and 15 mo of age. Open bars (PDI) show a shift of the normal curve to the left with a mode at 96 whereas the mode of the MDI's hatched bars is at 106. The means are 98.1 and 102.4, respectively. Both distributions are symmetric with very few performances under or above 2 SD.

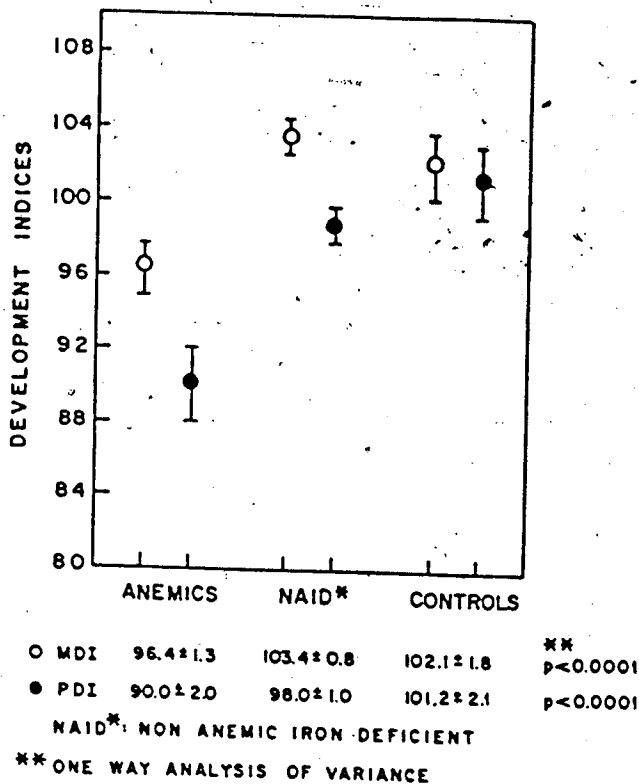


FIG 2. Effect of iron status at 12 mo (anemic, nonanemic ID, and control infants) on Mental Development Index (○) and Psychomotor Development Index (●). Only anemic infants differ significantly from the other two iron-status categories for both MDI and PDI.

values. Explanations for this phenomenon remain conjectural.

Effect of Hb on development scores

A sigmoid distribution is defined when MDI and PDI are distributed according to Hb level, the most common indicator of iron status (Fig 4). Three groups can be identified, those with Hb < 105 g/L, with Hb > 110 g/L, and an intermediate group at 105–109 Hb g/L. All are statistically distinct from each other by pair-wise and ANOVA comparisons. Thus, in anemic infants, Hb concentration was correlated with performance, so that infants with moderate anemia (Hb 84–104 g/L) had significantly lower scores than those with mild degrees of anemia (Hb 105–109 g/L). The latter in turn had poorer indices than infants with Hb > 110 g/L, with no graded improvements seen at higher Hb levels.

If the effect of ID on behavior is mediated by metabolic processes dependent on the presence of iron, it is understandable that overt anemia may be necessary to disclose these effects. Siimes et al (8) showed in an animal model fed graded amounts of iron that tissue-heme proteins were not affected until saturation of transferrin fell sig-

nificantly. Hb, as well as tissue cytochrome C and myoglobin, decreased steadily thereafter, demonstrating that availability of iron to the erythroid marrow is limited concomitantly to other tissues. Changes in iron stores (liver nonheme iron) did not influence Hb level or tissue-heme iron proteins. In humans the stage known as iron-deficient erythropoiesis—when iron availability becomes a limiting factor for Hb synthesis—corresponds to the moment when Hb concentration begins to decrease, presumably along with other tissue iron proteins. However, the stage of overt anemia is reached later, only when Hb values fall under 110 g/L for the infants in this study. Anemia ensues, therefore, after a rather protracted and severe period of iron lack, ensuring significant depletion of tissue iron proteins. Milder iron deficit may fall short of achieving sufficient tissue depletion to be reflected in behavior. On the other hand, we must consider that the psychologic tests available for this age group may be too crude to identify subtle deficits. These considerations may help explain the absence of cognitive effect seen in the nonanemic-responder population (subclassification of NAID, grade 3), which conceivably corresponds to the group of infants with limited Hb synthesis, soon to become anemic.

Effect of duration of anemia

For this purpose, we studied the effect of iron status at age 9 mo on development indices measured at 12 mo. The indices were distributed according to the Hb at 9 mo, ie, 3 mo before the BSID was performed, and showed a cutoff at 105 g/L (data not shown). To further pursue the effect of duration of anemia, infants who were anemic at both 9 and 12 mo, ie, those whose anemia had a duration of ≥ 3 mo (n = 19) were compared with those children who were anemic at 12 but not at 9 mo, ie, those whose anemia was presumed to be present for < 3 mo (n = 16). Of the 39 anemics at 12 mo, only 35 had Hb determined at 9 mo. Those infants anemic for > 3 mo had significantly lower development indices than those anemic for < 3 mo (Table 5). Nonanemic infants at age 9 mo fell mostly into the intermediate Hb range of 105–109 g/L at 12 mo, where development indices were less severely affected. Thus, infants not anemic at 9 mo tended to have milder anemia at 12 mo; of the 14 infants in the intermediate Hb group at 12 mo, 10 (71%) were not anemic at 9 mo. Of the 21 infants who fell below the intermediate Hb level, only 6 (29%) were not anemic at 9 mo (p < 0.015). It is apparent from these results that as anemia is prolonged and, consequently, as it increases in intensity, developmental indices also become more adversely affected.

In summary, infants whose anemia had a duration of > 3 mo had significantly inferior scores to anemic infants whose Hb levels were in the normal range 3 mo earlier. Infants without anemia but with marginal iron status at 9 mo became those who eventually, at 12 mo, fell into

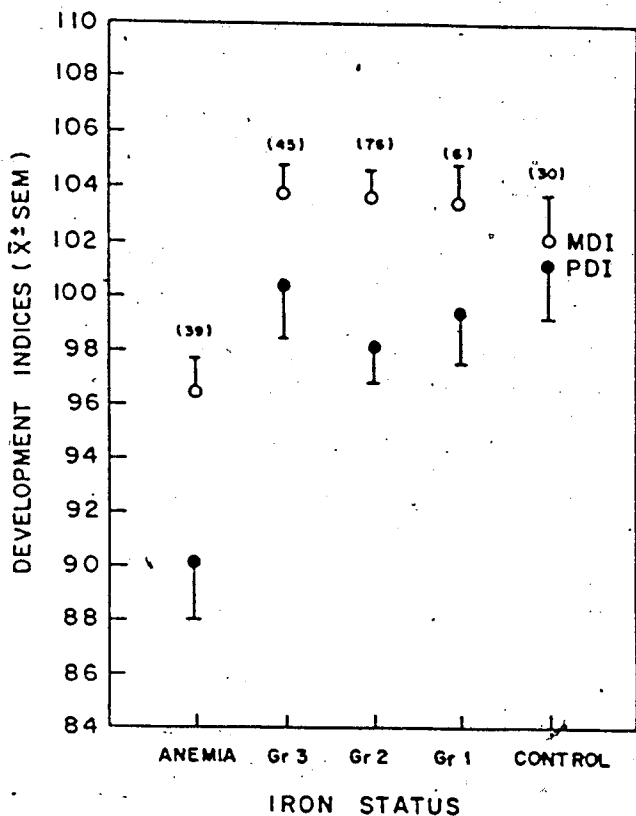


FIG 3. Effect of iron status at 12 mo, including subclassification of NAID (grade 3 = nonanemic responder to the therapeutic trial with a rise of Hb > 10 g/L; grade 2 = nonanemic, nonresponder to the therapeutic trial; grade 1 = nonanemic, iron depleted, only serum ferritin < 10 ng/mL). Neither subclassification of NAID showed differences from control values. All were significantly better than anemic subjects, showing that only anemia has a significant effect on performance of Bayley's Scale of Infant Development.

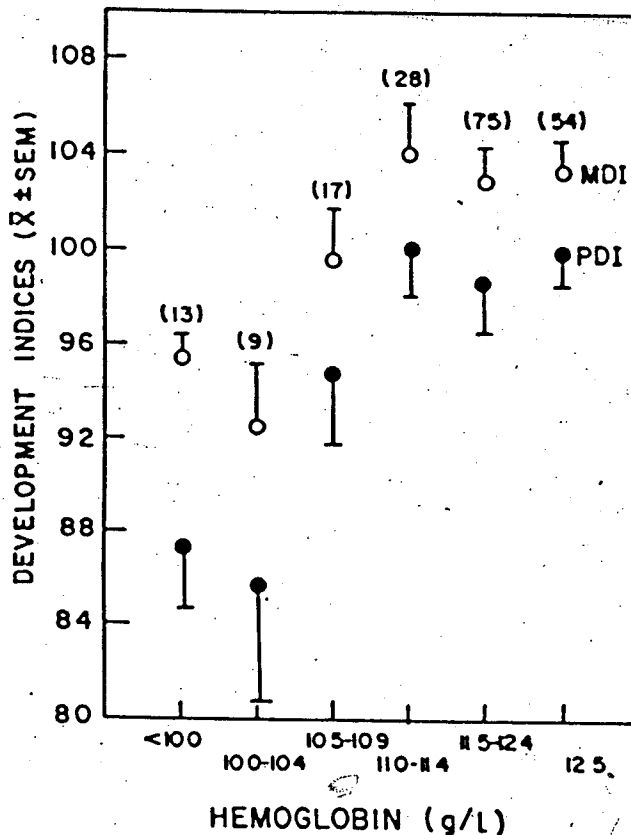


FIG 4. Effect of Hb concentration at 12 mo on Mental Development Index (O) and Psychomotor Development Index (●). Hb at 5-g/L intervals shows a sigmoid distribution of MDI and PDI. No differences of Hb < 104 or > 110 g/L are appreciable. The intermediate point at 105-109 g/L is significantly different from the lower and upper levels ($p < 0.05$) for both PDI and MDI.

the intermediate group (Hb 105-109 g/L), having anemia of shorter duration and lesser severity, associated with milder psychomotor derangements. It is reasonable to assume that if this group of infants (the intermediate anemic group) were to continue with an ID diet, their anemia would increase in duration (and severity) and their psychomotor performance would, hence, deteriorate further. Therefore, infants with anemia of longer duration were also more likely to have anemia of greater severity; these two characteristics of anemia, duration and severity, could thus not be completely individualized with the current design.

Effect of iron therapy

At 12 mo infants given 10 d of oral FeSO₄ were compared with those given placebo. Both groups significantly improved developmental indices in the mental and psychomotor scales to a similar extent, regardless of prior

iron status or therapy with iron or placebo, with average improvements of 4.4-8.9 points. After 3 mo of therapy, even though anemia was reversed in all infants and hematologic measures of iron status were completely corrected in 11 of the 39 formerly anemic infants, no significant improvement was detected between the scores at

TABLE 5
Effect of duration of anemia: development index scores*

	Duration of anemia		p
	9 mo Hb ≥ 105	12 mo Hb < 110	
Mental development index	99.4 ± 2.5	94.1 ± 1.1	<0.05
Psychomotor development index	93.7 ± 3.4	86.0 ± 2.6	<0.05

* Scores are $\bar{x} \pm SD$; p values from Student's t test of group means.
 † Hb < 110 g/L at 12 mo but ≥ 105 g/L at 9 mo.
 ‡ Hb < 110 g/L at 12 mo but < 105 g/L at 9 mo.

mod. Anemia Hb < 105
 mild Hb 105-109
 non anemic Hb > 110

TABLE 6
Mental-scale items at 12 and 15 mo*

Item	Description	Percent infants passing		p
		Anemic	Control	
99	Pushes car along	56	77	NS
101	Jabbers expressively	92	93	NS
102	Uncovers blue box	31	47	NS
103	Turns book pages	69	83	NS
104	Pats whistle-doll (in imitation)	44	63	NS
106	Imitates words (mama, dada)			
	At 12 mo	13	47	<0.01
	At 15 mo	75	100	<0.07
113	Says two words (with meaning)			
	At 12 mo	0	7	NS
	At 15 mo	42	93	<0.005
117	Shows own shoes, toys, clothing			
	At 12 mo	0	18	NS
	At 15 mo	25	60	<0.07

* Analyzed by chi-square test of absolute numbers.

12 and 15 mo. The control and the NAID infants showed no significant change in their MDI or PDI. Neither paired *t* comparison within groups (highest $p = 0.18$ for control infants) nor ANOVA of the differences of scores between groups showed significant changes from the baseline BSID scores.

Short-term therapy, before reversing the anemia, was aimed at correcting immediately the availability of iron to processes of neurotransmitter metabolism that are hypothesized to mediate the behavioral derangements seen during ID (9). In our previous study (10) we found significant improvement after 10 d of oral iron in the anemic infants whereas no difference was noted in the non-anemic infants. However, doubts were raised about the interpretation of the results that did not include placebo-treated infants because neither regression to the mean nor "ceiling" effects could be excluded. In the double-blind study protocol presented here, the increase of scores in iron-treated anemic infants was not significantly greater than that found in placebo-treated anemic infants. The same was true for the other iron-status groups. These findings stress the importance of appropriate controls and show that the plausible explanation for these changes is a practice effect.

After 3 mo of iron treatment many infants had completely corrected their hematologic status and all had improved it markedly; recuperating from anemia. This stage represents tissue replenishment of iron and ensures renewal of those iron enzymes that have turnover times similar or shorter than the Hb turnover time. If these tissue iron proteins were responsible for developmental-

test performance, anemic infants should have approached the scores of their iron-replete peers. Because this did not occur it is probable that other reactions not directly dependent on iron availability are responsible for the derangements seen, that those iron-sensitive behavioral processes may require a longer period of time for their correction, or that they are in fact irreversibly damaged.

Specific patterns of failure

We examined the specific test items involved in the raw score that, when normalized for age, yield the development index. The MDI scores were encompassed representatively by items 99-117 and the corresponding PDI items were 40-52. Examining the mental scale, the items that required comprehension of language but did not involve a visual demonstration were passed by significantly fewer anemic infants than control infants. On the BSID, language development at 12 mo is best marked by item 106, "vocalization of bi-syllabic words," an item that showed marked differences between anemic and control infants ($p < 0.01$). The following language item, 113, "says two words (meaningfully)," was passed by too few infants at 12 mo to be interpretable; however, at 15 mo it became significant ($p = 0.005$). At this age item 117, "shows shoes or other clothing or own toy," also became suggestive ($p < 0.07$). Item 106 continued to show a tendency ($p < 0.07$), partly because at 15 mo most infants passed the item (Table 6).

In the psychomotor milestones 40-52, it appeared that those items relating to balance in the standing position and body control (sits from standing, stands alone, and walks alone) were credited by significantly fewer anemic than control infants. Differences for items 47, "stands up from sitting," and 52, "stands on left foot with help," were not identified at 12 mo but became significant at 15 mo ($p < 0.05$) (Table 7).

TABLE 7
Motor-scale items at 12 and 15 mo*

Item	Description	Percent infants passing		p
		Anemic	Control	
42	Walks with help	85	97	NS
43	Sits from standing	67	97	<0.01
44	Plays pat-a-cake	82	97	NS
45	Stands alone	64	93	<0.02
46	Walks alone	38	67	<0.05
47	Stands up from sitting			
	At 12 mo	3	7	NS
	At 15 mo	42	80	<0.05
52	Stands on left foot with help			
	At 12 mo	0	10	NS
	At 15 mo	8	40	<0.05

* Analyzed by chi-square test of absolute numbers.

Effect of iron status on the infant behavior record

Of the 24 items evaluated by the IBR at 12 mo, 9 were rated significantly better by control infants compared with their anemic peers, as shown on responsiveness to the examiner ($p < 0.03$), responsiveness to the mother ($p < 0.0005$), general emotional tone ($p < 0.04$), goal-directedness ($p < 0.0005$), attention span ($p < 0.001$), activity ($p < 0.001$); as well as responsiveness to persons ($p < 0.001$), to vocalizations ($p < 0.05$), and to body motion ($p < 0.002$).

Further analysis of the IBR was performed, associating behavior items related to test affect and task orientation as suggested by Matheny (11). The test-affect combination rated significantly better in the control vs anemic groups ($p < 0.04$). Those anemic infants with abnormal test affect (14/39) commonly had more scores under the mean for MDI ($p < 0.04$) and PDI ($p < 0.04$) whereas control infants showed no differences, probably because very few (3/30) presented an abnormal-affect behavior pattern. Abnormal task orientation in the anemic children (12/39) was markedly associated with MDIs under the mean ($p < 0.01$); however, only marginal tendency was seen for lower PDIs ($p = 0.09$). Control infants showed nonsignificant associations because only 4 of 30 ranked inappropriate task orientation.

These IBR findings expand our previous experience (10) and those of other investigators (1, 4). There were no appreciable changes of the IBR after therapy, which correlates with the lack of change in development-index scores. This fact lends support to the hypothesis that the unfavorable behavior pattern of these infants may be a mediator of the poorer BSID-test performance. If this contention were true, it should probably have affected individual items in a random fashion. However, in this study item performance failures had a very consistent pattern, with preferences for language in the mental items and body balance-coordination in motor skills. The reason for this selective effect remains obscure. Moreover, these findings coincide quite closely with results from recent research in Costa Rica (12) presented

at this meeting by Dr Lozoff (12), with a different study design, show similar, selective failure patterns. This agreement is reassuring in that it ratifies our findings but worrisome because they also fail to show consistent improvement after short- or long-term iron therapy.

The consistency of these findings, corroborated by Lozoff's study in Costa Rica (1), warrants further investigation of possible persistence of adverse psychologic effects beyond infancy and, ultimately, of the underlying processes mediating these derangements. ■

References

1. Lozoff B, Brittenham GM, Wolf AW, et al. The effects of iron deficiency anemia and iron therapy on infant developmental test performance. *Pediatrics* 1987;79:981-95.
2. Walter T, De Andraca I, Chadud P, Perales CG. Adverse effect of iron deficiency anemia on infant psychomotor development. *Pediatrics* (in press).
3. Oski FA, Honig AS. The effects of therapy on the developmental scores of iron deficient infants. *J Pediatr* 1978;92:21-5.
4. Lozoff B, Brittenham GM. Behavioral aspects of iron deficiency. *Prog in Hematology* 1985;14:23-5.
5. Bayley N. Bayley scales of infant development manual. New York: Psychological Corporation, 1969.
6. Kopp CB, McCall RB. Predicting later mental performance for normal, at risk, and handicapped infants. *Life-Span Dev Behav* 1982;4:33-61.
7. Werner E, Bayley N. The reliability of Bayley's revised scale of mental and motor development during the first year of life. *Child Dev* 1966;37:39-50.
8. Siimes MA, Refino C, Dallman PR. Manifestations of iron deficiency at various levels of dietary iron intake. *Am J Clin Nutr* 1980;33:570-4.
9. Leibel RL. Behavioral and biochemical correlates of iron deficiency. *J Am Diet Assoc* 1977;71:398-422.
10. Walter T, Kovalskys J, Stekel A. Effect of mild iron deficiency on infant mental development scores. *J Pediatr* 1983;102:519-22.
11. Matheny AP Jr. Bayley's infant behavior record: behavioral components and twin analyses. *Child Dev* 1980;51:1157-67.
12. Lozoff B. Methodologic issues in studying behavioral effects of infant iron-deficiency anemia. *Am J Clin Nutr* 1989;50(suppl):641-54.

Comments

Frank Oski¹

In this prospective cohort study of 196 infants studied from birth to age 15 mo the authors conclude that when ID progresses to anemia—but not before—it produces adverse effects on both mental and motor performance of infants as measured by the Mental and Psychomotor Indices of the Bayley Scales. The Infant Behavior Record is also adversely affected by the presence of ID anemia (IDA). Oral iron therapy given for 3 mo failed to reverse

the changes in Bayley scores despite the fact that anemia had been corrected.

A skeptic could ask, what is the evidence that IDA is in fact responsible for the adverse effects described? The alterations in performance, as measured by the BSID tests, are closely related to the presence of anemia and, in fact, the degree and duration of anemia correlate best with the performance measures. There are no data provided on patients with other forms of anemia to demonstrate that this alteration in score is not a nonspecific effect of anemia from any cause. Iron therapy fails to cor-

¹ From the Children's Medical and Surgical Center, Johns Hopkins University, Baltimore.

rect the deficit. It could be argued, in the absence of any other information on control infants, that anemia occurring at a critical period in infancy results in permanent sequelae.

Let us suspend skepticism, however, and assume that the author was studying the effects of IDA. Is there some reason for the failure to discern differences in behavior until the subjects became anemic? It seems highly unlikely that the effects of ID without anemia are without consequence, and no continuum is evident. It can be argued that the test instrument is too insensitive to detect these changes. The classification system may also be at fault. We are told that a rise in Hb of > 10 g/L during a therapeutic trial was considered evidence of ID and that this superseded all biochemical measures. Perhaps these patients were not ID but merely represented replication variation in a laboratory study. We are not provided with any information on the results of duplicate analysis or whether any subjects actually demonstrated a decrease of Hb of 10 g/L between test periods. A misclassification of patients would confound the results and in this specific example would result in the group with true ID without anemia being given higher scores as a result of the inclusion of patients without ID.

The method of scoring might also obscure results during the first 10-d intervention trial. We are informed that the raw scores of the BSID tests obtained at the first and second test administrations were converted to indices based on the child's age at the first interview. This

Comments

Ernesto Pollitt¹

The questions I address in this critique are 1) do Dr Walter's data point to a causal relationship between ID and comparatively low performance in the Bayley Scales and 2) what is the nature of the dependent variables that were assessed? Based on the terminology of Cook and Campbell (1) these two issues can be defined as questions of internal and construct validity.

Internal validity

Three experimental interventions at two developmental periods were included in the study: one at 3 mo and the others at 12 mo and at 12 mo 11 d of age. The first was a fortification intervention involving four groups of subjects: breast-fed, receiving 1) a heme-iron fortified cereal or 2) solid food as recommended by a health clinic; and weaned by age 3 mo, receiving 3) iron-fortified formula or 4) nonfortified milk.

At 12 mo the subjects were subdivided into two groups

method of scoring could result in all infants, both placebo and control, showing spurious improvement between the two testing periods because of the way the scoring of the BSID is constructed. A real response might be obscured. Perhaps no response was observed because ID had not been adequately corrected. We are told that even after 3 mo of therapy only 11 of the 30 initially IDA infants completely corrected their measures of iron status. Unfortunately, all infants, regardless of iron status, were treated for a period of 3 mo, thus obscuring any opportunity to observe intragroup differences.

Finally, it is possible that this was not a homogeneous group of subjects. Some were initially breast-fed whereas others were not. Diets after weaning were not tightly controlled. The infants receiving a non-fortified formula may have developed other deficiencies. The presence of other deficiencies must be considered when ID is not completely corrected after 3 mo. Mejia and Chew (1), in a recent study from Guatemala, reported that the simultaneous administration of vitamin A and iron resulted in a better response of serum iron and transferrin saturation than when the supplement consisted of only vitamin A or iron alone. In some fashion, vitamin A has a beneficial effect on iron metabolism.

Dr Walter and his co-workers are to be congratulated for a comprehensive and ambitious study. The results are provocative and require explanation. ■

Reference

1. Mejia LA, Chew F. Hematological effect of supplementing anemic children with vitamin A alone and in combination with iron. *Am J Clin Nutr* 1988;48:595-600.

exposed, for a 10-d period, to two interventions: one received iron and the other received a placebo. Finally, all subjects received iron for 3 mo.

In terms of the developmental outcome of concern to us, the fortification was the most important intervention because "... most of the anemics (36 of the 39) originated in the unfortified groups" (2). Thus, in order to fully understand the developmental trajectory of these children, it is of critical importance to assess the nature and quality of the design in the 3-mo fortification intervention. Unfortunately, the reports available (2, 3) do not give full information on the field-study design. Dr Walter cited an International Nutritional Anemia Consultative Group (INACG) report for a detailed description of the trial at that time (4). After consulting the text from INACG, I understood that groups a and b, the breast-feeding groups, are part of one study (pp 43-6), where mothers continued breast-feeding up to 9 mo, and that groups c and d were included in a different study on formula fortification (pp 21-9), where mothers weaned their babies before age 3 mo. Thus, based on the INACG report, I am assuming that the two studies are separate

¹ From the Department of Applied Behavioral Sciences, University of California, Davis.

TABLE 1
Correlations between Bayley MDI scores across ages and with 36-mo Stanford-Binet scores*

Age (mo)	Age (mo)			
	3	12	24	36
3	—	0.23† [164]	0.01 [137]	0.06 [118]
12	—	—	0.16‡ [133]	0.12 [116]
24	—	—	—	0.55† [101]

* From reference 8. Values in brackets are *n* values.
† *p* < 0.01.
‡ *p* < 0.10.

events with different sampling and data-collection procedures.

Besides the problems of internal validity (1) related to meeting statistical requirements to pool data from two independent studies, there are at least two powerful covariates that need to be taken into account in the analysis of the data from this study: duration of breast-feeding and home-visiting. The dependent variables in the study are sensitive to these two covariates (eg, see ref 5) and might explain, in part, the between-group differences observed.

Thus, as presently reported, it would appear that the study falls short of meeting the requirements of internal validity. To reject the possibility that differential history between subjects accounts for the between-group differences observed, it will be necessary to know the history of these two covariates in all four groups and to partial out their possible effect on the dependent variables.

TABLE 2
Organization of skills: a second-order factor analysis*

	Factor 1	Factor 2	Variance accounted for†
3-mo factors			
Manipulation	-0.01	0.90	1 = 36%
Social	0.46	0.41	2 = 23%
Search	0.78	-0.18	
Auditory	0.69	0.25	
12-mo factors			
Means-end	0.81	-0.34	1 = 44%
Imitation	0.81	0.33	2 = 37%
Verbal	0.00	0.94	
24-mo factors			
Lexical	0.83	0.03	1 = 35%
Verbal-symbolic	0.79	0.14	2 = 26%
Spatial	0.07	0.76	
Imitation	0.04	0.72	

* From reference 8.
† Totals for the variance accounted for in 3-mo, 12-mo, and 24-mo factors were 59%, 87%, and 61%, respectively.

TABLE 3
Test scores as illustrative of limited infant-variability Bayley scores*

Chronological age	Mental age	Mental-Development Index	Ratio IQ
mo	mo		
12	11	84	92
12	13.5	116	113
12	9	68	75
12	8	52	67
12	21	84	88
24	27	116	113
24	17	68	71
24	15	52	63
6 y†	4 y	68	67
6 y†	3 y 2 mo	52	53

* From reference 7.
† These data are from Stanford-Binet (9); values in the middle column (68 and 52) are for deviation IQ, not MDI.

Whether this is indeed possible given the sample sizes of the four groups and the three iron levels is something that needs to be determined.

Construct validity

Construct validity refers to the possibility that the particular effect postulated in the study under scrutiny—that is, changes in cognitive development produced by changes in iron status—might be construed in terms of other, very different concepts (1, p 59). Cook and Campbell (1) based their discussion on construct validity on, and extended the meaning of construct validity advanced by, Cronback and Meehl (6).

Dr Walter has indicated that the inability of the Bayley Scale to predict later IQ and the fact that its global nature precludes an indication of which aspect of behavior may be influenced by iron lack were not considered enough justification for not using this test. He further stated that the "Mental Scale is designed to assess memory, learning, vocalization and language communication—in general—the basis for abstract thinking and conduct generally associated with intelligence" (2). A critical issue of construct validity, therefore, are the theoretical assumptions underlying the use of the Bayley Scale and whether the effects of iron on the performance in the BSID provide theoretical justifications to infer that ID in infancy has an effect on intellectual operations or mental processes.

In the construction of the developmental scale in question, Dr Bayley did not select the items of the scale because they tap or measure specific information processes or learning abilities. Most items are rooted in the items included in the Gesell Schedule, which were selected from a vast array of observational material and without any theoretical justification. The choice of items from

the Gesell Schedule and the addition of new items were guided by practicality and goodness of age fit. Thus, at age 12 mo, for example, there is no justification to suspect that pushing a car along is more important than lifting it or pulling a cube. Similarly, there is no theoretical reason to assume that closing a round box is more intelligent than intentionally throwing an object to the floor (7).

Accordingly, the lack of predictive validity of the developmental scales is not surprising. Moreover, the low test-retest correlations even during the infancy period are understandable. A study (8) of the development of mental abilities in infancy based on the Bayley Scale reports correlations of 0.23 between tests taken at ages 3 and 12 mo and of 0.16 between 12 and 24 mo (Table 1).

Principal-component analyses at 3, 12, and 24 mo yielded a sensory-motor-manipulation component at 3 mo, a means-end-imitation component at 12 mo, and a verbal-symbolic component at 24 mo (Table 2). None of these factors accounted for > 20% of the variance at any of these three ages.

Moreover, the highest correlation coefficient between components was 0.16. A more detailed factor analyses of the data led the authors to conclude, "Although related sets of skills that changed over development were identified, caution is necessary. First, the factors identified, although agreeing with those reported by others, account for relatively small amounts of variance. Exactly why this appears to be the case in several studies is not clear but the items of the Bayley do not seem to form either a strong coherent principle component or a coherent set of factors" (p 6; pp 350-1).

* Accordingly, there does not seem to be any support for the notion that the "Mental Scale is designed to assess memory, learning, vocalization and language communication—in general—the basis for abstract thinking and conduct generally associated with intelligence" (2). In fact, I would argue that if the BSID is going to be used at

all in studies on ID and development, there is an urgent need to specify the constructs that we are dealing with, rather than assuming that because the Bayley is a popular scale it tells us something important about development.

Finally, I would like to point at the developmental significance of score differences in the Bayley Scale. As observed in Table 3, at 12 mo of age relatively large differences in scores between subjects represent relatively small developmental differences in chronological age. As observed in Table 2, a MDI of 84—that is, 16 points lower than what is expected—is equivalent to a mental age of 11 mo. Similarly, a MDI of 116 is equivalent to a mental age of 13.5.

Thus, although a difference between groups of ~5.7 points is statistically significant, as reported in the Chilean study, we must recall that at age 12 mo this only represents a developmental difference of about half a month. The issue of concern is what is the true developmental significance of the difference. ■

References

1. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin, 1979.
2. Walter T, De Andraza I, Chadud P, Perales CG. Adverse effect of iron deficiency anemia on infant psychomotor development. Pediatrics (in press).
3. Walter T, Kovalskys J, Stekel A. Effect of mild iron deficiency on infant mental development scores. J Pediatr 1983; 102:519-22.
4. Stekel A, Monckeberg F, Beyda V. Combating iron deficiency in Chile: a case study. A report of the International Nutritional Anemia Consultative Group. Washington, DC: International Life Sciences Institute-Nutrition Foundation, 1986.
5. Taylor B, Wadsworth J. Breast feeding and child development at five years. Dev Med Child Neurol 1984; 26:73-80.
6. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955; 52:281-302.
7. Goodman JF. Infant intelligence: do we, can we, should we assess it. In: Reynolds CR, Kamphaus R, eds. Handbook of psychological and educational assessment. New York: Guilford (in press).
8. Lewis M, Jaskir J, Enright MK. The development of mental abilities in infancy. Intelligence 1986; 10:321-54.
9. Psychological Corporation. Stanford-Binet intelligence scale. 4th ed. Riverside, CA: Riverside Publishing Co, 1972.

Response to comments by Ernesto Pollitt

T Walter

Issues of internal validity

At age 12 mo the infants were randomized between treatment with FeSO₄ or placebo. However, these infants had been the subject of two previous interventions, one that was random (iron-fortified foods or nonfortified products) and one that was not random, whether they were on breast milk as the exclusive source of milk (breast-milk group [BF]) or were receiving > 50% of their estimated caloric requirements from sources other than breast milk (early weaned group [EW]) at the time of the fortification assignment.

At this point, let us describe the experimental setting in more detail. All the infants in this study were obtained

from the same clinic, at the same time, and with the same criteria and were treated in a standardized fashion. The study was carried out over a period of 2.5 y and the infants were assigned to a food product (with or without iron fortification) in an alternate fashion, according to order of recruitment to the study. The nonrandom stratification applied was the breast-feeding history, over which the investigators had no influence whatsoever. However, once the infant became part of the study, efforts to maintain or reinstitute breast-feeding were routinely carried out by the staff caring for these babies, as will be appreciated by the duration of breast-feeding in the EW infants. It is also important to note that the food products normally handed out by the National Food

TABLE 1
Fortification assignment in breast-fed and early-weaned infants according to iron status

	+Fe	-Fe	+Fe	-Fe	+Fe	-Fe
Breast-fed (n = 104)	10	8	34	34	12	6
Early-weaned (n = 92)	2	19	26	33	9	3

Supplementation Program, to which all these infants belonged, are not iron fortified. Thus, we were in fact not experimentally removing a benefit but rather experimentally adding a well-recognized potential need.

The fortification randomization assignment determined to some extent the iron status of the infant at age 12 mo. Iron fortification was more efficient in the EW group because milk is consumed in large and fairly predictable quantities throughout the first year of life. Thus, only 2/38 (5%) of infants in the EW fortified group became anemic, whereas 10/55 (18%) of the BF fortified group had a Hb < 110 g/L at 12 mo (Table 1).

The distribution of treatment with iron or placebo was fairly even, except for a large number of EW anemic infants getting iron vs placebo (Table 2). This skew potentially stems from the fact that the randomization between iron and placebo was determined with only the first available lab data, and the final classification of iron status was delayed until all the lab results were in, including the response of the 3-mo therapeutic trial.

To adequately interpret the result of the iron-placebo therapy trial, it is important to evaluate the influence of breast-feeding history on the performance of the BSID.

Infants in the BF group received breast milk as the exclusive source of milk for an average (\pm SD) of 275 ± 78 d and nonexclusively for 334 ± 55 d. Infants in the EW group had breast milk exclusively for 74 ± 61 d. By definition, no infant in this group was exclusively breast-fed for > 120 d; however, most continued to receive the breast for 171 ± 120 d. Thus, even though the differences between the duration of breast-feeding in the BF and EW groups are large and highly significant, all the infants in this study had ≥ 60 d of breast-feeding, and even in the EW group some were breast-fed—although not exclusively—for as long as 9 or 10 mo.

TABLE 2
Short-term-therapy trial assignment in breast-fed and early-weaned infants according to iron status

	+Fe	Placebo	+Fe	Placebo	+Fe	Placebo
Breast-fed (n = 104)	8	11	37	31	7	10
Early-weaned (n = 92)	13	7	29	30	5	8

TABLE 3
Mean Mental Development Index scores in breast-fed and early-weaned infants according to iron status

	Breast-fed	Early-weaned
Anemic	98.2 [19]*	95.1 [17]
NAID	105.9 [68]	100.9 [57]
Control	102.3 [17]	99.7 [11]
Weighted mean	103.8	99.7
ANOVA	<0.05	<0.05

* Number of infants in brackets.

We found no correlation between the duration of breast-feeding, be it exclusive or not, and developmental scores. Separating the infants into BF and EW or accumulating them, or taking breast-feeding duration in days or stratifying it as more or less than 4, 6, 8, or 10 mo did not influence the significance of this operation. However, when the development scores are examined at 12 mo, there is an appreciable difference between the BF and EW, favoring the former group (Table 3). However, the difference in MDI and PDI (Table 4) between the iron-status groups, showing that anemic infants have a poorer performance than all nonanemics infants, persists unaltered. Thus, even though the BF infants have overall better scores than the EW group, this effect is overcome by a still-persisting effect of iron status.

The next step was to examine the performance of the infants after 10 d of iron or placebo. Infants were analyzed as EW and BF, each segregated according to iron status (thus, six groups). After 10 d all six groups improved their performance significantly both in MDI and PDI (paired *t* test). The analysis of whether some infants improved their scores more than others failed to show differences (ANOVA), showing that neither breast-feeding nor iron status, or iron vs placebo, have an influence on developmental indices that cannot be explained as a practice effect (data not shown).

Finally, stepwise multiple-regression analysis was executed with MDI or PDI as dependent variables, including duration of breast-feeding and all the other intervening variables collected. By forcing Hb in first, all other variables lost significance when 0.2 was used as the limit-of-significance cutoff to enter the stepwise equation.

TABLE 4
Mean Psychomotor Development Index scores in breast-fed and early-weaned infants according to iron status

	Breast-fed	Early-weaned
Anemic	92.1 [19]*	87.3 [17]
NAID	100.3 [68]	97.1 [57]
Control	102.6 [17]	98.8 [11]
Weighted mean	99.1	95.4
ANOVA	<0.0002	<0.01

* Number of infants in brackets.

Issues of construct validity

All of the comments put forward by Dr Pollitt are fully endorsed and irrefutable. We agree that better methods of probing mental processes in infancy must be used in any future protocol in this area of research. Some of the techniques mentioned by Dr Frances Horowitz earlier in the meeting are certainly worth looking into.

However, other neurophysiologic measures searching for subtle maturational effects of iron lack should also be incorporated into this line of research in the next generation of studies. Sleep-wakefulness-cycle studies seem warranted, based on the patterns of sleep inversion found

in experimental animals. Measures of myelination of peripheral and central nervous system pathways and synaptic activity found altered in the parietal cortex in the rat with methods such as brain-stem evoked potentials are noninvasive and need to be further evaluated.

Notwithstanding, the Bayley Scales are the only tool that has been looked at extensively and such coincidences as have been found by independent investigators working in distant regions and with different study designs cannot be ignored. The BSID, therefore, have become the gold standard to which all future research in the effect of ID in infancy must be compared to, until a better tool becomes available. E