

Conceptual and methodological issues regarding the epidemiology of iron deficiency and their implications for studies of the functional consequences of iron deficiency¹⁻³

George H Beaton, Paul N Corey, and Cathy Steele

ABSTRACT This paper reviews the measures of iron status (hemoglobin, hematocrit, mean corpuscular hemoglobin concentration, mean cell volume, free erythrocyte protoporphyrin, serum iron, transferrin saturation, and serum ferritin) that are potentially available for inclusion in field studies of the relationship between iron and mental performance. The characteristics of these measures (sensitivity to iron status, specificity to iron, and diurnal and day-to-day variability) are reviewed and the implications of choice of variable for the design, analysis, and interpretation of studies are discussed. Brief consideration is given to the question of confounding variables and to sources of both false-positive and false-negative conclusions. The explicit message of the paper is that there is no perfect choice of measure of iron status but, given explicit definition of the research question, there are preferred choices that can most effectively combine the choice of variable and the design of the study. *Am J Clin Nutr* 1989;50:575-88.

KEY WORDS Iron, iron status, cognition, psychology, epidemiology

Introduction

The purpose of this workshop is to consider whether the relationship between iron nutritional status and cognitive development is known and, if not, to establish working hypotheses about the relationship(s) that can be subjected to explicit testing. Within that framework, the purpose of this paper is to consider those features of iron and of measures of iron status that may influence in important ways both the interpretation of past studies and the design of future studies. A later paper will address possible confounding variables. At this stage of the proceedings it is not clear whether the relevant design will be epidemiologic (observational) or experimental (interventionist). An important consideration in this regard will be the nature of the hypothesis and, more particularly, the question of whether it is current iron status that is deemed important or iron status at an earlier stage of growth and development. Given that these aspects of design have not been defined, the approach that will be taken in this paper is first to review some of the measures of iron status that might be included in studies of the functional outcome of iron deficiency (ID) and then to consider the implications of the various measures when they are used in epidemiologic and in experimental designs.

Available measures of iron status

Extensive reviews of the available measures of iron status have been published (1). Here we need be concerned

only with a few that have potential application in field studies of human populations. All measures discussed can now be applied to microliter samples of blood (2) and hence are minimally invasive and are potentially applicable to children, even infants. The list is effectively limited to hemoglobin (Hb), hematocrit, red cell count, and their various indices; free erythrocyte protoporphyrin (FEP); serum iron, serum transferrin, and transferrin saturation; and ferritin. These subdivide into measures that assess the adequacy of body iron supplies in relation to erythropoiesis and those that more closely relate to total iron storage. Each of the groups of measures has certain advantages and certain disadvantages as measures of iron nutriture. If and when a specific compartment of body iron is hypothesized to be causally linked to mental function, then indices of the particular compartment would be preferred.

In this paper the terms *sensitivity*, *specificity*, *validity*, and *reliability* are used with the following meanings. Sensitivity is taken to imply the degree to which the measure reflects the status of iron stores (iron depletion).

¹ From the Departments of Nutritional Sciences and Preventive Medicine and Biostatistics, Faculty of Medicine, University of Toronto, Toronto.

² Supported in part by grant 606-1468-41 from Health and Welfare Canada.

³ Address reprint requests to GH Beaton, Department of Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada M5S 1A8.

Specificity is taken to mean the degree to which the measure is unaffected by other factors and hence is a specific indicator of iron depletion or deficiency. Validity is taken to mean the assurance that the method used is actually measuring the variable in question. Reliability is usually taken to refer to both measurement error (the reproducibility of repeated measurements of the same sample) and within-person variation (reproducibility of the results from repeated samples collected at different times, within a day or across a period of days, from the same subject). In this paper, the components of reliability are discussed separately and the term is not applied in a generic sense.

Hemoglobin, hematocrit, red cell volume, and calculated indices

If the question being asked relates to the presence of ID anemia (IDA) or oxygen-carrying capacity of the blood, one or more of these measures is essential. In general population studies, Hb and hematocrit offer similar information, at least across the range of mild anemias. Anemia, marked by low Hb, is not specific for ID. Use of the combination of measures and the calculation of mean corpuscular hemoglobin concentration (MCHC) and mean cell volume (MCV) increase the specificity but not the sensitivity. In populations in which other forms of anemia are expected to be present, even these measures may offer insufficient specificity, and FEP values as well as measures of iron stores become very desirable (3, 4). This, of course is a diagnostic or classification issue rather than one of assessment of relative iron status. In the context of the present studies, such differential diagnoses take on more importance in explaining nonresponse of red cell indices to an iron intervention than in serving as a measure of iron nutriture in analyses of the association between iron depletion and functional outcome.

A small diurnal variation is seen in both Hb and hematocrit (Table 1). It is not significant in either the biological or statistical sense. Interestingly, the changes are not proportional to each other and there is a small reverse diurnal variation in MCHC. More important for field studies, the method of collection of the blood sample can affect the hematocrit and hence the Hb level. Both time of day and technique of collection should be standardized if the measurement errors are to be minimized. Because cell volume can change with holding time, precautions should be taken to determine hematocrit shortly after collection. Hb and hematocrit values differ with altitude. Although this presents a problem for classification of anemia, it should add no specific problem in either cross-sectional or longitudinal comparisons carried out at the same altitude.

There are four very serious limitations of Hb and hematocrit for the studies at hand. The first is relative insensitivity to iron depletion. Hb and hematocrit are measures of a severe deficiency of iron (1). If the question being considered relates to the association between iron depletion and functional outcome, then these measure-

ments may be necessary background information but are unlikely to be of major importance in analyses. Conversely, if the question relates to the effect of IDA on cognitive development, they are the critical measures and indices of iron stores are simply confirmatory. The second major issue for studies of underprivileged groups in developing countries is the nonspecificity of Hb levels. In particular, levels are reduced by the presence of chronic infection or inflammation regardless of iron status. A third very serious concern is the fact that there is a wide interindividual variability in physiologically normal Hb levels. Normality of the distribution has been reported for replete subjects (6-8). In the Nutrition Canada data (5) the Hb distributions for both males and females approximated the Gaussian except for the lowest few percent of the population, where true anemia may be exerting its effect. The coefficient of variation (CV) of Hb was 9-10%. As a special case of this interindividual variability, there is also clear evidence of a genetic difference between blacks and whites (8-11). The fourth problem, illustrated in Table 2, is that in children Hb and hematocrit change markedly with age and exhibit sex differences in the pattern of change. It is interesting to note that the Hb level in girls does not fall during adolescence. Rather, it fails to rise. In free-living adults, it appears that Hb levels are relatively age-stable in women but progressively decrease with age in men. This pattern is very similar to that reported by Yip et al (12) in US survey data.

The interindividual variability in normal Hb presents extremely serious problems in any approach that involves classification of individuals as anemic or nonanemic (6-8, 13, 14). The relatively rapid and sex-specific changes with age present serious problems for classification criteria (15) and for longitudinal studies in children (a rise in Hb level is expected with time). The occurrence of infectious disease can confound both cross-sectional and longitudinal comparisons. These considerations essentially limit effective use of these measures to experimental designs where change in the measure within the individual across the period of intervention can be examined and used as a variable (thereby circumventing in major part the problem of interindividual variation in normal Hb). Inclusion of a matched control group reduces the problem of expected trends with increasing age in longer-term studies. Obviously, in any group comparisons, age and sex must be matched or an adjustment procedure must be included in the analyses. Even in such designs it may be necessary to obtain some measure of the occurrence of infectious disease, particularly if there is any reason to believe that the intervention can itself affect morbidity or that these rates differ across populations being compared. (The concern with morbidity as a confounding factor applies to all of the measures under discussion).

Free erythrocyte protoporphyrin

FEP levels tend to rise as the supply of iron for erythropoiesis is constrained. In that sense they mark the same phenomenon as a decrease in hemoglobin. However, be-

TABLE 1
Observed diurnal variation in measures of iron status in adults aged 21-60 y examined in the Nutrition Canada Survey*

Group and time interval	n	Hb g/L	Hematocrit	MCHC g/L	Serum iron μmol/L	Transferrin saturation %
Adult males						
0930-1030	14	157.3	0.4836	325.2	20.48	33.14
1030-1130	33	154.3	0.4651	331.7	19.71	31.91
1130-1230	8	159.9	0.4800	333.0	18.62	29.13
1230-1330	131	156.9	0.4717	332.7	21.06	33.03
1330-1430	219	155.6	0.4656	334.2	19.59	31.48
1430-1530	156	156.4	0.4679	334.4	20.17	31.22
1530-1630	138	156.4	0.4682	333.9	18.95	30.40
1630-1730	301	156.4	0.4658	335.8	18.24	28.58
1730-1830	266	156.4	0.4608	339.6	17.76	27.65
1830-1930	56	152.2	0.4538	335.4	17.08	27.22
1930-2030	955	155.2	0.4551	341.1	16.80	26.57
2030-2130	137	154.4	0.4561	338.5	17.16	26.74
Adult females						
0930-1030	18	137.8	0.4317	319.2	19.33	31.11
1030-1130	71	138.7	0.4186	331.6	19.41	29.56
1130-1230	26	135.7	0.4146	326.9	19.16	27.23
1230-1330	179	136.8	0.4182	327.1	18.40	28.00
1330-1430	455	137.2	0.4194	327.2	18.04	27.19
1430-1530	421	136.8	0.4183	327.3	17.92	26.57
1530-1630	403	136.6	0.4149	329.5	17.29	26.00
1630-1730	454	137.0	0.4166	335.0	16.10	24.24
1730-1830	333	136.5	0.4129	330.6	15.47	23.16
1830-1930	67	134.2	0.4145	323.9	14.46	21.70
1930-2030	709	136.5	0.4103	332.7	14.72	22.35
2030-2130	95	136.8	0.4082	335.2	15.52	23.36

* Nutrition Canada (5) was a population survey with a stratified random design. Analyses presented in this paper have not taken into account sampling weights. Other analyses established that there was no consistent relationship between the diurnal change and body weight, reported iron intake in the previous 24 h, or deciles of age.

cause FEP levels are very low in the adequate situation, the rise tends to be a sharper marker than is a decrease in Hb. FEP is seen as an early measure of the effect of iron depletion on erythropoiesis, somewhat more sensitive than Hb (1, 16). FEP levels rise in the presence of infection, just as Hb levels decrease, even when body iron stores are adequate (17). In separating these effects from those attributable to iron depletion, FEP offers no real advantage. FEP levels decrease with age during childhood (12) and rise in the presence of lead toxicity (18). A difficult and time-consuming methodology delayed general acceptance of this measure. Even though current methods are more feasible in application, experience with this measure in studies of the type under consideration remains limited. The general impression one obtains from the current literature is that although FEP may be useful in a multifaceted diagnostic criterion of anemia, it offers little as a measure of the relative level of iron stores in the body. Recently, the within-person variation in FEP levels across days was examined in US adults (Table 3). It would appear that a distinct advantage of FEP is its stability within the individual (in con-

trast to the serious problems that arise for serum iron and transferrin saturation).

Serum iron, transferrin, and transferrin saturation

For many years the preferred technique of assessing iron status, in the absence of obvious anemia, was by measurement of transport iron in the serum. Much evidence was accumulated to demonstrate that serum iron levels fall before impairment of the hematopoietic function leads to a detectable deficit in circulating Hb (1, 4). In the range of modest deficits in Hb, the serum iron variables respond much more sharply than does Hb. Across this range of major depletion to moderate anemia, these measures appear quite sensitive to the body supply of mobilizable iron. At higher levels of body iron, serum iron is still responsive but with much less sensitivity.

The relative patterns of change in serum iron and Hb are consistent with the notion that they reflect at least two different body iron pools that deplete differentially. A detailed examination of Nutrition Canada data served to demonstrate the nonlinearity of the relationship (Fig 1) (generally similar patterns hold if serum iron is used

TABLE 2
Regression of iron variables on age for data from the Nutrition Canada Survey

Variable	Age interval	Males			Females		
		n	Regression	SE*	n	Regression	SE
Hb (g/L)	0-13	2148	119 + 1.4 y	0.06	1867	121 + 1.1 y	0.07†
	14-20	1064	94.3 + 3.4 y	0.02	1025	131 + 0.28 y	1.3‡
	21-	4097	162 - 0.17 y	0.03	4958	134 + 0.06 y	0.01‡
Hematocrit	0-13	2123	0.359 + 0.0038 y	0.0001	1841	0.363 + 0.0034 y	0.0002†
	14-20	1055	0.288 + 0.0096 y	0.0005	1010	0.396 + 0.00075 y	0.00045‡
	21-	4053	0.472 - 0.00026 y	0.00003	4903	0.408 + 0.00019 y	0.00003‡
MCHC (g/L)	0-13	2098	332 + 0.28 y	0.12	1821	333 + 0.02 y	0.13
	14-20	1041	332 + 0.25 y	0.29	998	330 + 0.21 y	0.48
	21-	4017	345 - 0.18 y	0.02	4849	330 + 0.01 y	0.02‡
Serum iron ($\mu\text{g}/100\text{ mL}$)	0-13	1636	14.7 + 0.14 y	0.036	1387	14.9 + 0.13 y	0.039
	14-20	1045	11.3 + 0.44 y	0.084	1012	16.4 + 0.0007 y	0.090‡
	21-	4040	19.1 - 0.27 y	0.005	4847	17.5 - 0.021 y	0.004
Transferrin saturation (%)	0-13	1636	21.3 + 0.30 y	0.05	1387	22.6 + 0.19 y	0.06
	14-20	1044	14.5 + 0.81 y	0.13	1012	24.4 - 0.019 y	0.13‡
	21-	4040	29.0 - 0.012 y	0.007	4846	24.2 + 0.021 y	0.007‡

* Standard error of the slope of regression. Regressions are unweighted.

† Slopes are significantly different between males and females, $p < 0.01$.

‡ Slopes are significantly different between males and females, $p < 0.001$.

in place of transferrin saturation and if hematocrit or MCHC is used in place of Hb). In these analyses individuals were clustered by $0.18\ \mu\text{mol/L}$ ($1\text{-}\mu\text{g}/100\text{ mL}$) units of serum iron and the mean Hb level was calculated for each cluster (for the 4056 females between ages 20 and 70 y, there were 185 such clusters; for the 3196 males, there were 192 clusters). Weighted-regression techniques (with number of subjects in each cluster as the weighting factor) were applied to fit the nonlinear, asymptotic model. The derived estimates of the parameters of the model are presented in Table 4. At lower levels of serum iron and at the higher levels, the number of subjects included in a cluster were small, often a single subject. To test whether aberrant values in this range might be exert-

ing undue effect, algorithms were developed to group the lowest 32 (25 in females) serum iron levels into intervals of 0.56, 0.72, or $1.43\ \mu\text{mol/L}$ (2, 4, or $8\ \mu\text{g}/100\text{ mL}$). Similar reclusterings were done for the upper 72 (64 in females) clusters. In total this yielded 16 new data sets. The estimates of A, B, and C were quite stable across these data sets.

A segmented linear-regression model was also fitted to the data. This model assumed a quadratic relationship between serum iron and Hb until a threshold value of serum iron is reached, after which a plateau in Hb was achieved. The results of fitting this model are shown in Table 5. The proportions of variance explained (R^2) were almost identical with those shown in Table 4. The

TABLE 3
Predicted impact of repeated measures on attenuation of correlation and regression analyses in data from Hispanic HANES adults*

Iron measure	Sex	Variance ratio†	Simple correlation		Bivariate regression	
			One measure	Three measures	One measure	Three measures
Erythrocyte protoporphyrin	M	0.05	0.98	0.99	0.95	0.98
	F	0.06	0.97	0.98	0.94	0.98
Serum iron	M	1.9	0.59	0.78	0.34	0.61
	F	3.9	0.51	0.66	0.20	0.43
Transferrin saturation	M	2.2	0.56	0.76	0.31	0.58
	F	1.4	0.65	0.83	0.42	0.68

* $n = 31$ males and 47 females, each with two measures.

† Variance ratios for logarithmically transformed data reported in reference 19.

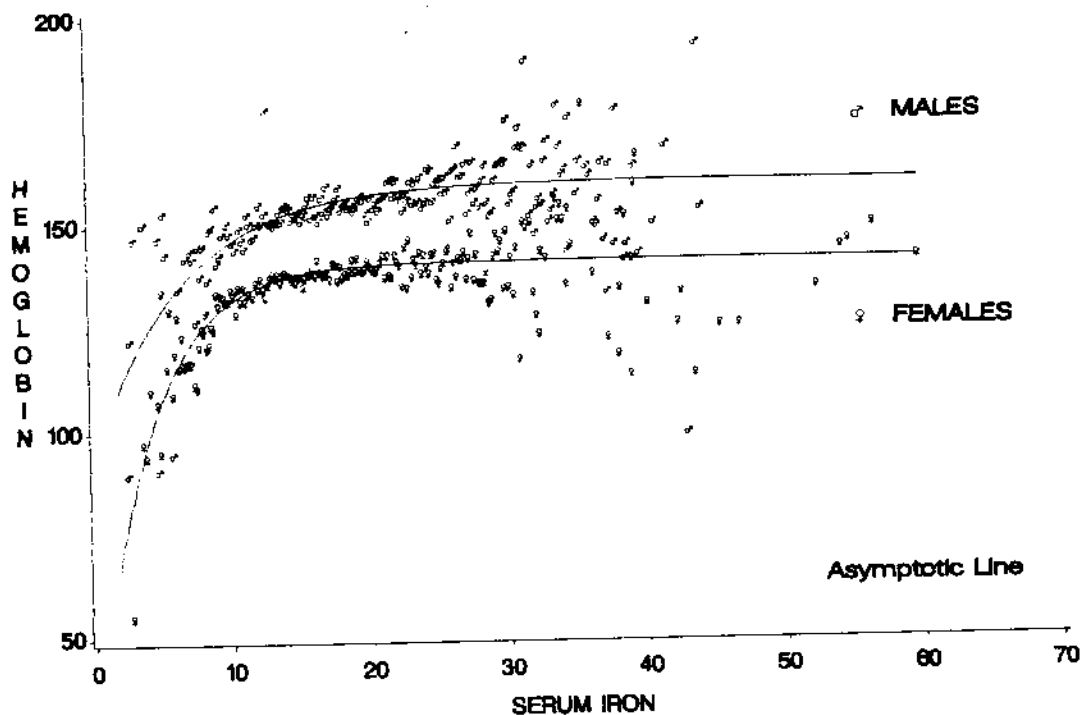


FIG 1. Observed relationship between Hb (g/L) and serum iron ($\mu\text{mol/L}$) in adults aged 20–70 y examined in Nutrition Canada (not controlled for diurnal variation or the small age effects seen in adults). Mean Hb levels for 0.18- μmol (1- $\mu\text{g/mL}$) intervals of serum iron are plotted. See text for fitted regression parameters. The lines shown are for the asymptotic model.

threshold values were 24.01 and 16.56 μmol iron/L serum for males and females. In the segmented linear-regression models the maximal Hb for males was 159 g/L compared with 160 for the asymptotic model. For females the values were 139 and 141 g/L, respectively.

Both models suggest that females reach their maximal Hb level at a lower serum iron level than do males. These results suggest that there is a true sex difference in physiological normality for serum iron just as there is for Hb, even though this is not commonly recognized in establishing diagnostic cut points. The display of data in Table 2 suggested that the relationship between serum iron and age was very similar in male and female children. It is

reasonable to assume that the true sex difference begins at or around menarche and hence is not a matter of major concern in studies focusing on children. Conversely, the analyses illustrate that Hb and related measures are insensitive to iron status except when the storage pool has been markedly depleted. It must be assumed that a similar relationship holds for children.

The strong perceived advantage of serum iron as a status measure is the increased sensitivity to iron depletion through the range representing major depletion through moderate anemia—a range that is of interest in studies of the association between iron and cognitive function. Initially, there were difficulties in standardizing laboratories and achieving acceptable validity of measures. These were overcome and measurement of serum iron and serum transferrin and thus calculation of transferrin saturation are now acceptable although it is clear that care must be taken (15). Current methodologies permit estimations with very small samples (2).

The serious limitation to the use of serum iron is its notorious variability both within days and across days in the same individual (1, 20–23) (see Table 3). These are mediated in part by shifts in iron release (24, 25), with iron uptake being influenced by the plasma level. Other, imperfectly identified influences play a major role. Because iron status is only one of these influences, specificity is a problem. Expressing serum iron in proportion to

TABLE 4
Relationship between hemoglobin and serum iron in adults*: A—
asymptotic model: $\text{Hb} = A - B \times e^{-C \times \text{SI}^\dagger}$

Sex	n	Parameter estimates			R^2
		A	B	C	
Males	3196	160 ± 0.9	66 ± 10.4	0.156 ± 0.020	0.53
Females	4056	141 ± 0.4	116 ± 15.4	0.251 ± 0.018	0.71

* Subjects aged 20–70 y, drawn from Nutrition Canada data.

† SI = serum iron.

TABLE 5
Relationship between hemoglobin and serum iron in adults*: B—
segmented regression model: if $SI \dagger < SI_0$, $Hb = A + B \times SI + C \times SI^2$;
if $SI \geq SI_0$, $Hb = D$

Sex	n	SI ₀	Parameter estimates				R ²
			A	B	C	D	
Males	3196	24.01	121.7	3.07	0.0654	159	0.50
Females	4056	17.03	83.1	6.81	0.2055	139	0.69

* Subjects aged 20–70 y, drawn from Nutrition Canada data.
† SI = serum iron; SI₀ = threshold value of serum iron.

the carrying capacity of the transport protein, transferrin, improves the situation somewhat but does not remove the problem (see Table 3).

As reported by many authors and portrayed in Table 1, diurnal variation is a major systematic effect for both of these measures. The cycles are not necessarily in phase across individuals; nevertheless, in total groups a clear pattern emerges. For the data in Table 1, simple linear regressions of half-hour-interval means (unweighted) on time yielded the following estimates of changes across the 0900–2100 period. As noted in Table 1, these could not be related to iron intake in previous day, body weight, or decile of age.

Serum iron for adult males

$$= 26.51 - 0.478 h; R^2 = 0.83 \quad (1)$$

Serum iron for adult females

$$= 25.66 - 0.545 h; R^2 = 0.90 \quad (2)$$

Transferrin saturation for adult males

$$= 42.23 - 0.78 h; R^2 = 0.86 \quad (3)$$

Transferrin saturation for adult females

$$= 38.50 - 0.81 h; R^2 = 0.90 \quad (4)$$

Whereas for Hb and hematocrit the range of diurnal variation represented 3–5% of the early-morning values, for serum iron and transferrin saturation the range was ~20% of morning levels. In women, oral contraceptive usage was another systematic factor. (This can be seen for both serum iron and transferrin saturation even after subjects are matched for menstrual blood loss to avoid differences in iron status [26]). Age- and sex-related differences are present (12, 27). The display in Table 2 suggests that although these changes present difficulties for classification criteria, they would present less of a problem in cross-sectional or longitudinal analyses than would Hb up to puberty. In the adolescent years again the levels rise sharply for boys but not for girls. Diurnal variation is seen in children as well as in adults (22). Infections, even minor infections, depress the levels quite

sharply (1), creating potential problems of interpretation of field data.

The major concerns with these measures, then, are 1) wide diurnal variation (which could be controlled); 2) wide variation within persons from day to day; and 3) nonspecific response to several other influences (which would be difficult to control under field conditions).

Serum ferritin

In the last decade or two, serum ferritin has gained acceptance as the preferred index of iron storage. Methodologic problems have hindered the validity of measurements done in different laboratories although this problem has been greatly reduced in recent years. It appears that this measure is directly related to the level of storage iron in normal subjects (28). It has been suggested that a difference of 1 $\mu\text{g/L}$ in serum ferritin is equivalent to ~8 (29) or 10 (30) mg of storage iron in normal adults. Serum ferritin level is sensitive to tissue iron level across a substantially greater range than any of the other indices and it is the only index that appears to be useful as an index of iron nutriture at tissue levels above those marking the beginning of impairment of hematopoiesis. This measure has the further advantage of having an apparently smaller day-to-day variation than serum iron but it does increase in response to infection (30). (Note that this is the only variable examined that moves in the opposite direction to that seen in ID during infection.)

Serum ferritin levels are high at birth, decrease rapidly during the first year, and then climb slowly through adolescence (31). As for other variables of iron, there is a major increase in males. For ferritin it appears to take place during the late adolescent and early adult years. There is then a progressive rise with increasing age. Females do not show this adolescent increase and levels tend to increase slowly to about the age of menopause and then increase relatively sharply (31, 32). The increases with age in older persons may represent accumulation of stores or be the effect of progressive development of inflammatory disease (32).

Implications for interpretation and design

In the foregoing discussion, emphasis was placed on three dimensions of the characteristics of measures of iron status: their sensitivity to change in iron status; their specificity as markers of iron status; and their variability within the day or across days (reliability as markers of iron status). The assumed analytical question is, What is the relationship between iron status and functional outcomes, in this case cognitive function? In this conceptual model, iron status is not measured directly and all the variables under discussion are proxy measures or estimators of the reality that is of interest. That being the case, there is yet another important dimension of the properties of the measures that warrants consideration—proximity to the true measure of iron status.

The analytical question underlying this paper, and indeed this workshop, is, What is the relationship between ID and brain function? To proceed further in discussing the utility of the various markers of iron status it is necessary to pause and consider the explicit meaning of ID. In the foregoing discussion three dimensions of the characteristics of available measures were considered: their sensitivity to changes in body iron stores, their specificity to changes in iron status as contrasted with other factors such as infection, and their variability within days and across days in the presumed absence of real changes in body iron stores. To this point it has been assumed that ID and *reduced body iron stores* are synonyms.

If this conceptual identity is accepted then a critically important issue arises. None of the measures discussed directly estimates the true reality under study. They are proxy measures, indirect estimators, indicators, whatever term you may wish to apply. In the foregoing discussion we have examined the degree to which the measures appear to relate to the reality—in group data. What must now be recognized is that there is every reason to accept that, among individuals, there is interperson variation in that relationship. This is clearly illustrated by the widely accepted interindividual variation in physiologically normal Hb. For a particular individual, achievement of his or her physiologic norm implies that body iron levels have risen to the point that hematopoiesis is no longer compromised—and that either the level of stores needed for this purpose differs between individuals or the hematopoietic rate associated with a given level of stores differs between individuals. Either way, we must accept that there is individual variation between the estimator of iron status, in this case Hb, and the reality of body iron stores. It is but an extension of the same logic to suggest that analogous variability exists in the relationship between measures of serum iron and true body stores, and between serum ferritin and true body stores. It also follows that, given a fixed state of the true reality, there must be interindividual variation in the relationships among estimators of that reality (33). Implicit recognition of this issue is to be found in the various efforts to estimate the prevalence of ID (eg, 3, 8, 9, 15, 32). These break down into two broad approaches: those that have used a single measure and then use a distributional approach to estimate the proportion of low values for the variable (whether that be a statistical estimator of shift in distribution, a probability approach, or the application of a fixed cutoff) and those that attempt to apply multiple criteria, each with its own cutoff, to diagnose the individual and then count the number of deficient. The important element to recognize is that in each case there is an implicit acceptance of the notion of individual variability of the marker in relation to the reality under investigation and of the notion that these interindividual variabilities may be independent, or semi-independent, across markers. This has very important analytical implications.

A critical question that must be asked in the planning

of analysis and interpretation of studies relating to iron and human function is, Is there need for, or advantage to, a classification of individuals? Should we dichotomize data or treat the measures as continuous variables? There is a long tradition of the development of classification criteria marking stages of ID. Undoubtedly this has arisen from the model of diagnosis and treatment of anemia in health care practice. Over the years the definitions have become more refined and more parameters have been included in the criterion. However it is also true that over this period there has been continuing concern about the recognized misclassification of individuals. Given the interindividual variation discussed above, coupled with the sources of within-person variation discussed earlier, it is inevitable that the classification of individuals will be very imperfect. In addressing the current questions, misclassification of individuals in a dichotomized (or trichotomized) classification will seriously compromise the ability to demonstrate effects of iron status. In most analytical situations there is little or no advantage to classification of a continuous variable and indeed there is often a loss of statistical power. This reviewer would agree wholeheartedly with the concept espoused by Dallman (4) when he said, "In relation to examining the role of iron in various aspects of behavior, the operational definition of ID, provided by the response of Hb concentration to iron administration, is a useful adjunct to the use of currently available iron tests." This reviewer would only extend the concept to all of the variables of iron. It is the approach of paired analyses of the impact of an intervention on the individual. It is reasonable to assume that the relationship between change in iron status and change in functional performance will show less interindividual variation than will the relationship between status and function.

There is a problem. The relationship between our variables of iron status and iron status are not linear. Indeed in the case of the red cell measures (Hb, hematocrit, and MCHC) the relationship is truncated—there is a maximum limit in response marked by the individual's physiologic, homeostatic norm (6). As was illustrated in Figure 1, the relationship among parameters changes across presumed levels of iron nutrition.

It follows that we should recognize that the preferred estimator(s) of iron status will differ with the range of status in which we are interested and that the relationship between the chosen indicator of status and a functional outcome will differ between measures as we move through the range of status. Surely this is a part of the confusion in the literature when investigators have used different measures of iron status and have applied them in populations having different ranges of iron nutrition. Some of these designs have to have been mismatches.

From the earlier part of this paper one can stipulate the measures that would be expected to be effective tools as a function of both the question being asked and the status of the population group(s) under examination. These measures are summarized in Table 6. The goal of

TABLE 6
Conceptual framework for choice of measures of iron status based on the responsiveness of the indicator to range of true iron status

Analytical question	Expected study-population condition	Potential indicators of iron status		
		Red cell measures*	Serum iron†	Serum ferritin
Does severe ID affect function?	Anemia very prevalent in study group with evidence of very low Hb in substantial numbers	Preferred but problematic	Desirable	?
Does marginal deficiency affect function?	Moderate prevalence of anemia in study group; few individuals with very low Hb levels	Necessary but not enough	Acceptable but problematic	Preferred
Does iron depletion affect function?	Low prevalence of anemia; indication of presence of low iron stores (eg, low serum ferritin levels)	Not responsive	?	Preferred

* Includes Hb, hematocrit MCHC, MCV, FET. FET may be particularly useful in questions concerning marginal deficiency.

† Includes both serum iron and transferrin saturation.

this assignment is to select indicators that can be expected to show a continuous response (preferably linear but this is unlikely) across the range of iron status involved. Hb is useless once one is above the range where anemia is seen. Serum iron is not very responsive once one leaves the marginal-deficiency stage. Serum ferritin may not be very effective in very severe depletion (or in very excessive accumulation). This table alludes to, but does not factor in, differences between the indicators in their variation with factors other than iron status. This is discussed below.

In reviewing the available measures, emphasis was placed on specificity and variability (diurnal and day-to-day) as well as sensitivity to changes in iron status. These two characteristics are important in the design of any study.

Consider first the question of specificity. All of the measures considered are affected by infection and/or inflammatory disease. In children we can focus attention on infection and its implications. The effect of infection will be to underestimate stores when Hb, hematocrit, FEP, serum iron, or transferrin are chosen; MCHC and transferrin saturation will be affected with the same direction of bias but the effect may be of different magnitude. The converse is true for serum ferritin: the bias is likely to be in the opposite direction, so iron stores will be overestimated.

If the design is cross-sectional then at least two types of statistical effects might be seen. If the prevalence and severity of morbidity differ between groups there could be a true bias in the comparative assessment of iron status of the groups. Perhaps much more important, because it can be postulated that measured psychologic and cognitive functions would associate with current or past morbidity, there could be serious bias in the examination

of the relationship between iron status and function. If one were to rule out the possibility of association between morbidity and outcome, then the presence of morbidity and its effect on the measures of iron status would serve to attenuate relationships (would create noise in the data). Similar considerations would hold if one suspected lead toxicity to be present in the study population; this would shift the levels of FEP and, if severe, other variables.

In a longitudinal design, error in the estimation of change in iron status is more serious than error in estimation of the absolute level of nutriture at the beginning or end of the study period. One would wish to have available measures that are specific and sensitive and that respond in a linear manner to change in iron status across the range of interest. None of the available measures fully meet these specifications.

As discussed above, morbidity affects all of the available measures of iron status. However, infectious disease (with some exceptions) does not affect response to iron. In infection depressed indices of iron status may not respond to iron supplementation but that is generally because true ID was not present, not because the infection blocked response. In addition to the implications for the cross-sectional design discussed above, there are three scenarios of the effect of morbidity on interpretation of the longitudinal design. If morbidity is a prevalent and essentially constant (in the individual) background in the study population, its major effect will be to bias the assessment of initial or final nutriture without bias in the estimation of change of status. As the prevalence of morbidity decreases, it becomes more likely to affect a particular individual at the time of both (all) measurements. The effect will be to generate noise in the estimation of change in iron status and this will decrease statistical

power. It can also bias the estimate of coefficients in regression analyses. Usually this bias will be one of underestimation of relationships but, in some circumstances, it can also lead to overestimation in multivariate analyses (34). Finally, as is generally the case in developing countries, if morbidity follows a cyclical pattern and if measurements fall into this cycle, there is a major chance for a true bias in the estimation of change in iron status. The direction and magnitude of the bias would differ with the nature and phasing of the morbidity cycle in relation to the measurement cycle and indeed with the particular measure of iron status being used.

Analogous effects would be created in longitudinal designs if the time of blood collection were not controlled. Diurnal variation could create noise or bias, depending on the design. Problems associated with diurnal variation are particularly major for serum iron and transferrin saturation but they exist also for other measures (see Table 1 for examples).

From the standpoint of study design, diurnal variation is more easily addressed than is morbidity. The obvious solution is to carefully standardize both the method and time of blood collection. In the case of morbidity, two possible approaches immediately come to mind. In a cross-sectional study one might exclude from the study, or from the analysis, children with evidence of current or recent morbidity. In a developing country this would result in a very major sampling bias and might well defeat the purpose of the original study. A much more demanding approach would be to collect data concerning current and past morbidity and include that information in multivariate or covariance analyses of the association between iron status and function.

In a longitudinal study there is another approach to minimizing the impact of random morbidity (or other factors) on the estimation of change in iron status. This involves repeated measures during the course of the study, coupled with regression analyses, for each individual, to estimate change across the period of the study. Here the question of linearity or nonlinearity of response curves becomes important in selecting the appropriate regression model. Of concern is the linearity of the relationship between the measure of iron status and iron status itself and the expected linearity of the response of iron status to the intervention over time. From discussions earlier in this paper, it is clear that one must select measures of iron status that are responsive across the range of status under study. Thus, for example, serious difficulties arise if Hb is the chosen measure and the expected iron status will range from manifest anemia to an iron-replete state. In such a situation serum ferritin would appear to present the preferred choice. Not only does it respond across that range but the impression is that it responds in a relatively linear manner.

With cross-sectional studies (measures at one point in time) the effects of short-term random variation (day-to-day variation) can be greatly reduced by collecting several samples of blood on different days clustered around the nominal date of measurement. The more measures

pooled, the smaller will be the remaining component of measurement error and the greater will be the statistical power (34). The phenomenon is illustrated in Table 3. The presence of a substantial day-to-day variation in serum iron and transferrin saturation would greatly attenuate the relationships seen in simple correlation or regression analyses—we would lose statistical power and we would incorrectly describe the nature of the relationship (slope of the regression line). Table 3 also illustrates the improvement that would result were three measures (collected on different days) available. In practice this is usually not a feasible approach for minimizing the effects of morbidity. The spacing of measures would have to be greater than the expected duration-of-illness effects.

In longitudinal designs extending over a period of months or years, the reported association between measures of iron status and age and the differences in these relationships between sexes become very important (see Table 2 for examples of effects). Here the presence of matched control groups, or at least a precise knowledge or expected change with age, is critically important to analysis and interpretation.

A very specific issue arises if the analytical design involves stratification of the individuals by the dependent variable. An example would be attempting to ask whether iron-depleted subjects show a different iron-status response to supplementation than do iron-replete subjects. If the measure of iron status has a large within-person variance compared with the expected true response, then the phenomenon of regression to the mean can cause very serious errors. This says simply that initial outliers can be expected to move toward the group mean when a second measurement is examined—the deficient can be expected to respond whether or not there was any real change. Such analytical approaches can be taken only with extreme care and full awareness of the possible spurious results. Pooling of data collected on different days, as discussed above, would decrease the error component and decrease, but not eliminate, the regression to the mean effect.

The preceding discussion has a theoretical tone. It is difficult, or impossible, to apply this theory retroactively to published studies for the simple reason that it has not been customary to report, or use in analyses, the estimated error terms. A strong plea is herewith issued to include a series of replicated measures in all future studies and to report the within-person variance of measurements. With additional knowledge about the expected behavior of the measures, both analyses and interpretations can be greatly enhanced (34).

With either cross-sectional or longitudinal designs, there is a conceptual issue that must be recognized—the fact that interindividual variation in the relationship between true iron status and any functional outcome must be expected (33). The present paper has focused on variations affecting the relationship between measures of iron status and the underlying reality of iron status. There is another variation for which we have no estimators. The level of iron nutrition required to achieve a defined level

of functional performance must be expected to vary among seemingly similar individuals. The issues are parallel to those we face in defining nutrient requirements. In the analysis of the results of human studies this fact of life will attenuate observed relationships. It becomes essential that the change (or difference) in true iron status and hence associated change (or difference) in psychological measures be substantial and that sample sizes must be adequate if we are to display the associations. Sample-size and statistical-power calculations must allow for the expected variation in response as well as the measurement errors discussed above.

Having solved all of the above problems by implementation of a near-perfect design, it is worth remembering that the probabilities are that we will have underestimated the magnitude of the relationship, the slope of the line relating iron status to function. That is, for the most part, the factors discussed are likely to attenuate that slope even though we may have designed the study to have sufficient power to detect the presence of an association.

Other cofounders: covariance

In any cross-sectional type of epidemiologic study, the very serious concern must always be that associations seen may be spurious in the sense that both iron status and brain function relate to other, perhaps unmeasured, variables but are not causally linked themselves. The epidemiologist attempts to address this problem in three ways. The first is by collection of data on all variables recognized as potential covariates and controlling for these variables through various types of multivariate analyses. The second is to search for consistency of relationships among studies carried out in various settings, perhaps with various measures. (In this case care must be taken to ensure that the individual studies were well designed and conducted and that the measures of iron status and brain function were appropriate to the condition of the population.) The third, of course, is the controlled, double-blind, experimental intervention trial, a design that is possible with pharmaceutical supplements (iron and carefully matched placebo).

The potential for covariance is very real in the research questions at hand. Measures of mental function and of iron status tend to change with characteristics of the social and biological environment. Pollitt (35) has discussed and exemplified several of these possible associations. He has discussed also the conceptual difficulties that arise when attempting to interpret analyses that have controlled for third variables when indeed all three are truly related. Here, the attempt may remove evidence of a true relationship, increasing the chance of a false-negative conclusion at the same time one is attempting to minimize the chance of a false-positive one.

Pollitt (35) makes another important point. The relationship between ID and mental performance may differ depending on the environmental circumstance of the population under study. The epidemiologist's search for

consistency across studies may not be appropriate if the study settings differ in ways that are important to the relationship.

Perhaps, however, the most revealing commentary is that of Lozoff and Brittenham (36): "Despite major methodologic strengths, virtually every study [reviewed] has been limited either by inadequate characterization of subjects' iron status, the lack of appropriate normal controls, the omission of a placebo-treated group, or the failure to demonstrate changes in behavior." There can be no question but that well-designed, well-controlled trials accompanied by careful analyses are desperately needed if we are to understand the relationship, if any, between iron status and brain function (35, 36).

The critical question: what is the hypothesized relationship?

The most important question that must be addressed before the design of a new investigation or before the planning of analysis on an existing data set is, What is the question that is being asked? In this paper sources of potential analytical and interpretational problems have been reviewed. There have been repeated hints or specific indications that the relative importance of the issue being discussed depends on the precise analytical question being addressed.

One dimension of this question, addressed above, is the range of iron nutriture over which the association is expected to exist. Only with the answer to this in mind can one begin to consider the criteria for selection of study groups or the parameters of iron status that should be included.

A closely related question is the age group, or age range, in which an effect is expected to be seen. Given the argument that a controlled intervention is the preferred design for several reasons, the question must be expanded to, In what age range is it expected that an effect of ID can be detected and a reversal of the effect by iron supplementation be demonstrated?

This then brings us to the final question: Is it current iron status that is expected to influence current mental function or was it iron status at either a critical point in brain development (physical development or functional development) or a chronic, long-standing iron deficiency that is expected to have affected the performance now being measured? The importance of this specification for the design of studies [must they be prospective from an early age or can they be over a short term (a few weeks) or longer term (a few months)?] and for the selection of data or published reports for critical analyses should be apparent (36, 37).

Finally, it is necessary to remind the reader of a fact that has been implicit in the preparation of this paper. It has been repeatedly demonstrated that controlled iron supplementation is an efficacious approach to the treatment of IDA (38, 39). It is yet to be clearly established that distribution of iron supplementation is an effective

of controlling anemia in the community. It must not be assumed that because iron supplements and placebos were distributed in the study population there was a resultant change in iron status. It is absolutely essential that the effect of the intervention on iron status be assessed and that this be made an explicit part of the analysis of the effects of intervention on the mental-performance outcome. The linkage must be built if causation is to be inferred—or if definitive refutations are to be obtained. ■

For the covariance discussion the author had access to a background literature review prepared by MW Fairchild, Division of Nutritional Sciences, Cornell University. This contribution is acknowledged with gratitude.

References

- Bothwell TH, Charlton RW, Cook JD, Finch CA. Iron metabolism in man. Oxford: Blackwell Scientific Publications, 1979.
- El Guindi M, Skikne BS, Covell AM, Cook JD. An immunoassay for human transferrin. *Am J Clin Nutr* 1988;47:37-41.
- Cook JD, Skikne BS, Lynch SR, Reusser ME. Estimates of iron sufficiency in the US population. *Blood* 1986;68:726-31.
- Dallman P. Biochemical and hematological indices of iron deficiency. In: Pollitt E, Leibel R, eds. Iron deficiency: brain biochemistry and behavior. New York: Raven Press, 1982:63-77.
- Health and Welfare Canada. Nutrition Canada: national survey. Ottawa: Information Canada, 1973.
- Garby L, Irmell L, Werner I. Iron deficiency in women of fertile age in a Swedish community. III Estimation of prevalence based on response to iron supplementation. *Acta Med Scand* 1969;185:115-7.
- Margolis HS, Hardison HH, Bender TR, Dallman PR. Iron deficiency in children: the relationship between pretreatment laboratory tests and subsequent hemoglobin response to iron therapy. *Am J Clin Nutr* 1981;34:2158-68.
- Meyers LD, Habicht J-P, Johnson CL, Brownie C. Prevalences of anemia and iron deficiency anemia in black and white women in the United States estimated by two methods. *Am J Public Health* 1983;73:1042-9.
- Beaton GH. The epidemiology of iron deficiency. In: Jacobs A, Worwood M, eds. Iron in biochemistry and medicine. London: Academic Press, 1974:477-528.
- Meyers LD, Habicht J-P, Johnson CL. Components of the difference in hemoglobin concentrations in blood between black and white women in the United States. *Am J Epidemiol* 1979;109:539-49.
- Garn SM, Smith NJ, Clark DC. Differences in hemoglobin levels between blacks and whites. *Am J Clin Nutr* 1975;28:563-8.
- Yip R, Johnson C, Dallman PR. Age-related changes in laboratory values used in the diagnosis of anemia and iron deficiency. *Am J Clin Nutr* 1984;39:427-36.
- World Health Organization. Nutritional anaemias. Geneva: WHO, 1972. (WHO tech report series #503.)
- Driggers DA, Reeves JD, Lo EYT, Dallman PR. Iron deficiency in one-year-old infants: comparisons of results of a therapeutic trial in infants with anemia or low hemoglobin values. *J Pediatr* 1981;98:753-8.
- Pilch SM, Senti FR, eds. Assessment of the iron nutritional status of the US population based on data collected in the second National Health and Nutrition Examination survey, 1976-1980. Report of the Life Sciences Research Office. Bethesda, MD: Federation of American Societies for Experimental Biology, 1984.
- Piomelli S, Brickman A, Carlos E. Rapid diagnosis of iron deficiency by measurement of free erythrocyte porphyrins and hemoglobin: the FEP/hemoglobin ratio. *Pediatrics* 1976;57:136-41.
- Dallman PR. Tissue effects of iron deficiency. In: Jacobs A, Worwood M, eds. Iron in biochemistry and medicine. London: Academic Press, 1974:437-75.
- Cavalleri A, Baruffini A, Minoia C, Bianco L. Biological response of children to low levels of inorganic lead. *Environ Res* 1981;25:415-23.
- Sempos C, Locker A, Liu K, Johnston C, Woteki C. Effects of within-person variability on prevalence estimates from biochemical indicators of nutritional status. *FASEB J* 1988;2:A631 (abstr).
- Bothwell TH, Mallett B. Diurnal variation in the turnover of iron through the plasma. *Clin Sci* 1955;14:235-9.
- Bowie EJW, Tauxe WN, Sjoberg WE, Yamaguchi MY. Daily variation in the concentration of iron in serum. *Am J Clin Pathol* 1963;40:491-4.
- Schwartz E, Baehner RL. Diurnal variation of serum iron in infants and children. *Acta Paediatr Scand* 1968;57:433-5.
- Statland BE, Winkel P. Relationship of day-to-day variation in serum iron concentrations to iron-binding capacity in healthy young women. *Am J Clin Pathol* 1977;67:84-90.
- Funk DD. Plasma iron turnover in normal subjects. *J Nuc Med* 1970;11:107-11.
- Lynch SR, Simon M, Bothwell TH, Charlton RW. Circadian variation in plasma iron concentration and reticuloendothelial iron release in the rat. *Clin Sci Mol Med* 1973;45:331-6.
- Myo Thein, Beaton GH, Milne H, Veen MJ. Oral contraceptive drugs: some observations on their effect on menstrual loss and haematologic indices. *Can Med Assoc J* 1969;101:678-9.
- Dallman PR, Siimes MA, Stekel A. Iron deficiency in infancy and childhood. *Am J Clin Nutr* 1980;33:86-118.
- Bothwell TH, Charlton RW, Motulsky AG. Idiopathic hemochromatosis. In: Stanbury JB, Wyngaarden JB, Fredrickson DS, Goldstein JL, Brown MS, eds. The metabolic basis of inherited disease. 5th ed. New York: McGraw Hill, 1983:1269-98.
- Walters GO, Miller F, Worwood M. Serum ferritin concentration and iron stores in normal subjects. *J Clin Pathol* 1973;26:770-2.
- Cook JD, Skikne BS. Serum ferritin: a possible model for the assessment of nutrient stores. *Am J Clin Nutr* 1982;35:1180-5.
- Siimes MA, Addiego JE, Dallman PR. Ferritin in serum: diagnosis of iron deficiency and iron overload in infants and children. *Blood* 1974;43:581-90.
- Cook JD, Finch CA, Smith NJ. Evaluation of the iron status of a population. *Blood* 1976;48:449-55.
- Beaton GH. Toward harmonization of dietary, biochemical and clinical assessments: the meanings of nutritional status and requirements. *Nutr Rev* 1986;44:349-58.
- Anderson SA, ed. Guidelines for use of dietary data. Report of the Life Sciences Research Office. Bethesda, MD: Federation of American Societies for Experimental Biology, 1986.
- Pollitt E. Effects of iron deficiency on mental development: methodological considerations and substantive findings. In: Johnston FE, ed. Nutritional anthropology. New York: Alan R Liss, 1987:225-54.
- Lozoff B, Brittenham GM. Behavioral aspects of iron deficiency. *Prog Hematol* 1986;14:23-53.
- Lozoff B, Brittenham GM, Wolf AW, et al. Iron therapy effects on infant development test performance. *Pediatrics* 1987;79:981-95.
- World Health Organization. Control of nutritional anaemia with special reference to iron deficiency. Geneva: WHO, 1975. (WHO technical report series #580.)
- International Nutritional Anemia Consultative Group. Guidelines for the eradication of iron deficiency anemia. New York: The Nutrition Foundation, 1977.

Comments

Peter Dallman¹

Dr Beaton's paper is very stimulating and thought-provoking. The need to have a clear conceptual model for ID may on first impression seem rather abstract but it turns out to be of practical value in placing the numerous tests for iron nutrition into an understandable framework. A conceptual model is also useful when we encounter apparent inconsistencies in clinical and epidemiological studies of iron nutrition.

One point on which Dr Beaton and I may differ is on the use of the terms *reduced iron stores* (or *low iron stores*) and *iron deficiency*. I find it difficult to consider reduced iron stores and ID as synonymous, especially in relation to the physiological handicaps associated with ID. It seems much more in line with what we know of iron nutrition to consider low iron stores as a state of vulnerability to ID. Only when low iron stores become associated with impaired production of the so-called essential iron compounds (compounds that fulfill essential physiological functions) is there a theoretic basis for abnormalities in physiological function that seem to be implied by the term iron deficiency. It is generally agreed that low iron stores by themselves represent no such handicap (1). In fact, low stores serve a homeostatic function in that they enhance the effectiveness with which the intestinal tract absorbs food iron. Thus young women, who typically have low iron stores compared with men, absorb about twice the percentage of iron from food as men.

I can describe the results of an animal experiment to further illustrate and support this conceptual model for distinguishing low iron stores per se from iron deficiency (accompanied by low iron stores). In this experiment 11 matched groups of weanling rats were fed diets for 3 wk in which the only variable was the concentration of iron (2). The 11 concentrations of iron used spanned an almost 100-fold range, from 7 to 500 mg iron/kg diet, and we measured liver iron stores, transferrin saturation, and three of the major essential iron compounds: Hb, myoglobin, and skeletal-muscle cytochrome c. Liver iron stores rose in a continuous curvilinear fashion as the concentration of iron in the diet was increased. By itself, this continuum of liver iron stores gives no indication where one might place a dividing line between physiological iron status and ID. Any value that one might select could be considered arbitrary. In human populations, estimates of prevalence based on such definitions would vary widely and could only be based on an arbitrary value for serum ferritin.

In marked contrast to the smooth curve for liver iron are the sharp breaks that occur in all the other curves at an iron intake of ~25 mg/kg. Above this intake, the level of dietary iron has no discernable effect on the concentra-

tion of any of the three essential iron compounds. Below this level, there is a sharp drop in concentration. Because the term ID connotes an abnormality, it seems most reasonable and least ambiguous to reserve it for those groups of rats whose iron intake was < 25 mg/kg diet and whose production of essential iron compounds was rate-limited by the availability of iron.

In man it is most practical to define the corresponding point in terms of Hb production. Hb is the easiest to measure of the essential iron proteins and its concentration accounts for two-thirds of body iron. Only when Hb production is rate-limited by iron availability is there the likelihood of an abnormality in the production of other essential iron compounds (3). If iron availability is decreased to the extent that the Hb concentration is below the 95% reference range that condition can be termed *iron-deficiency anemia*. If, on the other hand, Hb production is only slightly impaired, it is likely to remain within the 95% reference range, with values that are typically in the low-normal portion of that range. This condition is often termed *iron deficiency without anemia* or *preanemic ID*. It is this stage that is most difficult to identify in individuals, even with all the laboratory tests that have become available for evaluating iron status.

Low iron stores per se are a normal characteristic of infants, children, adolescents, and women during their childbearing years (1). In populations that are relatively well-nourished with respect to iron, most individuals in these groups remain in an equilibrium of low iron stores that rarely progresses to ID anemia. For example, in HANES II, the median serum ferritin was 17 $\mu\text{g/L}$ and the 25th-percentile value was 12 $\mu\text{g/L}$ in children between the ages of 5 and 10 y (4). Despite these values, which indicate very low iron stores, the prevalence of anemia was < 3% (5). The probable reason that the vast majority of children could maintain such a precarious equilibrium so successfully is that iron absorption plays an extraordinarily effective homeostatic function; absorption increases substantially when iron stores are decreased (1). The regulation of iron metabolism, therefore, works very successfully against the progression from low stores to ID if the diet supplies sufficient absorbable iron to allow some scope for this regulatory mechanism.

As Dr Beaton points out, laboratory tests are traditionally (and most simply) used in a dichotomized fashion, with values being considered either normal or abnormal. One reason why individuals near the boundary between low iron stores and ID are so difficult to classify by laboratory tests is that results for erythrocyte protoporphyrin, the ratio of serum iron to iron-binding capacity, and ferritin are also likely to yield borderline values. Whether these are classified as normal or abnormal is often determined more by sampling error, analytic error, and factors such as diurnal variation than they are by small shifts

¹ From the Department of Pediatrics, School of Medicine, University of California, San Francisco.

iron status. This helps to explain why one study showed that the prevalence of anemia in individuals with only one of the above laboratory abnormalities was 10.9%, only slightly higher than the 8.3% in the entire population (6). However anemia was found in 28% of individuals with two abnormal results and in 63% of those with abnormal results in all three tests. Results such as these were the basis for considering only those individuals with two or more abnormal values to have *impaired iron status* (4). It would be helpful if a method for using the full scale of the laboratory measurements could be developed. This might allow us to give appropriate weight to a low-normal or extremely abnormal value. However, the difficulty of developing such methods is enormous because each test behaves differently, usually in a nonlinear fashion, with abnormalities developing at individual rates and with different confounding variables.

I agree that the confounding effects of infection and inflammation on laboratory tests for iron nutrition represent a major problem. However, we can go a long way toward resolving the problem by obtaining a thorough medical history and by including acute-phase reactants in the laboratory tests in surveys and studies involving iron nutrition. Infants with mild infections either at the time of laboratory studies or even within the previous month are often anemic and have false-positive or false-negative results for the other laboratory tests (7). Anemia

is also far more common in children in whom an elevated sedimentation rate or C-reactive protein suggests a persistent or recent infection. Laboratory tests for iron status will yield more reliable results if such subjects are considered as a separate category.

The issues that Dr Beaton raises will remain with us for a long time. I do believe, however, that we are successfully chipping away at the problems of developing a conceptual model for iron nutrition and of classifying ID more reliably with laboratory tests. □

References

1. Bothwell TH, Charlton RW, Cook JD, Finch CA. Iron metabolism in man. Oxford: Blackwell, 1979.
2. Süimes MA, Refino C, Dallman PR. Manifestations of iron deficiency at various levels of dietary intake. *Am J Clin Nutr* 1980;33:570-4.
3. Dallman PR. Biochemical basis for manifestations of iron deficiency. *Am Rev Nutr* 1986;6:13-40.
4. Expert Scientific Working Group. Assessment of the iron nutrition status of the US population on data collected in the Second National Health and Nutrition Examination Survey, 1976-1980. Life Sciences Research Office. Bethesda, MD: FASEB, 1984.
5. Dallman PR, Yip R, Johnson C. Prevalence and causes of anemia in the United States, 1976-1980. *Am J Clin Nutr* 1984;39:437-45.
6. Cook JF, Finch CA, Smith NJ. Evaluation of the iron status of a population. *Blood* 1976;48:449-55.
7. Reeves JD, Yip R, Kiley V, Dallman PR. Iron deficiency in infants, the influence of antecedent infection. *J Pediatr* 1984;105:874-9.

Comments

Jean-Pierre Habicht¹

I have little to offer beyond what Dr Beaton has presented but this commentary does give me the opportunity to expand on an issue that seems particularly relevant: the nonlinearity of response of indicators of nutritional status to changes in iron nutrition. First, however, it may be useful to agree on some terminology.

The concepts outlined in Dr Beaton's paper are so rarely addressed by nutritionists that we do not yet have an agreed-upon terminology. It is satisfying to see that he and I came to a similar common-sense terminology about sensitivity and specificity (1). Unfortunately, these terms have a quantitative, well-agreed-upon meaning in epidemiology (2), where they refer to the proportion of diagnoses made correctly by a diagnostic test of ill and healthy subjects, respectively, and there is no quantitative way to relate this usage to our common-sense definitions. Because these terms are particularly relevant to clinical and field epidemiological research, I try to maintain compatibility with the accepted epidemiological terminology (2).

In epidemiology, the term *reliability* is used in the same way Dr Beaton uses it. We both (3, 4) further

differentiate between the component of unreliability that is due to measurement imprecision and that component due to other factors than imprecision and changes in nutritional status within an individual. This latter component I call *undependability*. Undependability of an indicator of nutritional status across individuals also reflects the degree to which the indicator of nutritional status is affected by nonnutritional factors. Thus *dependability* is what Dr Beaton calls *specificity*.

To avoid conflict with epidemiological usage I call *responsiveness* of the variable to changes in nutrition (5) what Dr Beaton calls *sensitivity*. Changes in nutrition status cannot be measured directly (6) but they can be inferred from changes in indicators of nutritional status in response to nutritional interventions. Thus, operationally, one estimates the responsiveness of the indicator to an intervention.

These quibbles about terminology are not important as long as we all understand what we are talking about when we use different terms for the same concept. However, that becomes more difficult as we try to operationalize these concepts quantitatively. Thus it is best if the same term is used to describe the results from the same quantitative formula, which is available for sensitivity and specificity (2), for responsiveness (5), and for reliabil-

¹ From the Division of Nutritional Sciences, Cornell University, Ithaca, NY.