Erythrocyte protoporphyrin or hemoglobin: which is a better screening test for iron deficiency in children and women?¹,²

Zuguo Mei, Ibrahim Parvanta, Mary E Cogswell, Elaine W Gunter, and Laurence M Grummer-Strawn

ABSTRACT

Background: Hemoglobin and erythrocyte protoporphyrin (EP) tests are commonly used to screen for iron deficiency. However, little research has been done to systematically evaluate the sensitivity and specificity of these 2 tests.

Objective: The objective of this study was to evaluate the sensitivity and specificity of hemoglobin and EP measurements in predicting iron deficiency in preschool children and in women of childbearing age.

Design: We examined data from the third National Health and Nutrition Examination Survey (n 2613 children aged 1–5 y and n = 5175 nonpregnant women aged 15–49 y). Children or women with blood lead ≥ 10 μg/dL were excluded from this study. We used the receiver operating characteristic (ROC) curve to characterize the sensitivity and specificity of hemoglobin and EP measurements in screening for iron deficiency, defined as having abnormal values for ≥ 2 of the following 3 indexes: mean cell volume, transferrin saturation, and serum ferritin.

Results: The ROC performance of EP was consistently better than that of hemoglobin for detecting iron deficiency in preschool children. However, in nonpregnant women, we found no significant difference between EP and hemoglobin in ROC performance for detecting iron deficiency. We observed the same results when we stratified the analyses by sex and race of the children and by race of the women.

Conclusions: For children aged 1–5 y, EP is a better screening tool for iron deficiency than is hemoglobin. However, for nonpregnant women, EP and hemoglobin have similar sensitivity and specificity for predicting iron deficiency. Am J Clin Nutr 2003;77:1229–33.

KEY WORDS Iron deficiency, hemoglobin, erythrocyte protoporphyrin, zinc protoporphyrin, mean cell volume, transferrin saturation, serum ferritin, receiver operating characteristic curve, preschool children, women, iron deficiency anemia

INTRODUCTION

Iron deficiency is the most common micronutrient deficiency in both developing and developed countries, particularly among children and women of childbearing age (1–3). Hemoglobin, hematocrit, mean cell volume (MCV), and red blood cell distribution width are hematologic tests that are commonly used to assess iron status. Commonly used biochemical tests include serum iron, total iron-binding capacity, transferrin saturation (TS), serum ferritin (SF), transferrin receptor, and erythrocyte protoporphyrin (EP).

A reticulocyte count (or reticulocyte index), which provides information about the rate at which the bone marrow is producing red blood cells, is usually performed when patients are evaluated for anemia or for response to its treatment. Of all of these tests, hemoglobin is the most commonly used to screen for anemia as a proxy for iron deficiency because of its low cost, the ease and speed of the procedure, and its better performance compared with hematocrit (4–6). However, as an indicator related to red blood cell population turnover, hemoglobin only detects the late stages of iron deficiency; mild iron deficiency may not affect the hemoglobin concentration (4–6). Binkin et al (4) reported that using hemoglobin to screen for iron deficiency in the United States detected only 25% of iron-deficient children aged 1–5 y and only 37% of iron-deficient women of childbearing age.

Biochemical tests are more specific than are hematologic tests for assessing iron nutrition status, but they are also more expensive and more complicated to use. Of all the biochemical tests, EP may have the most potential for low cost and simplicity because zinc protoporphyrin (ZP), a related metabolite, can be measured with a portable hematofluorometer (7, 8). Previous studies have suggested that EP and ZP may be more sensitive than hemoglobin for detecting iron deficiency (8–11). Margolis et al (10) reported that using EP (≥ 3.0 μg/g hemoglobin) detected 42% of iron-deficient children aged 6 mo to 17 y (as determined by a 1 g/dL rise in hemoglobin with iron therapy). In contrast, hemoglobin (≤ 2 SD) detected 35% of iron-deficient children. Serdar et al (11) reported that in a case-control study in infants, EP was a more sensitive but less specific test for iron deficiency anemia (sensitivity of 94% and specificity of 72%) compared with SF (sensitivity of 85% and specificity of 94%) among iron-deficient anemic patients. Siegel and LaGrone (12) reported that ZP (≥ 50 μmol/mol heme) had greater sensitivity than did hematocrit (≥ 33%) in identifying children aged 9–36 mo who responded to iron treatment; sensitivity values were 81% and 16%, respectively. ZP also had a higher positive predictive value (72%) compared with hematocrit (47%).

¹ From the Division of Nutrition and Physical Activity (ZM, IP, MEC, and LMG-S) and the Division of Laboratory Sciences (EWG), Centers for Disease Control and Prevention, Atlanta.

² Address reprint requests to Z Mei, Centers for Disease Control and Prevention, Mailstop K-25, 4770 Buford Highway, Atlanta, GA 30341-3724. E-mail: zmei@cdc.gov.

Received August 9, 2002.

Accepted for publication October 24, 2002.
Using data from the second National Health and Nutrition Examination Survey (1976–1980), Sudre and Yip (13) compared the sensitivity and specificity of EP and hemoglobin for detecting low TS using receiver operating characteristic (ROC) curves and found that EP and hemoglobin results were similar among children aged 1–5 y and women aged 18–44 y. The ROC curve method allows comparison of the sensitivity of a given test to that of another at the same level of specificity. We sought to refine this analysis using data from the third National Health and Nutrition Examination Survey (NHANES III). We excluded subjects with high blood lead values and defined iron deficiency as having abnormal values for ≥2 of 3 tests (MCV, TS, and SF).

### SUBJECTS AND METHODS

The Centers for Disease Control and Prevention (CDC) conducted NHANES III from 1988 to 1994 to provide representative data on the civilian, noninstitutionalized US population aged 2 mo to 74 y (14, 15). All procedures used in NHANES III were approved by the CDC National Center for Health Statistics Institutional Review Board, and written informed consent was obtained from all subjects.

NHANES III included standardized hematologic and biochemical tests for iron status, such as hemoglobin, MCV, SF, TS, and EP. Hemoglobin and MCV were measured in mobile examination centers as part of a complete blood count by using an automated electronic counter (Coulter S-Plus Jr; Coulter Electronics, Hialeah, FL). All the biochemical iron indexes (such as SF, TS, and EP) were measured in the NHANES laboratory at the CDC. SF was measured with the Bio-Rad QuantImmune Ferritin IRMA (Bio-Rad Laboratories, Hercules, CA). TS was calculated from serum iron and total iron-binding capacity, which were measured with a modification of the automated AALI-25 colorimetric method (Alpkem TFA analyzer; Alpkem, Clackamas, OR), and EP was measured via fluorescence extraction by using a modification of the Sassa method (16). Details of the assay methods for these laboratory parameters have been published elsewhere (16, 17).

NHANES III included 4745 children aged 1–5 y and 6014 women aged 15–49 y. For this analysis, we excluded 345 children and 52 women with blood lead levels ≥10 μg/dL to avoid the confounding influence of lead on hemoglobin and EP results. We also excluded 351 women who were pregnant and 47 who had missing data for pregnancy status. NHANES III subjects whose records had missing data for EP, hemoglobin, MCV, TS, or SF values (1787 children and 441 women) were also excluded. Thus, our final analyses included 2613 children aged 1–5 y and 5175 nonpregnant women aged 15–49 y. We did not take into account the original sample weights assigned to the data because we did not intend to perform population-based analyses.

We defined iron deficiency as having abnormal values for ≥2 of the following 3 tests: MCV, TS, and SF. To be considered iron deficient, an individual has to have an abnormal value for ≥2 indicators. The rationale for the use of multiple tests was the finding that populations with only 1 abnormal test of these 3 tests had scarcely more anemia than did populations with normal values for all 3 tests (3, 18). Cutoffs for TS and SF were determined on the basis of those recommended by Looker et al (3), which were derived from NHANES II data by an expert panel (18, 19). Thresholds for MCV were determined on the basis of values indicated in the Healthy People 2000 Final Review (20). The thresholds for defining the abnormal values for these 3 indexes of iron deficiency are shown in Table 1.

We used ROC curves (21) to characterize the sensitivity and specificity of EP and hemoglobin measurements in screening for iron deficiency. The ROC curve is constructed by first calculating the sensitivity and specificity generated when using a series of EP or hemoglobin cutoffs against the definition for iron deficiency. Then, the series of sensitivities are plotted on the y axis against the corresponding values of 100−specificity on the x axis. In general, the farther away the curve is from the diagonal line of chance, which extends at 45° from the origin (x = 0, y = 0), the better the indicator’s performance. MEDCALC software (22) was used to test the significance for the areas under the ROC curves (AUCs).

Because liver disease, infection, and inflammation can influence biochemical indexes (23), we repeated the above analysis on a subsample that did not include women who had signs of infection or possible liver disease. Signs of infection were defined as an abnormal white blood cell count (> 10.0 × 10^9/L) or elevated C-reactive protein concentration (> 6 mg/L). Signs of possible liver disease were defined as abnormal elevations (> 2 times the upper limit of normal) of alanine aminotransferase (> 70 U/L; EC 2.6.1.2), aspartate aminotransferase (> 70 U/L; EC 2.6.1.1), or both (24). However, we were unable to exclude preschool children with these abnormalities to control for a possible confounding effect on biochemical indexes because C-reactive protein, alanine aminotransferase, and aspartate aminotransferase were not measured in children.

### RESULTS

Basic demographic and biochemical characteristics of our sample are shown in Table 2. Among preschool children, the sensitivity and specificity of EP in detecting iron deficiency were significantly and consistently better than those of hemoglobin (P < 0.001; Figure 1 and Table 3). For example, at cutoffs yielding an 80% specificity (100−specificity of 20%), the sensitivities were 65% for EP but only 50% for hemoglobin for the detection of iron deficiency (Figure 1). However, there was no significant difference in ROC performance between EP and hemoglobin in the detection of iron deficiency in nonpregnant women (P > 0.05; Figure 2 and Table 3). For example, at cutoffs yielding an 80% specificity, the sensitivities were 70% for EP and 67% for hemoglobin for the detection of iron deficiency. In comparing Figures 1 and 2, we observed that the ROC performance of EP was equally good for children and women, but the ROC performance
TABLE 2
Basic demographic and biochemical characteristics of the selected sample of children aged 1–5 y and nonpregnant women aged 15–49 y from the third National Health and Nutrition Examination Survey.

<table>
<thead>
<tr>
<th></th>
<th>Children</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 2613)</td>
<td>(n = 5175)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>50.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Non-Hispanic white (%)</td>
<td>28.7</td>
<td>31.5</td>
</tr>
<tr>
<td>Non-Hispanic black (%)</td>
<td>29.2</td>
<td>33.5</td>
</tr>
<tr>
<td>Mexican American (%)</td>
<td>36.7</td>
<td>30.1</td>
</tr>
<tr>
<td>Other racial group (%)</td>
<td>5.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Anemic (%)</td>
<td>9.6</td>
<td>16.5</td>
</tr>
<tr>
<td>Iron deficient (%)</td>
<td>6.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Age (y)</td>
<td>3.1 ± 1.16</td>
<td>31.1 ± 9.5</td>
</tr>
<tr>
<td>EP (µmol/L)</td>
<td>0.975 ± 0.475</td>
<td>1.07 ± 0.813</td>
</tr>
<tr>
<td>MCV (FL)</td>
<td>80.46 ± 4.86</td>
<td>87.83 ± 6.00</td>
</tr>
<tr>
<td>TS (%)</td>
<td>18.82 ± 9.07</td>
<td>22.42 ± 11.59</td>
</tr>
<tr>
<td>SF (µg/L)</td>
<td>28.93 ± 21.93</td>
<td>50.46 ± 53.67</td>
</tr>
<tr>
<td>Lead (µg/dL)</td>
<td>3.97 ± 2.19</td>
<td>2.21 ± 1.60</td>
</tr>
</tbody>
</table>

1 EP, erythrocyte protoporphyrin; MCV, mean cell volume; TS, transferrin saturation; SF, serum ferritin.
2 On the basis of the Centers for Disease Control and Prevention hemoglobin cutoffs (6).
3 Defined as having abnormal values for ≥2 of the following 3 tests: MCV, TS, and SF.
4 ± SD.


of hemoglobin was worse for children than for women. We observed the same results when we stratified the analyses by sex and race for children and by race for women (Table 3).

When we examined the ROC performance for those children whose blood lead concentrations were ≥10 µg/dL (n = 345; mean blood lead = 15.39 ± 6.89 µg/dL), we found no significant difference between EP and hemoglobin in the detection of iron deficiency (AUC for EP = 0.888, AUC for hemoglobin = 0.881; P = 0.876). We did not examine the ROC performance for women whose blood lead concentration was ≥10 µg/dL.

FIGURE 1. Comparison of the receiver operating characteristic curves of erythrocyte protoporphyrin (thick line) and hemoglobin (thin line) in detecting iron deficiency in children aged 1–5 y from the third National Health and Nutrition Examination Survey (1988–1994).

FIGURE 2. Comparison of the receiver operating characteristic curves of erythrocyte protoporphyrin (thick line) and hemoglobin (thin line) in detecting iron deficiency in women aged 15–49 y from the third National Health and Nutrition Examination Survey (1988–1994).
because the sample size \((n = 52)\) was too small to make meaningful comparisons.

The repeated analysis of the women’s subsample after excluding women who had signs of infection or possible liver disease \((n = 4476)\) shows that there was still no significant difference between EP and hemoglobin in the detection of iron deficiency \((\text{AUC for EP} = 0.839, \text{AUC for hemoglobin} = 0.819, \text{P} = 0.071)\).

**DISCUSSION**

Our study found that EP is a better screening tool for iron deficiency than is hemoglobin in US children aged 1–5 y. However, in nonpregnant US women, EP and hemoglobin performed similarly in predicting iron deficiency.

The strengths of the present study include the large number of standardized measurements and the use of ROC curves, which can summarize all the sensitivities and specificities in one diagram and can identify which indicator has the highest sensitivity and specificity for the predictor variable. However, the conclusions must be viewed with some caution because our data may be difficult to generalize to other countries. In the United States, iron deficiency is an important cause of anemia among young children and women of reproductive age \((1, 25)\). In other countries, however, anemia may result from a variety of underlying conditions besides iron deficiency, such as other nutritional deficiencies \((\text{eg}, \text{vitamin B–12 or folate deficiency})\), thalassemia, malaria, and other infections that are uncommon in the United States. In countries where the prevalence of thalassemia is high, EP may be a better predictor of iron deficiency than of anemia among women of reproductive age because it is not increased by thalassemia but is increased by iron deficiency, infection, inflammation, and other conditions that interfere with heme synthesis \((\text{such as lead poisoning})\) \((26)\).

A definitive diagnosis of iron deficiency can be made by bone marrow aspiration or by measuring hemoglobin response to a therapeutic trial of iron \((27)\). Biochemical indicators of iron deficiency are also helpful, but in the presence of infection, inflammation, malignancy, malnutrition, alcoholism, or liver disease, their interpretation can be problematic. To control for these potential confounding effects, we repeated the analysis on a subsample of women after excluding those who had signs of infection or possible liver disease, and the results confirmed our initial findings that EP and hemoglobin are equally sensitive and specific for this population. Also, our definition of iron deficiency included having abnormal values for \(\geq 2\) of 3 tests \((\text{MCV, TS, and SF})\), which is generally considered to be a more accurate measure of iron status than is any single indicator alone.

Anemia \((\text{low hemoglobin concentration})\) is a late-stage indicator of iron deficiency and has serious negative consequences on health, mental development, and work capacity \((28–32)\); this shows the need for a screening tool that is effective at an earlier stage of iron deficiency without anemia. EP has potential as a screening test for this purpose because it can detect iron deficiency at early stages \((8, 9)\) and \(\text{ZP}\) can be performed relatively easily and very inexpensively in the field or clinic setting \((\text{ie, capillary blood is tested in a hematofluorometer})\). EP was recommended by the CDC as the preferred screening test for detection of lead poisoning in pediatric populations \((33, 34)\). However, the sensitivity of EP for the detection of lead poisoning is poor at the acceptable blood lead threshold of 10 \(\mu\text{g/dL}\) \((35, 36)\). At low lead concentrations, EP has been used for assessing iron status \((9, 37)\). Our study has shown that EP is at least as sensitive and specific as is hemoglobin in the detection of iron deficiency in nonpregnant women and is more sensitive and specific than is hemoglobin in young children. These findings apply to the United States and other areas in which the prevalence of elevated blood lead concentrations is not high.

Lamola and Yamame \((38)\) studied the fluorescence of protoporphyrin in whole blood and thus were the first to determine, in 1974, that the major molecular form of EP is ZP. Smith et al \((39)\) confirmed in 1980 that EP usually exists as ZP \((\sim 90\% \text{ of EP is ZP})\). However, this ratio may vary in several disease conditions because of the higher amount of metal-free protoporphyrin present in reticulocytes \((37, 40, 41)\). Thus, the techniques for measurement of EP can be divided into 2 basic categories: direct measurements and measurements after solvent extraction \((37, 42)\). Cell suspensions or lysates and whole blood \((\text{hematofluorometer})\) techniques are the direct measurements. Measurements after solvent extraction include acidified solvents \((\text{zinc is detached})\) and nonacidified solvents \((\text{zinc is retained})\). These methods are used less frequently than is hematofluorometry but have advantages with respect to direct standardization and specimen stability \((37, 42)\). Although ZP measurements obtained with a hematofluorometer reflect the fraction of blood protoporphyrin that exists as the zinc chelate, whereas EP measurements reflect the total protoporphyrin, the results for EP and ZP \((\text{in all situations except erythropoietic protoporphyrin})\) should be very similar if the hematofluorometer is properly calibrated and maintained and a standardized procedure is followed \((37, 42)\). Hematofluorometers are small, dedicated instruments that are relatively simple to operate and offer some degree of portability \((37)\). Hematofluorometers use a drop of whole, unprocessed blood obtained by capillary sampling, and the results are available in \(\approx \text{1 min}\). Also, the results obtained with hematofluorometers are independent of the blood volume used, and the cost of the hematofluorometer test is relatively low \((37)\).

Our study confirmed the sensitivity of EP as a screening test for iron deficiency and showed that a carefully executed ZP measurement \((\text{which should correlate well with EP})\) makes a convenient field assay for this purpose. Hematofluorometry is currently the most widely used technique for measuring ZP. It was reported that an increase in ZP or EP is the first measurable biochemical change in erythrocytes after a decline in iron status \((8, 43)\). However, the use of EP or ZP for detecting iron deficiency before anemia is underused. Our study confirms that the use of EP or ZP to assess individuals and populations for iron deficiency deserves further consideration in the United States and in settings where multiple micronutrient deficiencies are common.

We thank Ray Yip for reviewing preliminary drafts of this manuscript and Patrick Parsons for his valuable advice on the hematofluorometry technique.

ZM designed the study, analyzed the data, and wrote the manuscript. IP, MEC, and EWG assisted in the study design, data interpretation, and revision of the manuscript. LMG-S assisted in the study design, data interpretation, analyses supervision, and revision of the manuscript. None of the authors had any financial or personal interest in any company or organization connected in any way with the research represented in the article.

**REFERENCES**


1233

SCREENING FOR IRON DEFICIENCY IN CHILDREN AND WOMEN


33. Increased lead absorption and lead poisoning in young children. A statement by the Centers for Disease Control. Atlanta: Centers for Disease Control, 1975.


