Measurement of Iron Deficiency in Infants

Implications for Screening in a High Risk Population

by

Clare A. Hawkins MD.

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfilment of the Requirements for the Degree of

MASTER OF SCIENCE

Department of Community Health Science
Faculty of Medicine
University of Manitoba

March 31, 1997

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

MEASUREMENT OF IRON DEFICIENCY IN INFANTS

IMPLICATIONS FOR SCREENING IN A HIGH RISK POPULATION

BY

CLARE A. HAWKINS

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

Structured Abstract

Objective: To evaluate the measurement properties of hematological and biochemical indicators of iron deficiency (ID).

Design: Secondary data analysis of a longitudinal data set collected for a randomized controlled trial of iron supplemented formula conducted in 1991. Analysis of a data-set of four venous samples drawn from 225 infants at ages 6, 9, 12 and 15 months.

Setting: Aboriginal children living in the urban core of Winnipeg.

Patient Selection: Non-breastfeeding mothers who agreed with informed consent to participate in an infant formula iron-fortification trial < 2 months of age.

Measurement and Results: Serum Ferritin (SF), Transferrin Saturation (TS) and Free Erythrocyte Protoporphyrin (FEP) were compared with Hemoglobin (HB), Mean Corpuscular Volume (MCV) and Red Cell Distribution Width (RDW). Non parametric receiver operator characteristic curve (ROC) analysis was conducted as a cut-off independent method to compare the iron tests. The areas under the curve were measured to contrast the diagnostic strength of each test. The analysis was repeated with a series of surrogate "gold standards". At the same assessment age, SF and hematological tests were used . The longitudinal nature of the dataset was used to determine if a low value of a test

at one sample interval predicts low values of other tests at the next sample interval. Predicting which infants would have a HB rise in response to iron administration was also evaluated.

Conclusions: HB performed well at predicting the response to iron supplementation. TS was the most accurate early indicators of ID in infants. Age appropriate cut-off values for FEP and MCV are recommended. Recommendations are made for screening and primary prevention in high risk, remote populations.

Dedication

This thesis is dedicated to my wife Laura and to our children, Teo and Peter.

Acknowledgments

My sincere gratitude to my committee: Dr. Charlyn Black for her persistence over several years as my supervisor, Dr. Mike Moffatt who encouraged me to pursue this specific area of research and Dr. Margaret Fast for critically reviewing this thesis.

My deepest appreciation also goes to Dr. Rob James for his technical assistance with data management, editing and reflection. Most of all I have benefitted from his tireless encouragement.

For patience throughout the extended time required to complete this project and for the sacrifices that they have had to make, I am indebted to Laura, Teo and Peter.

Thanks to God, who has made the completion of this project possible.

Table of Contents

Index of Figures

Index of Tables

Introduction

Introduction to the Thesis:

Iron is a critical element in major physiological body functions. It is a component of hemoglobin, the oxygen carrying protein in blood, and it is also a component of cytochromes and enzymes which are required in multiple important body processes. Severe deficiency measured as iron deficiency anemia, (IDA), has been shown to cause delays in infant psychomotor development which are reversible with iron administration. It has also been associated with reduced immune function.[1] Therefore, iron deficiency (ID) is a significant health problem for infants under two years of age.

Certain populations of socioeconomically disadvantaged infants have a higher prevalence of ID. This is common in many developing countries but also in North America within poorer communities. In Manitoba, aboriginal communities both urban and remote suffer from the many health effects of poverty including ID. The effort to define the magnitude of the problem has been hampered by methodologic problems in the measurement of ID. Hemoglobin, (HB), which is most reliably assessed with automated equipment such as a coulter counter, has been the measure most often used when trying to count the numbers of anemic children. In remote communities, however, this technology is not available. A recent population health study of Inuit communities in Keewatin used a portable photometer for the measurement of HB and red blood cell count on capillary (finger prick) blood samples.[2] It confirmed a 5-20% prevalence of anemia (low hemoglobin), in different communities but the finding of significant macrocytic anemia

raised etiologic questions as to whether a deficiency in iron, B_{12}, folate or a genetic hemoglobin defect explained the low hemoglobin values. The macrocytic cells could also have been explained as a measurement error. A reliable method for detecting iron deficiency in isolated northern communities would contribute greatly to the delineation of this question as well as assisting in answering the main question of the prevalence of ID.

Many tests are available to measure iron levels in the human body. These tests fall into two categories, biochemical and hematological. Of the two categories of tests, biochemical tests attempt to assess iron status more directly. The more familiar hematologic tests measure the components or dimensions of red blood cells involved in iron-assisted methods of oxygen transport. Much research has been conducted in the previous 20-30 years to evaluate the relative strengths of each of these tests but no individual test has been chosen as the superior measure of iron status.[3]

This thesis addresses the question of which test or series of tests can best be employed in the population assessment of iron deficiency in young children? And which would be best applied to the problem of screening within aboriginal communities?

The study used data from a prospective randomized trial of the iron-fortification of infant formula and its effect on mental and motor development that was conducted on a population of urban aboriginal infants. It contained serial measurements of six different iron tests at 3 month intervals from ages 6 to 15 months. Standard methods of clinical epidemiology are used to address issues of test selection and interpretation.

Literature Review

Introduction to Iron Deficiency:

An understanding of iron deficiency is predicated on an understanding of the currently accepted division of iron metabolism into three physiological compartments. Each of these compartments are depleted in turn and each have separate measures of iron status:

Three Compartments of Body Iron:

1. Storage iron, accessible for incorporation into HB, is composed of insoluble hemosiderin (which is the stainable iron seen during bone marrow examination) and the soluble protein, ferritin. These proteins exists mostly within the bone marrow of adults but also the liver and spleen of infants. The balance of "storage" iron which is not accessible for HB synthesis is in "parenchymal" or "tissue iron" as muscle protein, myoglobin or as enzymes, cytochromes. 2. Transport proteins carry iron to the sites of blood cell formation. This transport iron accounts for a relatively small part of the total iron in the body. 3. The majority of body iron exists within the circulating protein, HB, (Fig. 1).

Three Stages of Iron Deficiency:

Heinrich in 1970 divided iron deficiency into three sequential stages.[4] (Fig. 2) Although HB has been the traditional measure of ID, it is likely the final stage in a progressive reduction of body iron. When a body is deprived of iron, there is first a

reduction in the body's storage iron compartment. As the iron stores become depleted one can measure a reduction in transport iron. This stage is called "latent iron deficiency". Finally, when there is insufficient iron to produce HB there is a measurable reduction in HB called iron deficiency anemia, (IDA). Anemia is the outcome which has been well correlated with reversible psychomotor problems of development in infants.[5-7] Earlier states of deficiency also seem to have some relation to adverse outcomes and hence these precursor states may be important to identify.[8-10]

Iron Tests for each Compartment and Stage:

Biochemical tests measure compounds within the blood which fluctuate in proportion to the iron status of the individual. Hematological ones measure features of the red blood cell and its oxygen carrying protein, HB. Each individual test is most relevant to a specific compartment and stage of deficiency.[11-14] (Figure. 3)

Of the biochemical tests, Serum Ferritin, (SF) is the best measure of iron stores. It has been shown to be an accurate reflection of bone marrow iron.[15] The next test, Transferrin Saturation, (TS) reflects transport iron, which is measured as serum iron and it's carrier protein transferrin. TS is a calculated index of the percentage of the protein which is bound to iron, (and by extension, the amount of unbound protein deficient in iron). When a significant reduction in the delivery of iron to the blood forming areas of the body occurs, other parameters become affected. Protoporphyrin is a chemical which when bound to iron forms heme, the oxygen carrying compound of the hemoglobin protein. If iron is unavailable to produce heme there is a resulting increase in the amount

of unused protoporphyrin circulating in the blood as Free Erythrocyte Protoporphyrin (FEP). These, then, are the three biochemical tests and the order in which they are thought to decrease with the development of ID.

Hematological tests include HB, Mean Corpuscular Volume (MCV) and Red Cell Distribution Width (RDW). When hemoglobin production begins to be compromised, but before the actual HB value declines, the size of the newly produced red blood cells becomes smaller and the consequent variability in cell size is measured as Red Blood Cell Distribution Width, (RDW).[11] Another index of red blood cell size pertains to the actual volume of the individual cells,. mean corpuscular volume, (MCV).[3] The compromise in production ultimately results in a drop in the absolute amount of hemoglobin (HB) but this stage of "frank anemia" is only the final outcome.

In reality these stages and tests are not as discreet as outlined, many of the tests drop concurrently as deficiency develops. Some authors feel that TS declines later than RDW and FEP. A more detailed diagram including four body pools is adapted from Bothwell 1966 in Figure **4**.[16,17]. It would appear from these categorizations that SF would be the earliest and perhaps best measure of iron deficiency.

Measurement Issues:

In the struggle to select a superior measure of iron deficiency it is necessary to define three terms which explain the measurement properties of interest: accuracy,

validity and reliability. Definitions are taken from A Dictionary of Epidemiology by John Last.[18]

Accuracy:

Accuracy is defined as: *"The degree to which a measurement, or an estimate based on measurements, represents the true value of the attribute that is being measured".* In the context of this thesis project it must be noted that there is an underlying state of the body's iron status which none of our tests can fully and completely measure. (True value of the attribute.) Each of the iron tests attempts to identify the individual who is iron deficient and describe the magnitude of the deficiency but none of them are completely and consistently able to identify all individuals or reflect the magnitude of the deficiency. When attempting to find the superior iron test, its ability to reflect the underlying iron status is what is being measured.

Validity:

Closely related is the concept of Validity. John Last quotes the Oxford English Dictionary when defining this as *"sound and sufficient".*[19] He goes on to explain " if *a test measures what it purports to measure (it is sufficient) then the test is said to be valid".* There are two types of validity which are relevant to this thesis, construct and criterion validity. These will be itemized below.

Construct Validity:

Construct validity is "*the extent to which the measurement corresponds to theoretical concepts (constructs) of the phenomenon under study*". Each of the iron tests has been identified with a compartment or stage of deficiency. The extent to which they actually measure this stage is construct validity. For example, since we postulate that iron is a critical component of the hemoglobin molecule, protoporphyrin which is the other component of the heme compound, has a valid pathophysiological explanation for varying in proportion to iron deficiency.

Criterion Validity:

A more useful evaluation concept is criterion validity; "*The extent to which the measurement correlates with an external criterion of the phenomenon under study*". The comparison of a test with a well accepted reference test is the most desirable. Accurate and valid tests must be compared with a "gold standard".

There have been several attempts to find research measures of iron status which are valid. The most valid of these measures are often laborious or even injurious to patients. As with many other cases of test selection the trade-off between accuracy and practicality becomes real. These accurate tests are considered to be "gold standards" and can function as an external reference for the evaluation of tests which may not be as accurate.

Garby, when studying adults speculated that "a criterion based upon the individual's feelings of well-being and capacity to perform pertinent social functions would theoretically be the most relevant."[20] A parallel concept for infants might be an

assessment of neurodevelopmental function. Although IDA has been well correlated with poor mental and motor performance this would not be easy to measure nor would it be precise.[21]

There is no true proxy for the direct measurement of underlying iron status and the "gold standards" used in some research studies such as bone marrow iron, liver biopsy, ^{57}cobalt excretion, or even absorption of radioactive iron salts, are not feasible for application on a large scale.[17,22,23] The most direct measure involves weekly phlebotomy to determine the amount of hemoglobin iron removed and the induced iron deficit.[14,23] Because of the practical and ethical difficulties involved in performing these on a population of infants it is difficult to compare standard measures to these references.

Response to iron therapy has been the only population-based gold standard used in many large scale studies. The minimum significant magnitude of this HB rise in several studies is accepted as 5-10 g/L.[24-27] A study evaluating the variability of hemoglobin reported that an increase of 6 g/L would exclude 95% of the differences caused by analytic factors and biologic variation.[28] Since taking blood is not as invasive or painful as bone-marrow aspiration or repeated phlebotomy it is well accepted by the population. Although compliance with iron therapy in most studies is only about 50 to 75% the effect of iron administration is still quite measurable. In most of the studies the iron tests, (HB, SF, TS ● FEP) are contrasted in their ability to predict the effect of supplementation. This is called predictive validity. As this is the most commonly used standard, 5 studies will be reviewed.

In 1969 Garby published the results of a supplementation trial involving 121

healthy women aged 18 to 48.[20] He documented an approximately 30% rise in HB and TS compared to placebo, but found that a low HB was a better predictor than TS of the effect of supplementation. Dallman et al, reported on 188 one-year-old children of US Air Force personnel who had low HB values then completed iron therapy of 3 mg/kg as ferrous sulfate for three months.[24] They found that pretreatment HB was the best predictor of response although TS was the next most useful single test.

Margolis et al, studied 344 Inuit children between 1977-78.[27] After 3 months of supplementation to deficient individuals, A low value of HB (<1.5 SD from normal) was the most efficient test at predicting a response to iron. They described this predictive property as sensitivity. The other tests, SF, TS and FEP all had between 40-60% sensitivity and specificity. Idjradinata and Pollit published a trial of iron supplentation of infants aged 12 to 18 months from Indonesia in 1993. They documented the reversibility of delays in mental and motor development after iron administration. Although those with documented IDA had test results which increased dramatically, many of the children in each of their three groupings, IDA, ID and iron sufficient responded by increasing several measures of iron status.

After reviewing each of these trials it is apparent that HB at time=1 is more able to predict a response in HB at time=2 than the other tests used alone at time=1. One must remember, however, that this gold standard is not golden. There is more than a bit of room for scepticism when we see the possibly circular logic in finding HB as the best test for predicting a HB response. Serum Ferritin and TS, however, increased substantially in some individuals whose HB did not rise. This indicates that HB as an initial test fails to

identify all children who will respond to iron. Pretreatment HB may not, then, be a sufficient test.

This discussion has parallels within the bacteriology literature when comparing new immunological tests against the "gold standard" of culture for any fastidious organism. Although a positive culture is very specific it is not always sensitive. Using a positive culture as a reference standard underestimates a new test which may be correctly identifying culture-negative individuals as having true infections.[29]

Although not a "gold-standard", the use of other common tests applied simultaneously to the same subjects can serve as a reference point with "concurrent validity". For many population studies, SF, TS or a combination of these and the other tests are used to measure deficiency.[30-32] Whether these multiple tests are used in sequence or parallel will determine what fraction of the population is identified as deficient. (Figure 5)

Reliability:

Reliability is defined as "*a quality that is sound and dependable*". Epidemiologically it is a result or measurement which is stable, i.e., when repetition of an experiment or measurement gives the same results". When variation exists as a property of the phenomenon under study rather than an artefact of the measurement it is described as "day-to-day" variation or sometimes "within-day" variation, (diurnal variation).

Repeatability and reproducibility are synonyms for reliability in many usages but are "*actually referring to the measuring procedure rather than the attribute being*

measured". Some of the literature reports situations in which each of the iron tests have been repeatedly performed on the same blood sample or subject. When very similar results ensue from serial measurements, the test is said to be reliable. Variability of measurements are often calculated with the statistic, coefficient of variation.

Precision: "*is the quality of being sharply defined through exact detail. A faulty measurement may be expressed precisely, but may not be accurate*".

In summary, in order to measure the underlying iron status of an individual or a population we are obliged to utilize proxy measures which may have content validity and some degree of concurrent and predictive, (criterion) validity.

These tests, however, have specific advantages and disadvantages which have been documented in previous studies and will be outlined below. It will become apparent to the reader that there is no agreement as to the "best" iron test or the appropriate cut-off values. Large meta-analyses comparing iron tests have not been conducted chiefly because each of the studies use different definitions of deficiency, different cut-off levels and have been conducted on different populations.

Iron Tests:

Hematologic Variables HB, MCV, RDW:

Testing Methods:

There is currently little variation in the measurement of each of these tests between different labs and between different urban studies because they employ

standardized, automated equipment such as an electronic particle counter, (Coulter Counter).[28,33] This is considered to be the best technique for hematological analysis with determinations being both accurate and reproducible.[28,34] Details of the methodology and variation are presented in Appendix I.

Portable hematologic equipment has great appeal for remote settings and has been used extensively. This equipment however, has more problems with measurement variation. (Appendix II).

Hemoglobin (HB):

This is the test which is used in most studies since it is simple, the laboratory error is small and there is much less biological variation on repeated sampling than for tests such as serum iron.[35] The range of HB diurnal variation is 3-5% and there is an insignificant amount of day-to-day variation in HB values, and especially as compared to the variation in biochemical tests.[3,28,36] There is more variation possible when capillary rather than venous samples are obtained and a careful set of procedures are necessary to minimize the observed error seen in some studies which have resulted in falsely low values.[37,38] This is a significant limitation when working with a pediatric population since phlebotomy is often difficult for both technical reasons and withheld consent.

The major problem with this measure is that a drop in HB is a relatively late finding in the spectrum of ID. One observes a Gaussian distribution of hemoglobin values in a healthy population; however in populations with a high prevalence of anemia, this may not be the case.[39] One of the difficulties with using HB as a measurement of ID is

that normal and abnormal values overlap significantly and so it is difficult to clearly separate those children with iron deficiency.[20,40]

The prevalence of certain medical conditions in the population under study may affect HB values even in the absence of ID. Infections or inflammatory disease may lower hemoglobin values in the absence of ID.[41,42] Several hemoglobinopathies such as α or β thalassemia or sickle cell disease or trait may lower HB values.[43]

Hemoglobin values also change significantly with age. At birth the HB can be over 200 but drops to lower values within the first weeks, and by 2 months of age reaches its lowest value, (110 g/L).[44] The normal HB range for infants 6 months to 2 years is 110-130 g/L.[33,45-47]

Normal values of hemoglobin for black adults are considered by some to be slightly lower than values for Caucasian adults.[48,49] There is still some disagreement about whether racial differences in HB values exist. An analysis of data from the National health and nutrition examination survey (NHANES) I (1971-72), found that most of the racial difference between hemoglobin values for white compared to black children was explained by differences in TS and therefore likely true differences in iron status rather than racial differences.[50] A study of children from Minneapolis found no differences in hematocrit between aboriginal and white children.[50]

Mean Corpuscular Volume (MCV):

Inadequate delivery of iron to the bone marrow results in less hemoglobin production and consequently smaller cells.[51] In this way a drop in MCV can be seen prior

to the development of anemia in the presence of ID. This volumetric measurement is made in small fractions of a liter, femtoliters, 10^{-15} L, (fL). Measurement by other than an electronic particle counter can result in unacceptable variation.[2]

MCV is slow to reflect the deficiency imposed by experimentally removing a subjects blood at frequent intervals [52-54] MCV also can be elevated as a result of a nutritional deficiency in either vitamin B12 or folate even when there is no ID. MCV and HB may both be lower in individuals with thalassemias. This explains some of the discrepancy in the literature about whether MCV is a sensitive measure of ID. In a young population for whom blood loss and hemoglobinopathies are unlikely MCV can be both sensitive and specific.

MCV values rise steadily between 1 year of age and early adult life. The median value and 95% reference range for ages 1-2 years is 79 fL (67-88).[33] Many studies have used as the lower end for their cut-off to be the 10th percentile, 72 fL.[55] Commonly used cut-off values range from 70 fL[24,44,45,56] to 75 fL.[57]

Red Blood Cell Distribution Width, (RDW):

RDW is calculated from the volume distribution curves produced by automated hematologic equipment. The method of measuring red cell size variation was described in 1974 by England and Down.[58] It has been advocated as a sensitive and early measure of iron deficiency. Unlike the other iron tests, RDW correlates inversely with body iron status, (as iron levels fall, RDW increases).[13]

RDW has been used with RBC in a population with a high prevalence of β-

thalassemia trait to identify those who have true ID.[59] RDW has some limitations relating to the fact that RBC size, like MCV is slow to reflect ID. This may be significant for children.[41]

Biochemical Tests:

Introduction:

Since hematological tests are not reliable when the sample is transported .biochemical tests have some appeal for the remote testing situation. Methods used in the published studies are more diverse than the uniform results published with Coulter Counters. Because of this, different studies are not as directly comparable unless using the same chemical assay.

Free Erythrocyte Protoporphyrin, (FEP):

FEP and RDW are the only tests to correlate inversely with declining levels of iron. Although its role in heme synthesis was elucidated in the 1930s its use in the measurement of ID did not begin until the 1970s.[60,61] The initial method involved spectrophotometry which was cumbersome and therefore not widely clinically applicable.[54]

The newer spectrofluorometric method, used in to analyze the blood tests in this study, was described by Piomelli[62], showing good correlation (r = .97) with a more classical method. It showed good reproducibility with a standard error of 2.3% of the value over multiple paired testings and 5% over the entire range. It was also found to be

stable over repeated determinations weeks apart, refrigerated or at room temperature and even when exposed to laboratory light. It was shown to have a higher correlation with reduced bone marrow iron than did TS.[63] In another study FEP was found to have the least within-person variance.[35]

FEP testing became popular in the 1970s used as a screening test in the US for detecting both iron deficiency, and lead poisoning.[63-65] As the acceptable level for Iron decreased from ≥50 μg/dl to ≥35 μg/dl, FEP became less effective as a dual test.[66] FEP is also elevated in the anemia of chronic illness, certain infections, severe hemolytic anemia, and rare conditions such as certain types of sideroblastic anemia and protoporphyria.[42,54,64]

FEP is also thought to measure a pre-anemic state of iron deficiency.[54] In a 1983 study of 4,160 children including 10% aboriginal children FEP was 88% sensitive when SF < 15 μg/L was used as a reference standard. FEP even identified those without low ferritin values.[64] In a recent study of 26 Danish infants, the highest value was found at 6 months then falling again.[67] Cut-off values for FEP can describe as iron deficient any subject with values above 0.62 to 0.80 μmol/L blood.[60,68]

Serum Ferritin, (SF):

SF is one of the newest iron measures and the one most closely resembling a "gold standard". It is the most direct measure of storage iron and has been shown to reflect the magnitude of iron stores throughout a wide range of developmental changes in infants and children.[22,69]

The radioimmunoassay was first described in 1972 and there have been many validations of its usefulness in the diagnosis of ID in hospitalized patients.[23,70,71] It has also been favorably assessed in many population studies of children and pregnant women.[32,57,72-77] SF was slightly more sensitive than MCV at identifying ID in children >2 years in the NHANES II study.[31] Although therapy with iron can cause an immediate SF increase prior to the complete resolution of deficiency this does not invalidate it as a measurement tool.[78,79]

A major limitation is that SF levels in adults are known to rise up to 50 μg/L in patients with inflammatory conditions even though there is concurrent iron deficiency.[31] Similarly, SF is a less accurate iron test in pregnancy and fever.[80,81] When infection has been present, SF is the only iron test whose values trend away from the direction which indicates deficiency, (ie. SF increases, while HB and TS decline).[42]

In an effort to conclusively evaluate the magnitude of the effect of inflammation, Guyatt et al. conducted a meta-analysis of 55 studies which used a bone marrow aspiration as a gold standard.[15] This study confirmed that in the assessment of hospitalized anemic patients, SF is a reliable estimate of body iron stores and therefore useful in the distinction between iron deficiency anemia and the anemia of chronic disease. Using Receiver Operator Characteristic Curve (ROC) methodology the authors established that with an area under the curve (AUC) of 0.95 it was by far the most powerful test. Being more powerful than the other indicators in their evaluation, they recommended that it be the only test necessary to evaluate iron stores. This recommendation has been taken up by the Ontario Association of Medical Laboratories in

their testing guidelines for adults[82] and in other recent publications.[83]

Some have had reservations about how well ferritin reflects iron stores during the first months of life.[84] Others have reservations about the ability for SF to measure iron status accurately in a population susceptible to frequent infections.[85] In several large studies it has not correlated well with the other measures of deficiency, and specifically it failed to identify a significant number of subjects with anemia.[27,30,85] Although it had reasonable performance in the NHANES II, there were limited SF values obtained in children under two years of age.[31]

A frequency plot of SF values in a population is log-normally distributed.[86] As with the other tests, there is no distinct separation between normal and abnormal values.[87] The lower limit of normal for SF defined as the -2SD value was found to increase progressively from 12 to 21 μg/L with increasing age.[88] Suggested cut-off values range from 16 μg/L[64,86] to 10 μg/L.[55] The most common threshold value used in pediatric studies is 12 μg/L.[57,68]

Several studies have assessed the variability of SF measurements.[89] In a study of adults, both SF and TS had unacceptable day-to-day biological variation. They estimated that 3-10 independent measurements were required to accurately determine values.[36] Another recent study found less day-to-day variation than had been previously suggested.[84] As with HB, capillary measurement of SF have shown only slightly more variation than venous samples.[90,91]

Transferrin Saturation, (TS):

For many years it was the only non-hematological method to test iron status.[92] Consequently it has been well studied in the hospital setting, [93] and in the community.[27,73,78,87,94] It clearly measures ID in a stage prior to the development of anemia and is a valid test across a range of deficiency severity. Some investigators have found it to provide additional testing benefit over serum ferritin.[30,95] However many studies have shown some diagnostic validity of SF over TS.[26,32,96] As with many of the other indicators, iron and TS increase with age.

As a composite indicator, it is more reliable than either of its two components, Iron and TIBC. Iron is the most labile of tests with both significant day-to-day and within days variation. TS results can vary by approximately 20% of the morning level.[3] In children less than 3 years of age, though, the amount of diurnal variation has been shown to be limited.[97]

Perhaps some of the discrepancy in TS performance between studies is related to different choices of the deficiency threshold. The most widely accepted cutoff in adults is TS <16% by the International Nutritional Anemia Consultative Group, (INACG).[68] In NHANES II a cut-off value of the 10th percentile was selected which for the 1-2 yr age group was <12%.[31,33] Other studies use a cut-off value of <10% for infants,[13] or 7%.[98]

Test Summary:

As one can see with a host of iron-tests, each with strengths and limitations compared against different reference criteria, the selection of "best test" is problematic. (A summary of each of the test properties is presented in Table 1.) In this confusing array

it is difficult for either the epidemiologist or the clinician to know which test to apply to populations or individuals. Reviews have suggested that since independent information is available from each test, population surveys should employ a battery of laboratory measurements of iron status.[87]

Physiology of Iron Deficiency:

Iron Absorption:

An understanding of the physiology of an infant's iron metabolism is critical to enable one to follow trends in iron status. It also assists in an understanding of iron interventions such as formula-fortification. With this knowledge we can better evaluate the relative strengths of iron tests.

Most infants begin life replete with iron received through the placenta from their mother. It was previously thought that this transfer takes place regardless of the mother's iron status.[99,100] It has now been shown that although iron is absorbed throughout pregnancy a maternal deficiency in the third trimester as measured by SF is most predictive of low iron stores in the infant.[101-103] A study using radio-labeled iron at one year of age indicates that 70% of body iron exists from what was transplacentally acquired prior to birth.[104]

Iron metabolism is maintained in a delicate homeostasis with little iron lost. The conservation of iron is facilitated in several ways. When a blood cell dies after an average lifespan of 120 days the HB is metabolized into bilirubin and other products which are

then mostly reabsorbed. In this way there is seldom a significant net loss of iron. Of the 5,000 mg of body iron in the adult probably less than 1 mg is absorbed from the diet per day even though 25 mg of iron per day turns over within the hemoglobin cycle.[105]

The recommended amount of iron required by an infant has changed a great deal over the years.[106-108] Currently 0.5 - 0.9 mg of iron per day is advised.[44,109] There are three ways of administering iron to an infant, supplementation, (providing iron as a medication), the fortification of formula and the fortification of infant cereal. At first, breast-fed infants do not seem to require extra iron. Although breast-milk only contains small amounts of iron (0.2-0.4 mg/L), 49% of this iron is absorbed.[110] By six months of age, however, most authorities recommend the addition of iron rich foods to the diet of breast-fed infants.[111,112]

The iron-fortification of infant cow-milk formulas has been recommended as a universal strategy since the amount and bio-availability of iron is limited. Although the range of fortification is wide (12.8 vs 1.1 mg iron / L), it has been shown by many studies to be efficacious.[44,113-119]

Sometimes infants are low in iron because of a major hemorrhage which may have occurred during birth or from having been born prematurely. For these infants and the infants born to iron deficient mothers, early supplemental iron is important.

For the most part, the administration of these physiological amounts of iron is seen as innocuous. A theoretical problem, however, with the administration of large amounts of iron to a deficient individual is the potential to promote infectious disease because the iron may be more bioavailable to pathogens than to the host.[120] It has even

been speculated that iron deficiency is a protective mechanism against infection in tropical countries.[121]

Prevalence of ID:

Prevalence measured by HB:

For much of this century the health and nutrition status of infants and children has been partially estimated through a measurement of iron deficiency. Most of these measurements have been made using hemoglobin measurements of various types. The historical and geographical context of ID will be reviewed here.

The disadvantaged children of urban centres have long been a population with nutritional deficiency. Helen Mackay in 1928 was the first to find an almost universal prevalence of anemia in the poor children of London's east end. She also demonstrated the efficacy of iron-fortification.[122] Guest found that 30% of American urban children were anemic in a study conducted on 1735 children living in Cincinnati in the 1930's; this was unchanged when reinvestigated in the 1950's.[123] Katzman, in 1968 found a 12.5% prevalence of anemia in the 10-36 month age group of ethnically diverse infants from New Haven Connecticut and noted a relationship to ethnic group as well.[124]

In response to these trends, governments implemented public health programs to address iron deficiency among disadvantaged children.[117] In addition to iron rich food and supplemental iron being administered, an evaluation of the effectiveness of the programs was conducted. Children from six states were consistently monitored by the Centre for

Disease Control Pediatric Nutrition Surveillance System. Socioeconomic status as measured from birth records of maternal age, education level and employment were examined to ensure the comparability of the populations at both intervals. These programs were shown to have had some effect on this problem as documented in 1987 by Yip et al.[125] In this study the prevalence of IDA among low income US children declined from 7.8% in 1975 to 2.9% in 1985.

Unfortunately this improvement has not been seen elsewhere in the US, perhaps because of less nutritional intervention. In a group of 1 year old San Francisco children in 1981 the rate of anemia, as measured by HB <110g/L, was 25%.[24]

Prevalence as Measured With Other Tests:

For many years the quantification of iron deficiency was limited to a measurement of HB or hematocrit (HCT). These measurements only quantify the numbers of children with iron deficiency anemia (IDA). It has been increasingly recognized that counting only those with frank anemia misses the number of children with suboptimal iron stores. Because ID typically precedes frank anemia there will always be more children who are iron deficient than those with actual anemia.

By seeking to identify those with a reduction in iron stores without actual anemia we see a much larger problem. In a study of 109 infants and children of low income families from Michigan, Haddy et al used TS <16 as well as HB <110 g/L and found a prevalence of ID of slightly more than 50%.[126] In Montreal in 1989 a study of children found a 25% prevalence of anemia (Hemoglobin below 115 g/L) but actually 37% of

children had evidence of iron deficiency, (SF <10 μg/L).

There are many international studies which show similar results, mostly in developing countries or socioeconomically deprived children. (Table 2) Research conducted within developing countries assists in an understanding of ID within disadvantaged populations of North America such as the urban poor or those in Aboriginal communities.[127] Although co-morbid conditions such as parasite infestations and hemoglobinopathies are uncommon in native communities, the nutritional deprivation is analogous.

Prevalence within Aboriginal Populations:

The archaeological record of North American Indians possesses evidence of iron deficiency within individuals of all ages from time to time. The unaided eye examining skeletal remains, usually the scull, can see small holes of various sizes called Porotic Hyperostosis. There is general agreement that this represents acquired iron deficiency anemia but there is some debate as to its cause or possible evolutionary benefit. Traditionally the scientific community has maintained that the observation of porotic hyperostosis represents a period of nutritional stress. A newer explanation is based partly on the assumption that the presence of iron provides substrate for not only the individual (host), but also the infectious agent. In this situation the withholding of iron is seen as a successful adaptation to pathogen stress. Regardless of the explanation, ossuary samples frequently present evidence for a high prevalence of ID within aboriginal groups. In the Fairty ossuary the prevalence among children was 38%.[128]

Data have been collected and analysed from adult Inuit and Indian populations in North America even other parts of the Arctic. Many of the studies involve the measurement of HB and presentation of mean HB rather than prevalence. (Table 3)

Valberg et al in 1979, when building on the Nutrition Canada 1973 survey studied aboriginals in theYukon and NWT. They used HB, TS and SF to profile iron deficiency in Inuit populations, finding a lower prevalence in the former and higher prevalence in the latter. No Inuit children ages 1-4 years met their strict criteria for IDA but 21% of 19 individuals in this age group had a low SF and 47% had a low TS while only 5% were anemic.[32]

An early study of Indians from reservations in Manitoba in 1952 showed that 73-85% of infants under 2 yrs of age were anemic.[129] Although the extraordinary prevalence of anemia has declined since then, there is still a greater than 50% prevalence of iron deficiency. (Personal Communication, M.E.K. Moffatt. 1996)

In summary, the prevalence of ID has been disproportionately high within groups of disadvantaged infants. Children living within the urban core and aboriginal children from urban or remote areas have had an alarmingly high rate of ID and IDA. Programs to improve their nutritional condition are required and will need to be monitored and evaluated using an accurate and reliable iron-testing method.

Clinical Epidemiology:

Introduction:

Comparison of the relative strengths of each test to measure iron deficiency is assisted by the tools of clinical epidemiology. Initial comparisons between tests can involve the capacity of each test to measure the prevalence of ID within the population. Evaluation is based on juxtaposition of these prevalence estimates. These estimates, however are dependent on the cut-off values chosen.

For a proper evaluation, a test needs to be measured against an external reference point. If the test measures the individual as iron deficient and the external standard agrees, then the test result is a "true positive". The number of tests which are "true positives" divided by the total number of subjects with ID gives a result called sensitivity. "True negatives" divided by the total number without ID is called specificity. For screening purposes the test usually needs to be quite sensitive and as specific as possible but less importantly because later confirmatory tests can be applied. Sensitivity and specificity calculations are limited by the following factors: choice of gold standard (reference criterion) and choice of a cut-off value.

The first problem is encountered because there is no authoritative external reference, or "gold standard" for comparison. If, for example, SF measurements were chosen as the gold standard for measuring ID, the performance of other tests could be measured for how frequently they agreed with the determination made by SF criteria. In order to make a dichotomous determination of positive or negative, a cut-off level needs to be applied both to the test under study and to the reference test.

While sensitivity and specificity can help to describe the capacity of a test to work within a screening context, their utility must be interpreted in light of the prevalence

of the condition within the population under study. The actual predictive value of an iron test is proportional to the prevalence of ID in the population, (pretest probability). If we know that there is a high prevalence then the positive test is likely to be a "true-positive". If there is a low prevalence, even if the test is very sensitive, the test is likely to be a false positive, except under conditions of perfect specificity.

Another limitation of sensitivity and specificity measurements is that they are dependent on binary decisions of positive or negative based on cut-off values. An assessment method independent of these cut-offs has been developed which involves the plotting of sets of paired sensitivity and specificity combination for the test. The resulting line is called the receiver operator curve, (ROC). The area under this curve can be used to assess the relative merits of different tests.

Each of these limitations and solutions will now be discussed in turn:

Normal Values and Cut-offs:

When tests are to be applied to large populations, it is necessary to define a normal range. Studies for this purpose utilize the observed frequency distribution in a large group first to make conclusions about normality. If the curves suggest that it does not follow a gaussian distribution, logarithmic transformation and the use of a geometric mean are required.[86] The more representative the sample is of the general population and the larger the study population, the more generally applicable the estimated normal values are. The distribution of a variable in a population is usually an overlapping bimodal one. A cut-off point chosen to identify the maximum number as iron deficient while labeling

only the minimum number of normals as deficient will always encompass those with and without ID. (Figure 6) For a unimodal variable the same tension exists to chose a lower end of normal which appropriately selects those in need of treatment.

Usually the mean but sometimes the median and 2 standard deviations (SD) of a group of healthy individuals are used. The exclusion of "unhealthy" individuals, however, is somewhat arbitrary in that exclusions must be on the basis of individuals having values below subjective cut-off levels of other iron tests.[45,47] A specific example of the calculation of normal values is found in Appendix IV for NHANES II. Similar normative data have been obtained from various sources. Nutrition Canada Survey in 1973.[39] Preschool Nutrition Survey, Ten State Nutrition Survey[49], NHANES I and patients within the Kaiser Permanente health care group in San Francisco.[45]

Receiver Operator Curve, (ROC) Methodology:

When using an iron test, one is interested in the ability of the test to correctly identify a subject as deficient or not. The capacity to make this determination is partially represented within the concept of sensitivity. The opposite characteristic of a diagnostic test's performance is specificity, which describes the alternate situation in which individuals are correctly identified as "not deficient". Since few tests typically yield binary outcomes; results must be categorized as either positive or negative. When dealing with a continuous measurement a "cutoff" point needs to be chosen as a threshold below which we can define individuals as unaffected and above which we can define the individuals as affected. The sensitivity and specificity of a test usually vary inversely

according to the cutoff value chosen. The more inclusive the definition of deficiency the more sensitive but less specific. Hence an understanding of a testing properties involves a contrast between the sensitivity and specificity of tests at different cutoff values.[130] The Receiver Operator Characteristic analysis (ROC) is a method to evaluate the diagnostic accuracy of a test independent of its threshold values. ROC methodology will be outlined below.

General Principles of ROC Analysis:

ROC analysis was first used to separate signal from noise in the early development of radar during the second world war.[131] It was later developed to assist in the assessment of variability in the interpretation of radiographic films. In the last decade it has been widely applied in the assessment of diagnostic testing.

In 1986 Brownie, Habicht and Cogill were amongst the first to use ROC methodology to evaluate iron tests. In their paper they asserted, "*The performance of an indicator of health or nutritional status depends on sensitivity and specificity properties over a range of cut-offs*".[132]

There are three reasons to construct ROC curves. One is to display a visual representation of the testing value of an indicator. (Figure 7) First, the shape of the curve is instructive. Second, the area under the ROC allows us to statistically compare the tests without reference to the specific cut-off point. The final benefit is to assist in the selection of a cutoff point. It is possible that ROC analysis may suggest a different point than has been determined by population normal studies. There are weaknesses of using ROC to

select a cutoff point, chiefly that the empirical use of the curve doesn't allow one to factor-in the relevant costs and benefits inherent in the choice of a cut-off level.[133] Because of this, alternate methods exist, such as decision analysis which contrasts the implications of net cost, net benefit and prevalence.[66,134,135]

The selection of this threshold is problematic because one has to weigh the cost of not detecting individuals because of an insufficiently sensitive test against the consequences of incorrectly identifying individuals as affected by using an insufficiently specific test. The first step in selecting a cutoff point is to construct a graph of the true positive ratios (sensitivity) against the false-positive ratios (1-specificity). The resulting plot is known as the receiver - operating - characteristic curve, (ROC).[136]

Not only can the testing characteristics of the curve be graphically demonstrated for the empirical choice of a cutoff value, (operating point) but two curves can be compared against one another. Since the scale of the ROC plot of sensitivity vs 1-specificity is on a scale of 0-1 on both axes it is independent of the units used in the individual tests and therefore different indicators can be compared. The perfect test would result in a curve going immediately from 0,0 to 0,1 and then drawing a horizontal line to 1,1. Realistically curves are convex with better tests pointing to the 0-1 point. The weaker the test, the more it describes the 0,0-1,1 line of "no information". The areas under the curves (AUC) can also be compared and the statistical significance of the difference measured.[137,138] The AUC also represents the probability that a randomly chosen diseased subject is (correctly) ranked with greater suspicion than a randomly chosen non-diseased subject.[133,139]

Calculations of the AUC can be made using methods which assume a normal distribution of the tests. The maximum-likelihood method developed by Hanley and McNeil has been widely employed in the calculations of individual test areas under the curves which are presented with the standard error (and symmetrical normal 95% confidence limits).[137] In this way curves can be compared one with another. Small deviations from normality are not thought to invalidate this method.

Curves can also be compared with the "line of no information", a diagonal line connecting the coordinates 0,0 and 1,1. Calculations involving the non-parametric Mann-Whitney U statistic are used to compare the AUC with the line of no information. The "Z statistic" is used to determine the degree to which the test is superior to the line of no information.

It has been shown that non-parametric formulations have statistical power that is equal to the model outlined above.[140] In addition, they do not make assumptions about the distribution of the data.[141] Although the nonparametric "trapezoidal rule" for calculating AUC usually yields a slightly smaller AUC than the parametric smoothed curve, this difference is not clinically significant if the non-parametric curve is composed of five or more cutoff points. The AUC of different tests can be compared with one another using a method which takes into account the correlated nature of the curves. Because different iron tests were performed concurrently on blood taken from the same individual we can capitalize on the inherent statistical power of drawing curves from "correlated" samples.[142-144] This is similar to the use of paired t-tests when using the same subjects as a way of minimizing variation. These curves are constructed using the method

of DeLong, DeLong and Clarke-Pearson.[138] The resulting test has a chi squared distribution which can be used to test the statistical significance of the differences between the AUC of our different tests. Iron tests with larger AUC measurements have superior discriminative capacity independent of any chosen cut-off values.

Screening:

Screening Principles:

Screening is intended to identify unrecognized disease in apparently well individuals. A screening program is not intended to be diagnostic, but rather to identify those individuals who might benefit from further testing or intervention.[18] In 1968 Wilson and Jungner outlined principles for early disease detection:[145]

1. The condition sought should be an important health problem
2. There should be an accepted treatment for patients with recognized disease
3. Facilities for diagnosis and treatment should be available
4. There should be a recognizable latent or early symptomatic stage
5. There should be a suitable test or examination
6. The test should be acceptable to the population
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood
8. The should be an agreed policy on whom to treat as patients
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10. Case-finding should be a continuing process and not a "once and for

all" project

The importance of ID in relation to infant psychomotor development as well as the efficacy of an iron intervention have been well described.[6,7,44,146,147] The choice of a suitable and acceptable test as well as the policy of whom to treat has been the main focus of the literature review. Decisions about the economic justification of a screening program will be addressed in the discussion. The approach to these decisions is conditioned by the recommendations of certain expert groups.

Screening Recommendations:

Canadian recommendations for the screening of iron deficiency anemia are based on principles of evidence-based medicine. New evidence regarding the psychomotor effects of IDA published since the original 1979 recommendations caused the Canadian Task Force on the Periodic Health Examination to reexamine the evidence. The limitations in sensitivity of HB were noted, the low prevalence of IDA as well as the mixed success of iron therapy on mental and motor milestones. The Task Force concluded that there was insufficient evidence to support a general screening of Canadian infants.

Instead they promoted the selected screening of infants from families of lower socioeconomic status. Those with suspected poor diets should receive nutritional assessment and the consideration of HB testing between 6 and 12 months. Where there is suspicion of poor iron intake or poor general nutrition then both HB and SF were

advocated.[148]

Early detection was contrasted with an alternate approach of primary prevention. They also advocated dietary recommendations similar to those of the Nutrition Committee of the Canadian Pediatrics Society who recommended iron-fortified formula for non-breast-fed infants. They outlined the need to postpone whole cow-milk until one year of age as well as the use of iron-fortified cereals for all term infants. They made no recommendation regarding detection of IDA.[149]

The U.S. Preventive Services Task Force in 1989 made recommendations that a screening venous hemoglobin should be offered once to all infants and that parents should be encouraged to include iron enriched foods. They explained that the cost of SF, TS, and EP made them unsuitable for screening.[150] The American Academy of Pediatrics recommended at least one HB or HCT in infancy and at least one ages 1-4.[151] HB was the preferred test followed by a therapeutic trial of iron for those below threshold rather than further testing.

Although American recommendations seem to favor a more comprehensive screening approach, the Canadian guidelines do advocate a testing program for high-risk individuals and populations. A population based approach may involve a selected screening program according to knowledge of risk-factors or prevalence. Some of this information is determined by population health surveys. Wilson and Jungner made a careful distinction between screening programs and population health surveys.[145] The attempt to delineate the prevalence of a disease and its natural history within a population, while it may involve the identification and treatment of individuals, is more

directed at the quantification of the extent of disease within the community. The conduct of a population health survey is usually more intensive and may include an evaluation of the effectiveness of an intervention. In the context of iron testing, multiple tests are advocated for population surveys because of the limitations of each of the individual tests.[3]

This information is then used to assist in the health planning process which may or may not suggest a screening strategy or one of primary prevention. Primary prevention describes the systematic effort to reduce risk factors for disease in the entire population. In this situation it would mean trying to eliminate any dietary iron defecits through a program which includes breast-feeding promotion and the administration of additional iron-fortified infant cereal at four to six months. For those not choosing to breast-feed, it would include the uniform administration of iron-fortified formula and iron-fortified infant cereal.

Methods:

Summary of the 1992 RCT of Formula Fortification:

In order to understand the context of the data and the population from which the data for this thesis are drawn it is necessary to first present a brief overview of the methods and results of the original study. The RCT was designed to evaluate the impact of iron-fortified formula on infant mental and motor development.[113]

From June 1988 to April 1992 infants were selected from the general population of infants attending the Winnipeg Health Sciences Center (HSC), outpatient clinics for entry into a study. Inclusion into the study required that infants be less than two months of age and have mothers electing to use bottle feeding. Infants with conditions which could affect growth, nutrition or influence a developmental assessment were excluded.

Informed, signed consent was obtained by the research associate after an interview on the maternity ward, at the HSC outpatient clinic or at one of two nearby community clinics. Breastfeeding mothers were not approached for entry into the study. The study was approved by the University of Manitoba Faculty of Medicine Committee for the Study of Human Subjects in Research.

The effect on children of the consumption of iron-fortified formula (12.8 mg iron per liter) versus regular formula (1.1 mg iron per liter) was measured. Attention was paid to conventional blinding and randomization procedures. Analysis was conducted on the descriptive variables by treatment group to determine the effectiveness of randomization. The equivalence of treatment groups was confirmed in the original paper.

Data Collection:

Figure 8 presents a summary of the timing of data collection and the numbers of infants available for assessment. Although not all subjects had the complete set of 6 iron tests at each follow up visit there was a cohort of 225, 204, 186 and 154 at the 6, 9, 12 and 15 month assessments respectively. (Table 4)

For the original study two types of data were collected: 1) variables describing the bio-demographic characteristics of the infants and their families and 2) variables related to their iron status or psychomotor development. Demographic and descriptive data were collected during interviews conducted prior to the infants' third month of life and before randomization. Sociodemographic variables were collected and the information was supplemented by a review of the hospital records.

Sociodemographic Variables:

Since these visits did not actually take place on the infants' birthdays, the range of ages are described at each "assessment age". Age was recorded in months as a decimal. (Table 4) The study group was composed of 55.4% (124) males and 44.6% (100) female approximately evenly divided across both arms of the study.

The original study examined other sociodemographic variables and measures of infant stimulation to examine for inequality between groups and with the exception of a slight excess in history of maternal alcohol abuse (17/33 vs 8/25) none was found. The study employed Hollingshead's "Two Factor Index of Social Position" as a way to profile the economic and social status of the infants mothers.[152] This index was chosen to detect more subtle deficiencies in the original randomization.

Clinical Measures of Iron Status:

Although descriptive variables were collected at approximately 2 months of age there were no blood tests taken until several months after the intervention with iron-fortified formula had begun. There were four follow up assessments during the study which were spaced three months apart. Because these took place at approximately 6, 9, 12 and 15 months of age they will be referred to as the "six month assessment" or "nine month assessment" etc.

During each of the four visits, after an assessment of infant development, the study nurse drew venous samples from the infant. These were taken from 9:00 to 4:30 p.m., though most were in the morning. The samples were transported the same day to the laboratory and were handled according to protocol. All laboratory testing was conducted in the clinical laboratory services of the Health Sciences Center of Winnipeg which takes part in a rigorous internal and external quality control program.

Coulter Counted, Hematologic Variables:

The HB, MCV and RDW were automatically measured using a Coulter Counter (model STKR; Coulter Col, Hialeah, Fla.). Results are reported in SI units; HB g/L, MCV fl., and RDW %.

Biochemical Variables:

Free Erythrocyte Protoporphyrin, (FEP) μmol/L was measured using the extraction and fluorescence method of Piomelli ;[62] Serum Ferritin, (SF) μg/L., was

measured with the Abbott IMx method (Abbott Laboratories, Abbott Park, Ill.); Transferrin Saturation (TS) % was calculated using the formula: Serum Iron / Total Iron Binding Capacity (TIBC) x 100. Serum Iron was measured with the Kodak kit (Eastman Kodak Col, Rochester, N.Y.) and TIBC with the Kodak kit (with magnesium citrate precipitation).[113]

Methods Used in This Thesis:

Introduction

This thesis used secondary data analysis of a longitudinal data-set containing multiple measures of iron status on a cohort of aboriginal children to answer the main question: Which test or tests of iron deficiency are most appropriate to apply to the measurement of iron deficiency in high risk and potentially remote aboriginal populations?

Data Management:

Data Transfer: A data file was transferred from the mainframe of the University of Manitoba computer via diskette and installed on a 586 desk top computer running Windows 95 and Sas 6.11. The file contained unique subject ID numbers and variables connected to each individual, however no personal identifiers were included in order to maintain confidentiality.

Using SAS software (SAS Institute, Inc., Cary, N.C.), data were manipulated to

examine the relationships between different measures of iron status and the extent to which they measure the outcome of interest, iron deficiency. New variables created to answer the specific questions are outlined below.

Ethics:

Ethics approval was obtained from the University of Manitoba Faculty of Medicine Committee for the Study of Human Subjects in Research for the secondary analysis of this data. Confidentiality was ensured through the absence of personal identifiers in the dataset.

Data Perspectives:

The structure of the data permitted the examination of information in three different ways. First, cross-sectional analysis was conducted, (**Figure 9**). It was possible to make four distinct sets of observations at ages spaced three months apart. At each of the intervals age-specific information about iron status within this population was examined. Secondly, comparison of the data from each of the intervals allowed for the assessment of changes in the group over time. Within each of the cross-sectional studies, differences between the testing properties of each of the iron tests was examined.

The third type of analysis involved the examination of the relationships between an individual's iron test at one assessment and the same individual's iron test(s) at a subsequent assessment. Finally, iron test results were analyzed according to treatment

group. The group receiving iron-fortified formula represented a lower prevalence population. An assessment of the performance of indicators within this lower prevalence population was therefore possible. The data from infants receiving regular-formula permitted the assessment of test performance in a higher prevalence population. Data from both groups combined contained a complete range of iron status values.

Analysis of Iron Status:

Means:

Since values for the iron tests are continuous variables. Mean values and standard deviations, (SD) were calculated for all of the study subjects together as well as separately for those receiving iron-fortified formula and those receiving regular formula. Logarithmic transformation of the serum ferritin values and calculation of geometric means was necessary because results did not follow a normal distribution.

Correlation Coefficients:

In order to evaluate the degree to which tests seem to measure the same property, Pearson Correlation Coefficients were calculated for all tests at the six month visit and presented in tabular form together with "p" values as a test of statistical significance.

Mean Values Within Ranges of Hemoglobin Values:

Hemoglobin values provide a familiar frame of reference for clinicians and "anemia" has been the traditional proxy for ID. We therefore grouped study subjects according to their hemoglobin values by ranges of 10 g/L. The expectation was that those

individuals within the lower deciles should have lower values of SF, TS and MCV and higher values of FEP and RDW. It was also expected that the mean test values from each HB group would parallel the HB increments. The mean value of each test within the HB=100 cluster was examined in order to suggest an emperical "cut-off" point for each test.

Operationally, all HB values were divided by 10, made into an integer, then multiplied by 10, dividing the study subjects into the following 6 groups:

80 = those with HB values <90 g/L

90 = those with HB values <100 g/L

100 = those with HB values <110 g/L

110 = those with HB values <120 g/L

120 = those with HB values <130 g/L

130 = those with HB values > 130 g/L

Selection of "Cut-off" Points:

Having reviewed the literature regarding the wide range of suggested "cut-offs" for each of the tests it was difficult to select the most reasonable point for application to this population and age group. This was especially true for MCV and FEP. In order to verify the chosen threshold value, the effect of varying this "cut-off" point on respective prevalence estimates at the 9 month assessment was examined.

Estimates of Prevalence Based on Published Criteria:

The prevalence of iron deficiency was calculated for the total study group, for those who received iron-fortified formula and for those who did not. The estimated

prevalence of ID was defined as the number of individuals possessing a value below the accepted threshold for each indicator according to published criteria. These widely accepted thresholds were determined using population normative data from large studies. In most cases the mean and two standard deviations of values for populations of well children define the "normal" range, below which subjects are defined as deficient. Using these criteria each test was compared to the others according to the number of individuals that each test identified as iron deficient.

For HB an individual was designated deficient if below 110 g/L. Most studies support this level though some accept cut offs at 105 and some 115.[33,45] For MCV the choice of cut off from the synthesis of literature is between 70 and 73.[33] In a study of 2,314 white children near sea level without other evidence of anemia, Dallman et al determined the lower limit to be 70 for children 0.5 to 2 years of age.[45] Other sources also specifically support a cutoff of 70.[13,14] The utility of this value, however, seemed limited as evidenced by results of the data analysis. Operationally many pediatricians working with populations of aboriginal children to detect ID employ the 74 fL cut-off, (Personal Communication, January 1997, M.E.K. Moffatt). This study therefore utilized a cutoff of 74 fL.

The accepted threshold for RDW is also controversial and like MCV may vary with the actual automated machine making the determinations. While the general cut-off is recommended as 18%[13,58], at the laboratory where samples for our study were analyzed 11.5-14.5% is the accepted range for all ages. For this study it was decided that any infant with a value above 14% was to be considered iron deficient and this is consistent

with other published studies.[153,154]

For FEP, there is some debate about whether the upper limit of normal should be 35 or 45 μg/dl based on population studies of children.[88] In 1993 the American Academy of Pediatrics (AAP) Committee on Nutrition recommended a threshold of 30 ug/dL.[108] For this study the determination of deficiency is based on the accepted threshold of "any value greater than 35 μg/dl".[13,66] Because the laboratory values used in this study were in S.I. units, the conversion factor 0.0177 was used to arrive at the threshold >0.62 μmol/L.

For SF, many studies use a cut-off value of 10 μg/L for this age group, [24,55] but the more widely accepted deficiency definition for children in the first two years of life is any value less than 12 μg/L.[87,88] For TS any infant with value less than 10% was considered to be iron deficient.[13]

Estimating Sensitivity in Absence of a Gold Standard:

For most published studies which evaluate diagnostic tests any new test is usually compared to an established test, ideally one that is the most accurate and therefore referred to as a "gold standard". "Sensitivity" is a measure of the true positive fraction of the new test. In the context of this thesis "sensitivity" is ideally the proportion of individuals identified as iron deficient who are *truly* iron deficient.

In the absence of a true "gold standard", to arbitrate which individuals are correctly identified as deficient from those falsely identified, this study was limited to comparing each test against all others. We began by sequentially allowing each test at an accepted cutoff value to be the surrogate "gold standard". If we consider an example in

which MCV is the "gold standard" and attempt to calculate the sensitivity of HB we can construct a 2 x 2 table with MCV on the horizontal axis and HB on the vertical axis. Study infants can fall into one of four cells;

a: true positives (iron deficient according to both tests)

b: false negatives (iron deficient according to HB but not according to MCV)

c: false negative (iron deficient according to MCV but not according to HB)

d: true negative (normal values according to both tests)

The "sensitivity" would be a/a+b.

The specificity would be d/b+d

If we then alter our perspective to consider the sensitivity of MCV with HB as the "gold standard" along the horizontal axis the 2 x 2 table rotates about the diagonal axis from the previous sample calculation. The values in the cells change because although the numerator is the same, the denominators change. The totals for the rows now become totals for columns.

To present the multiple combination of 2 x 2 tables of sensitivity and specificity would be unwieldy. Unfortunately there is no combined measure of "average sensitivity" and so the series of sensitivity/specificity calculations are presented in a larger composite table patterned after Mahu, 1990.[13]

The reference test is presented along the top row, with the next row indicating the numbers of subjects identified as deficient by that test at the accepted threshold value.

Below that, the estimated sensitivity values of the comparison tests are presented in turn. By scanning across a row for each of the tests one can gain an appreciation of the relative sensitivity of the test as well as which test is most concordant with which.

The false negatives, false positives and the total numbers of subjects are not displayed in the sensitivity or specificity table in order to simplify presentation. An evaluation of sensitivity requires that both iron tests were conducted on each subject. The total number of subjects under consideration in each cell varies with each cell of the table. It is a function of how many individuals had a laboratory determination for both tests.

A table for estimated sensitivity and one for estimated specificity was constructed for values at the 9 month assessment, which is the conventional age selected for ID screening.

ROC analysis:

Sensitivity and Specificity Measurements at Different Cut-off Values:

Introduction:

As discussed in the previous sections, the selection of cut-off values, while governed by scientific methodology, is limited by inherently empirical and subjective choices. Even the estimated sensitivity calculations involved the use of a surrogate gold standard which is only as accurate as the cut-off value selected. It was therefore desirable to utilize a technique which could evaluate the relative strengths of different tests

independent of cut-off values. Consequently the remaining analyses were conducted using receiver operator curve analysis, (ROC) to compare the size of each tests' area under the curve, (AUC).

Because one of the tests, SF, has values which do not follow a normal distribution and because non-parametric methodology does not significantly reduce the ROC AUC when more than 5 points are available, a non-parametric ROC method was chosen. The method of Hanley and McNeil was used to compare individual curve's AUC with the line of no information.[143] The resulting test is called a Z score. Computer software, (ROC 200), was used to assist in the complex plot and calculations involved in working with nonparametric distributions of the indicator. This software performs chi squared calculations to provide two-tailed p values in order to test the significance of differences between each of the tests' area under the ROC curve and the theoretically minimum area which would provide no discriminative power, (the area under a diagonal line on the ROC graph, the line of no information).[155]

It was also possible to compare each of the iron tests with one another for statistically significant differences. Because the data structure was such that each laboratory determination was made on the same subjects, the between subject variability was limited and the method of "correlated ROC curve comparison" was utilized.[138,139,142,143] This analysis is similar to paired t-testing and allows for greater inference to be drawn from smaller subject numbers. Comparisons were conducted between two and three tests at a time.

The software employed for this analysis required that several changes be made to

the variables. The program required the use of integers between 0 and 200. In order to supply only integers for analysis the original data values were "rounded up" by adding 0.5 and taking the integer. (This was necessary for MCV, RDW, FEP and TS). In the case of FEP, the values had to multiplied by 10. Because the method involves rank-ordering the sensitivity-specificity pairs, changes in absolute test values do not affect the comparison. There is no bias resulting from reducing the number of significant digits in the laboratory value or in "rounding up" since the rank order of test results remain the same. Although the loss of significant digits results in less unique laboratory values, it does not effect the number of points on the sensitivity/specificity graph as long as the number of points is greater than 5.[133,139]

The program also assumed that higher scores, rather than lower, are associated with a higher risk of having the disorder in question. Since our outcome of interest is iron *deficiency* and lower values for HB, MCV, TS and SF are associated with deficiency, in order to make ROC curves sweep in the proper direction, data values were transformed by subtracting their respective values from the maximum value obtained for each iron test at that assessment age. For RDW and FEP this transformation was not required.

In each of the analyses subsequently described we compared the test under consideration with a reference test or series of tests. Again, because no "gold standard" was available, it was necessary in *each* of the following sections to define a "gold standard" according to the specific objective of the analysis at hand. These alternate reference standards were derived from whichever tests were not being compared at that assessment age. In two analyses the longitudinal nature of the dataset was used to

demonstrate the ability of tests to predict an outcome three months later at the next assessment age. The outcome of interest, an ongoing drop in HB or a rise in HB in response to treatment became the de-facto "gold standard".

Gold Standards:

The first analysis in this series used evidence of a rise in HB in response to treatment as presumptive evidence of prior ID. Even those individuals with a normal HB value, if they increase by a significant amount after receiving iron are deemed to have been deficient. The second analysis, based on the theory that iron deficiency happens in stages considered that a drop in HB could be anticipated by low values of other tests which are thought to precede anemia.The third analysis, capitalizing on the generally accepted position of ferritin as the best measure of iron stores, identified individuals below a threshold value of SF as deficient. This served as a concurrent gold-standard.

In the final analysis a composite definition of deficiency was employed using three hematological tests in parallel. Individuals were designated as deficient if they had any of the following: anemia, (HB<110 g/L), microcytic cells, (MCV <74 fL) or anisocytosis, (RDW >14.5 %). In this analysis the relative strengths of FEP, SF and TS were compared.

ROC analysis with "Gold Standard": Responders to Iron Therapy:

Previous studies have demonstrated that one of the best ways of truly identifying which individuals have ID is to document an increase in the HB value after treatment with iron. Responders to iron therapy are retrospectively defined as having been iron

deficient. This study utilized the longitudinal nature of the data-set to select those individuals from the treatment group whose hemoglobin rose by 0.5 g/L between the first and second or the second and third assessment age and tried to determine if the other indicators identified them as iron deficient at the previous visit. All subjects within the fortification arm of the study were considered for this analysis because regardless of the amount of iron received the objective was to determine if an iron test at time-1 would predict that a rise in HB should take place at time-2 with iron administration. ROC analysis was used to examine the sensitivity and specificity of iron tests separately at 6 and 9 months to identify those who would respond to iron therapy at 9 and 12 months respectively.

ROC analysis of Hemoglobin Trajectory:

Hemoglobin levels which in some subjects fall over time should be preceded by changes in the other indicators. The longitudinal nature of this data set provided the opportunity to observe the natural evolution of hemoglobin levels in all subjects, but especially in those who were not receiving iron-fortified formula. This analysis was conducted on the whole population rather than on just the unsupplemented arm of the study because even some individuals within the iron-fortified arm still exhibited declining HB values. Within the fortified arm of the original study the amount of supplementary iron was much smaller than the dosage conventionally used for the actual treatment of anemia. Five individuals with profoundly low hemoglobin values (HB <90) received treatment doses of iron in the original study. These individuals were included in this

analysis because although the treatment with iron may have created "noise", the results of the analysis with them included provided a clearer separation between AUC values. These results are presented in tabular form with AUC and Z statistics. Separate calculation of correlated ROC curves are presented including the χ^2 value with a $p_{2tailed}$ test of significance..

ROC Analysis with Serum Ferritin as a Reference:

The literature suggests that SF is the first parameter to become abnormal in the development of a deficiency state. It is possibly also the first to correct when iron status is being rectified. This may happen even ahead of the hematologic indicators (HB MCV and RDW). We therefore examined the performance of the non-ferritin tests, (HB, MCV, RDW, FEP and TS) with a ferritin standard in the unsupplemented group. Those with a ferritin value below 12 μg/L were defined as deficient.

ROC Analysis of Biochemical Variables Compared with Hematological Variables:

Ultimately, this paper has set out to make recommendations regarding population surveillance in aboriginal communities. Clusters of aboriginal people live in urban areas as well as areas which are remote or isolated from major centers and associated diagnostic facilities. Because it is impractical to move large, expensive, computerized hematological equipment to these communities and because alternate hematological methods are not as reliable, it is useful to assess the biochemical indicators in comparison to the hematological ones. The biochemical tests are less susceptible to deterioration with

time and transport. If comparable with the hematological ones, in spite of their higher expense, they could be well applied to the evaluation of iron stores in these isolated settings. The "maximal" hematologic definition used for the ROC analysis considered as deficient all subjects with the following values: HB < 110 g/L **or** MCV < 74 fL **or** RDW >14%. (because "or" rather than "and" was chosen in order to maximize the number of subjects which would be defined as iron deficient). Using this composite definition it was possible to compare the performance of the other tests, FEP, SF and TS identifying these iron deficient individuals. The data requirements necessary for the computer program to compare correlated ROC curves required that there be equal numbers of subjects for each test curve being compared and they each be presented in sequential order with "y" or "n" values for deficient or not. Because of this, the data were presented for analysis according to their subject ID number and were removed from consideration if there was not a complete set of determinations for each of the three biochemical tests.

In order to evaluate the numbers included in this "maximal definition", a form of "sensitivity analysis" was undertaken. The analysis was repeated five times using deficiency definitions incrementally inclusive to demonstrate the effect of higher cut-off points. Initially using the conventional "anemia" definition, "HB less than 110", then including those with both HB < 110 and MCV < 70. Next it was recalculated including any of three hematological parameters in the definitional statement; HB<110, MCV <70 and RDW >18. The fourth analysis increased the RDW cut-off to RDW > 14. Finally the "maximal" definition was employed as outlined above shifting MCV to <74.

The significance of the difference between the AUC values for SF, TS and FEP

and the line of no information were calculated in the standard method, the difference between test areas under the curve was compared using the method taking into account the correlated nature of the ROC curves. Power calculations were conducted to estimate the minimum sample size to find a difference between AUC values assuming independence of variables. Calculations were made for a range of alpha and beta errors.

Results:

Results of the Study From Which the Data Originated:

Sociodemographic Variables:

The economic situation of this population of children is perhaps best described with the statistic that (94%) 189 had parents who were unskilled laborers or were unemployed. There was a range of family structures represented in the study population (**Figure 10**). The majority involved the support of extended family, usually parents or grandparents. Eight percent of the infant's mothers were single parents; twenty five percent of mothers were living with a partner. For 35% of the study infants the mother was not living with the child.

Since educational level is a correlate of socioeconomic status, a measurement was made of the number of years of school completed. This categorical variable begins with 1 for grade one and could continue beyond 12 for post-matriculation training. (**Figure 11**) The population of infants studied in this trial came from an underprivileged segment of society. This is evidenced by the fact that only 50% (121) of their mothers had completed junior high school (grade 9) and only 7.4% (15) had completed a high school education. This is in stark contrast to the educational attainment of the general Manitoba population for whom 83% of women between age 20-29 have completed high school.[156] The mean age of the mothers in this study was 22.6 ● 5.5 for the iron-fortified group and 21.8 ± 4.8 for the regular formula group.

Results from this Thesis:

Analysis of Iron Status Tests:

Mean Test Values:

Figure 12 presents the mean values of each iron test for the total group and Figure 13 contrasts each treatment arm at each follow up visit. Two types of patterns in the data are apparent. First, there are trends which can be seen over time and second there are differences between the fortified and unfortified arms of the study.

For the HB series it is evident that the arm receiving iron fortified formula is uniformly higher than the regular formula arm. The other trend is an increase over time in hemoglobin values in both study arms. For MCV, RDW, SF and FEP there is a persistent difference between arms of the study but there is no evidence of a temporal trend. For TS there is a similar difference between study arms and a step-wise increase similar to HB over time.

It is consistent with our understanding of iron deficiency and the measurement properties of each of the six tests that the differences observed between the iron-fortified and unfortified arms are positive (infants receiving fortified formula have higher scores than those receiving unfortified) for the tests HB MCV SF TS. The differences are negative (eg. the results in the unsupplemented group are higher than the supplemented group for the tests RDW and FEP).

Horizontal lines have been drawn at the level of the cutoff value defined for use in this thesis. Until the means are presented separately for each arm of the study, the cutoff

value is not close to the mean. However, when we separate the unfortified from the fortified group, the means from the group receiving unfortified formula are sometimes beyond the cutoff value, (ie. six month mean for HB, two RDW means and all FEP means). It is worth noting that each test seems able to measure the effect of formula fortification as seen in the difference between the means for each group.

Correlation Coefficients:

Pearson correlation coefficients are presented in Table **5**; Log SF was used for the comparison. While all tests have some degree of correlation and the direction of the correlation is as expected (with RDW and FEP opposite to the others), there is a lesser degree of correlation amongst the biochemical tests than the hematological tests. There is more correlation between MCV and SF or TS than between SF and TS. Most of the test are significantly correlated ($p > 0.05$ to $p > 0.001$). Only FEP and HB or log SF do not correlate.

Mean Values Within Hemoglobin Clusters:

The five tests, MCV, RDW, FEP, SF and TS were compared with hemoglobin values at the 6 month assessment age by observing their increase in mean value with increasing hemoglobin values grouped in clusters of 10 g/dl. (This assessment age was chosen because there was the most separation between iron tests) Bar graphs of this comparison are presented in Figure **14**. The best separation between contiguous groups is seen between TS mean values. TS also has the most consistent parallel increase with

hemoglobin as HB groups increase. All tests have a distinct separation between the 80 and 90 g/L groupings, but the means of tests other than TS tend to become more equal in the higher HB groups.

The mean test value within the 100 g/dl group could be considered as a suggested cut-off value for that test. This HB group represents the cluster of individuals hovering just below the accepted anemia threshold for this age, (110 g/L). In the case of MCV it comes at 75, RDW 13.9, SF 20, FEP 0.8 and TS 12. For MCV and SF these are higher than expected.

Selection of Cut-offs:

Figure **15** displays the effect of progressively incremental cut-off values on prevalence estimates. This provides support for the selection of MCV <74 fL and FEP >0.61 μmol/L. Using these cut-off levels 30-40% of the total group is identified having ID at the 9 month assessment compared with 10 to 20% with the literature based cut-off levels. One can see that MCV <70 fL and FEP < 0.8 μmol/L likely underestimate the prevalence of ID.

Estimates of Prevalence Based on Chosen Cut-off values:

These results are presented in two figures. Figure **16** presents the various prevalence estimates as measured by different tests in four separate bar graphs, one for each follow up visit. Figure **17** displays the same information grouped by test in order to present the evolution of prevalence over time. Estimates of the prevalence of ID vary according to which test is selected and the cut-off value that is chosen. There is a wide variation in the estimated prevalence of ID at each follow up visit. For example at the

nine month visit the HB test identified 30.8% of the infants as iron deficient whereas if the MCV cutoff value was chosen to be 70 fL only 18% would have ID (when the threshold was increased to 74 fL the percentage was 55). However at the 15 month visit HB identified 12.9% as iron deficient whereas FEP identified 34.3% and SF identified 39.2%.

Estimated Sensitivity:

Table **6** introduces a sample calculation of a pair of 2 x 2 tables for MCV and HB and explains how the denominator shifts when the defining test changes. Cell "a" in this table will be the numerator, (and the number in brackets) within each of the sensitivity cells seen in the composite table of estimated sensitivity for the 9 month assessment (**7**). For example, of the 60 infants who were identified as iron deficient by the HB test, only 30 of those, (50.9%) were also identified as iron deficient using SF. Moreover of the 85 identified as iron deficient by SF only 30 (36.1%) are also judged iron deficient by HB thus leaving 45 subjects whose iron status is equivocal. HB and MCV have the lowest numbers for sensitivity although not dramatically different from the other tests.

Table **8** reports on the estimated specificity at the 6 month follow up visit. Here we find evidence of greater agreement. This makes clinical sense since there are more individuals who are not iron deficient and therefore more chance for agreement between tests. All of the tests are approximately 70-80% specific except SF which is between 60 and 70% specific for all of the tests.

ROC analysis: Sensitivity and Specificity Measurements at Different Cut-off Values:

Hemoglobin Rise as Evidence of Prior Deficiency:

Table **9** presents the data from this analysis. HB appears to be the best predictor of subsequent rise in HB with treatment based on AUC values. AUC values for TS however are almost as high. Tests of significance between AUC values were not conducted as they are relatively close for HB, MCV (9 mo), RDW and TS. The Z scores, (difference from the line of no information), are only significant for HB.

Hemoglobin Trajectory:

Table **10** presents the results of this analysis. There were 34 children identified as anemic at the 12 month visit. The denominator is slightly smaller with some of the other tests because of missing values, (which are presented in the second column). The smaller n should not have affected the calculation of AUC values or Z scores.

HB has the highest AUC (0.7941) and therefore HB at 9 months is superior to the other tests at the prediction of a low HB at 12 months. MCV has a similar value, (0.7366). The TS AUC is the next highest, 0.7346. Using the method of comparing correlated ROC curves in a three way comparison, HB is significantly different from TS or SF, (χ^2 value = 6.6430 with two tailed p of 0.0361).

By comparing individual AUC values with the line of no information, all except FEP achieve statistical significance. Although they are each significantly different from the line of no information they are not significantly different from each other.

ROC Analysis with Serum Ferritin as a Reference:

Because ferritin is thought to be the best measure of iron stores the analysis in this section considers ferritin as the "Gold Standard". ROC curves were drawn and AUC calculations made for the comparison of all non-ferritin indicators with ferritin at the same time period. The cut-off value for ferritin as a reference was any value less than 12μg/L. The two types of observations which can be seen from Table **11**, which is a summary of these calculations, are the values of each of the curve areas at one point in time and the cross-sectional trends in curve area with each successive assessment age.

The accuracy assessments of the non-ferritin iron tests as measured by their AUC are not of a great magnitude. (The AUC values, like the ones in the last two analyses are small, ie. between 0.56 and 0.72). MCV has the highest values, which is consistent with the high degree of correlation between SF and MCV as seen above with Pearson Correlation Coefficients, (Table **5**). Calculations were done for values at all 4 sampling ages to assess for any age related effect. Although average AUC values were not calculated one can see that all values are between 0.6 and 0.7 and although mostly significantly different from the line of no information they are not different from each of the other tests.

With only one exception the AUC values become smaller at each assessment age. (TS at 15 months is slightly higher). The z scores are similarly higher at the first visit except for MCV and RDW where they are slightly higher at the 9 month visit. The AUC values become less over the follow-up time of the study.

Biochemical Variables Compared with Hematological Variables:

The AUC comparisons were conducted on the curves for each biochemical test at each of the follow-up intervals including a chi squared test of significance (Table **12**). The gold standard was a maximal hematologic definition of deficiency (most inclusive). This utilizes a parallel combination of either HB <110 g/L **or** MCV<74 fL **or** RDW >14%. The AUC values are most significantly different at the 6 month follow-up assessment. At the 6 month visit the maximum separation of AUC values is seen in the ROC curves presented in Figure **19**. TS has the highest AUC 0.7764. In a 3 way comparison the χ^2 is 5.5113 with 2 df and a $P_{2\ tailed}$ 0.0636. In a 2 way comparison TS vs SF at 1df, χ^2 is 4.5231, $P_{2\ tailed}$ 0.0334.

This, however, was the only significantly different AUC value. All of the others were between 0.65 to 0.70, (TS at 15 months was slightly higher). Even the AUC value of 0.78 is not in the range of AUC values seen in other evaluations of diagnostic tests where 0.8-0.9 is more convincing. There is probably little to recommend one test over the others based on this analysis.

In order to assess the choice of the reference definition we altered the definitional statement in the gold standard, (this is a form of "sensitivity analysis"). This is best done by confining consideration to the tests taken at the 6 month assessment where the maximal separation of AUC values is seen and the maximum number of individuals are iron deficient. Table **13** displays the incremental benefit of making the definitional statement more inclusive. In the third row RDW >18%, 77 individuals are identified as deficient. In the fourth row however, establishing a more inclusive threshold for RDW

(any value greater than 14%) results in 97 individuals identified as deficient (20 more!). This translates into higher AUC values for each biochemical test and also increases the significance of their difference. Shifting the MCV threshold to 74 fL added an additional six deficient individual at the 6 month visit. The observation that using the maximum definition increases the number identified as iron deficient by a third supports the proposition that there are many individuals with pre-anemic ID who need to be identified.

Power calculations were conducted assuming variable independence in order to consider the minimum sample size to reliably measure the difference or lack of a difference between ROC curves. We were not able to factor correlation into this power calculation. Table 14 presents the comparison between TS and SF. The sample sizes required for even a 5% alpha and 20% beta are more than were present in the data available. The data however can assume test results to be correlated with a statistical benefit similar to the paired-t test. Because the correlated nature of the ROC curves enhances the power of our study, the lack of difference between some of the curves presented is unlikely to be a result of an insufficient sample size.

Summary of Results:

1. Several of the analyses within this study re-confirmed what was known from the literature and the RCT from which the data originated: that there is a significant benefit from the administration of iron-fortified compared with regular formula to non breast-fed infants.

2. An interesting finding of this study was the high prevalence of ID at the 6

month assessment when infants are still thought to have only begun depleting their maternally acquired iron stores. Recommendations for iron screening late in the first year of life may be unnecessarily late for individuals within a high risk population.

3. The selection of cut-off values for MCV at 74 fL and FEP at 0.62 μmol/L was demonstrated to be more useful than the more restrictive cut-off values suggested in the literature.

4. Biochemical tests for ID, especially SF, have been advocated as being significantly superior to a HB assessment. From the analysis of predicting HB trends there is no clear benefit to the use of biochemical tests for the detection of infants with pre-anemic deficiency. TS and FEP are as equally useful as SF. Where available, hematologic tests such as MCV and RDW may be used together with HB to increase the numbers identified as deficient. How many of these additional individuals are truly deficient could not be determined by this study.

Discussion:

Study Limitations:

The availability of secondary data permitted the comparison of iron tests. One is forced to be opportunistic with secondary data rather than constructing data collection or an intervention to answer specific questions. The chief limitation of this thesis research is the lack of an external gold standard. This resulted in some circular logic as tests were, by necessity, compared with reference to each other rather than an external reference test.

While the number of subjects recruited for this study was carefully calculated on the basis of the numbers needed to provide sufficient statistical power to demonstrate the effect of iron-fortified formula on infant development, there may not have been sufficient numbers to compare iron tests.[157] In spite of this, most studies comparing several tests have used the same number or fewer infants. Some of the studies, including NHANES II have even suggested normal values for children under age 2 based on from smaller numbers of infants. When this study used ROC methodology to assess differences in tests with reference to SF there was only a slight difference in AUC between tests. Direct power calculations of the number of subjects needed to demonstrate that no difference exists show that at least 100 more infants would have needed to be recruited. This, however, failed to include the statistical benefit of performing the calculations on paired observations.[138] The SF and other tests were determined from the same subjects and therefore some variability was removed. The statistical benefit is similar to the paired t-test which could not be directly calculated for the AUC comparisons. One can reasonably

assume then that the study numbers were likely sufficient to have discovered any difference if it was present.

Iron tests conducted at birth, two months of age or from the mothers during pregnancy would have provided additional information. It is understood that this would have adversely affected subject recruitment but would have contributed to the understanding of the maternal influence on ID and better assess the effect of iron fortification on those who were deficient from birth.

Multiple measures from the same infant over time were helpful to follow trends in iron and assess predictive validity. We were unable to assess within-person variability because each of the blood tests was taken three months apart. We were unable to assess the effect of inflammation on SF as there were no routinely measured markers for concurrent illness.

The general trend for an infant's iron status to increase with age may have affected our analysis of HB response to treatment or limited the numbers of subjects available within the HB trajectory analysis. There were, sufficient numbers to find some differences.

Suggestions for Future Research:

To more definitively answer the question about the best iron test, an external reference test would be helpful. Because of the invasiveness of most research methods, only the response to iron administration would be practical. The use of multiple biochemical markers of iron status within a remote population of aboriginal infants

(where a coulter counter is unavailable) taken at ages 6 months to two years would be helpful to compare to their response to an iron intervention. For reasons mentioned above the only hematological test available would be photometrically measured HB. Concurrent measurement of markers of inflammation, such as C-reactive protein or even ESR would assist in assessing the magnitude of this effect on SF measurement. Multiple blood tests from the same individual at one time would not be possible for reasons of consent and participation but multiple laboratory determinations on the same blood sample would allow an assessment of technical variability. It might even be helpful to perform hemoglobin electrophoresis on all, or a subgroup, to document the prevalence of hemoglobinopathies within the population.

The use of a new test, Transferrin Receptor has shown promise. It is available on a research basis and the purchase cost is not available. Like the other non-HB tests it is able to detect ID prior to anemia. It has the additional benefit of not being affected by inflammation. Future studies may employ this test for diagnostic or survey applications.[80,158-160]

Conclusions From This Study:

Effect of Fortification:

From the analysis of means and estimated prevalence, each of the iron tests studied were able to document the high prevalence of ID in this population, especially in the group receiving unfortified formula. These same tests were also able to show that the administration of iron-fortified formula provided benefit prior to 6 months even though iron stores are thought to be adequate during this period. These findings support the

recommendations from the original study and the literature which advises the administration of iron-supplementated infant formula to non-breastfed infants. (And by extension, for the administration of supplemental iron to breast feeding infants after the age of 6 months.)

The Superior Iron Test:

This study also demonstrates the difficulty in selecting an optimal measure of ID for infants. Although there is considerable overlap between groups of individuals identified as deficient by each test, this overlap is not complete. Unique individuals are identified by each test and although some of these individuals may be falsely identified as having ID, (false positives), each test has some additional merit. Even though the data do not allow the confirmation of each subject's underlying iron status it is reasonable to conclude that many of the additional individuals identified by each test do, in fact, have a deficiency of iron. Although there are liabilities associated with falsely labeling an individual as positive, in case of ID the problems associated with providing a relatively innocuous iron treatment to some individuals who may not need it outweigh the liabilities associated with leaving untreated those who may have ID but are not identified by an insufficiently sensitive test. (See the discussion below about Decision-Analysis.)

None of the tests is clearly superior. This is consistent with recent reviews which admit there is no clearly superior test.[3] In fact for population surveys, multiple tests are recommended.[87]

The ROC analysis was conducted in order to provide a method of cut-off

independent verification of test performance. The use of a parallel hematologic definition of deficiency, (any of 3 tests) was chosen to maximize the number of individuals identified with ID. Even though some may not have had serious ID, they were likely close to being iron deficient. Though there were always differences between each of the iron tests' AUC and "the line of no information" and some differences in AUC between iron tests were demonstrated, they were very modest and not sustained over the 4 assessment ages.

The comparison with HB trends was performed because in the literature this is the only available population "gold standard". As in the test comparison conducted by Margolis et al, all of the tests were about 60% sensitive.[27] The utility of four tests will be discussed individually, HB, MCV, SF and TS.

General Considerations about Hematological Tests:

Data for this thesis were based on hematologic analysis conducted using an automated electric particle counter (Coulter counter) for HB, MCV and RDW determinations. Because there are technical difficulties in the analysis of hematologic assessment transported from remote communities inexpensive portable machines called photometers can be used for capillary or venous HB determinations. These instruments are unable to reliably measure MCV or RDW and therefore the complete spectrum of hematological information is unavailable. For isolated areas then, one can use only HB, or transport the biochemical specimens.

For measurements made of subjects residing in an area with a Coulter Counter,

the composite hematological test (HB, MCV, RDW), would be the superior, economical test . The use of this multiple test policy may result in a small number of false-positive tests. Even so, treating an infant who is not iron deficient with iron is less harmful than failing to treat an iron deficient infant who was not identified.

Hemoglobin:

For this population with a negligible amount of hemaglobinopathy, hemoglobin is likely a reliable measure of iron status. There is some contribution to lower HB levels from the effect of concomitant illness on hemoglobin levels. Most of the HB decline, however is likely due to ID. This study demonstrated that HB was the best predictor of a subsequent rise in HB with iron fortification. This is consistent with the findings of others.[27,161,162]

Mean Corpuscular Volume:

The results for the MCV analysis are important because of the clinical application of a meaningful cut-off point in the detection of ID. The widely accepted cut-off value for infants at <70 fL would seem to be too low to use for the identification of deficient individuals. One possible explanation for the need for a higher cut-off for MCV within this population could be that there may be more individuals with macrocytosis, (larger red blood cells). If this was the case we would need to speculate about folate, B_{12} deficiency, or some other explanation for the macrocytosis. NHANES II, presented a mean and 95% CI of 79 fL, (67-88) for children one to two years of age while this study

showed a mean of 75.9 fL for the total group.[33] It would appear then that the poor performance of the lower cut-off is not that there are larger red blood cells within this population.

In the study of Inuit children from Keewatin, the unexpected amount of macrocytosis was thought to be due to a technical limitation of the method involving the calculation from hematocrit divided by red blood cell count.[2] For the data used in this thesis, MCV was measured directly with a coulter counter, and therefore technical problems are very unlikely to provide an explanation. One can conclude therefore that the cut-off value used for children in much of the literature may not be inclusive enough.

General Considerations about Biochemical Tests:

Serum Ferritin, (SF):

The observation that different tests seem to measure different body compartments should be of help in the choice of the best iron test. According to many authors, SF is considered the most sensitive measure of iron stores and should therefore be the best indicator of those in need of iron supplementation. If SF measurement detected all those who were anemic and those with pre-anemic deficiency it should have identified more infants as deficient than the other tests. The other tests, however, often identified more individuals as deficient than SF and there is no reason to believe that these individuals were not deficient, (false-positives).

SF may have failed to identify those who were identified by other tests as having low iron status because their values were elevated on the basis of inflammation. We know

that there is a high prevalence of upper respiratory tract infection, otitis media and skin infections in this population. It is also possible that in all of the comparisons, SF performed less well than expected because the reference tests were each limited in their ability to discriminate the children with ID from the others. One could argue that the power of a superior test is hidden when measured with reference to inferior tests. This is unlikely to be the case, however, because a series of tests were used in combination to identify the most number of deficient individuals for reference purposes, (ie. hematological definition of deficiency, using HB, MCV, RDW).

Another explanation for the poor performance of SF is that iron-fortification administered before the first blood sample was taken likely elevated the SF level to higher values before the other tests responded. Perhaps even in the non-fortified group, dietary iron from other sources could elevate the SF even before there was a sustained supply of iron necessary for hematopoesis (hemoglobin and red blood cell building). This is unlikely because the iron was administered from at least four months before the first blood sample.

Transferrin Saturation, (TS):

On the basis of the estimated prevalence, TS actually identified more individuals than SF at the 6 month assessment. It identified almost the same number as SF at 9 and 15 month assessments and only 15% less of sample at the 12 month assessment. There is no obvious explanation for this variation based on the body compartment theory since SF should always be the first to fall and we wouldn't expect a differential "inflammation

effect" at the age of 6 months.

TS has the most similarity with HB as seen in the display of hemoglobin groups, the correlation coefficients and the ROC analysis of biochemical versus hematological tests, where it was seen to be statistically superior at any cut-off value for the 6 month assessment. TS values have the benefit of possibly being less expensive than SF, and unlike hematological specimens, can be transported for analysis. Both SF and TS posses day-to-day variability although TS is known to be more variable.

TS also demonstrated the effect of iron-fortification when different arms of the RCT were compared with each other. Its estimated sensitivity and specificity were consistent between tests and over time. (The only test slightly more consistently sensitive was RDW.)

These findings are consistent with Valberg et al, who found that TS identified the largest number of children as deficient.[32] It is also similar to the study by Hershko et al, who found TS to be the superior test and found a spread of Pearson Correlation Coefficients almost identical to the ones presented in this study.[30] Others have also found TS to be the best measure of decreased iron supply.[163] Based on this assessment then, it would appear to be a superior iron test.

Cut off Values:

Decision Analysis:

It became obvious in this study how almost arbitrary "normal values" and cut-offs can be, especially with MCV and FEP. In addition to the basic statistical principles

discussed above the choice of a cut-off value is made clearer by a method called Decision Analysis.[66] Space will necessitate only a brief outline of the general principles.

The best cut-off point for transition from a negative to a positive test is a function of three quantities.[134] 1. NetCost of a mistaken intervention after a false positive test result for someone who is healthy. 2. NetBenefit, which is the gain to someone with disease from being correctly treated compared to remaining in the untreated state. And 3. Pretest Probability (p), (or prevalence). The formula which governs these three factors is:

$$SlopeofROCcurve = \frac{NetCost}{NetBenefit} x 1 - \frac{p}{p}$$

There have been no published studies using Decision Theory to assess an appropriate cut-off level for iron tests. The estimation of pretest probability, (prevalence) is straight-forward but the calculation of NetCost and NetBenefit is exceedingly difficult. Within a high prevalence population, the chance that a positive test is a true-positive is high, (high positive predictive value). Partly because of the effect of high prevalence, direct costs of testing and treating an individual with a false positive result are minimized. Also, the negative effect on those falsely identified as deficient is relatively small since the treatment is well tolerated and inexpensive. There is more problem in estimating the beneficial cost since it has not been conclusively demonstrated that ID as well as IDA is associated with psychomotor delay which may or may not be reversible. The projection of the future impact of these possibly permanent neurologic sequelae is difficult to factor

but could be high. Regardless of the specific economic details however, one could safely argue for a very inclusive cut-off so that a maximum number of children who could potentially benefit receive the innocuous iron supplementation. By extension, a strategy where all children receive iron-supplemented formula or breast-milk, followed by iron fortified foods will be discussed below.

Cut-off Independent Methods of Comparison:

The role of a ROC methodology which is a cut-off independent assessment technique was demonstrated. The curve comparison may have been more conclusive if a truly external "gold standard" were available.

Specific Cut-off Calculations:

The literature describes a range of cut-offs for FEP, (30, 35, 45 and 65 μg/dl or 0.53, 0.61, 0.80 and 1.15 μmol/L).[60,68,164,165] The data from this thesis support the use of 0.61 μmol/L for children of this age. As discussed above, the MCV cut-off for infants within this study is suggested to be 74 fL.

External Validity:

Socioeconomic Deprivation:

The results of any study are best applied to populations with similar characteristics. The objective of this study was to make recommendations regarding screening for ID within urban and remote aboriginal communities. The differences between these groups must be accounted for when considering health programs. As a

health determinant, socioeconomic conditions play a role in predicting co-morbidity and nutritional status.

A clear profile of the infants participating in this study was presented to assist in knowing to which population these results can best be applied. This group of infants has a much higher prevalence of ID than the general Canadian population. As discussed earlier, decisions about whether to mount a screening program, choice of a cut-off value of a test and which test to use are dependent upon the prevalence of the disease within the population. The social and nutritional environment of these infants is substantially more deprived than other segments of the population and multiple socioeconomic factors contribute to ill health including anemia.

While this study was conducted on aboriginal infants, its results are applicable to any population of infants of low socioeconomic status. Although the data originate from an urban-core area population of infants, those from rural or isolated areas would likely have had similar test results as long as their dietary pattern was not significantly different.

Ethnic Variation:

We have no reason to expect an ethnic difference in the performance of the iron tests with the exception that there is still debate about the normal hemoglobin levels of black children.[50] For Asian, Mediterranean, Black or other infants with a higher prevalence of hemoglobinopathies, screening with hematological tests may not be as specific as the biochemical tests and therefore these results are not directly applicable.

Groups with Similar Nutritional Profiles:

The performance of iron tests has been evaluated in a population of infants receiving artificial cow-milk formula essentially from birth. The group of study infants included those with a wide range of iron status. Half of the infants in the study received iron-fortified formula and the other half non-fortified formula. There was a difference in the prevalence of deficiency in each of these two groups but the ability of the tests to evaluate iron status in each of the treatment arms was not different as seen in both the mean values, (and the prevalence estimates). It is therefore most reasonable to assume that similar test performance can be expected amongst any infants receiving artificial formula.

Traditional aboriginal diets were high in some iron-containing foods although the iron content was often sporadic.[128] The experience of children within urban and isolated aboriginal communities is a tendency toward highly pre-prepared, iron-poor foods.[76,166] This situation is sometimes more severe in remote areas because of the more limited food choices available.[167] Because solid foods which contain iron were introduced to the infants participating in this study, the results of this thesis can be generalized to infants with a diverse range of intakes.

Because we know that breast-feeding reduces the likelihood of iron deficiency, the rate of breast-feeding in a community would affect the prevalence of iron deficiency. This population of urban aboriginal infants has a low breast-feeding rate. Although the overall Canadian rate of women breast-feeding at 6 months is less than 50%, the rate in urban and remote aboriginal communities is even lower. The feeding method of choice

for aboriginal infants for economical and convenience reasons, is evaporated milk, which, while heat denatured and pasturized, does not possess any significant amount of absorbable iron. Aside from the effect of the difference in prevalence there is no strong reason to conclude that these tests would perform much differently in infants fed evaporated milk or in breast-fed infants.

Screening Program Recommendations:

Cost Considerations:

The reader should consider the potential costs of implementing a screening program. In addition to the laboratory costs of testing, other expenses include the physical and human resources necessary to coordinate individuals to attend for testing, have blood samples drawn and report the laboratory result. Further costs include those incurred once a sub-population with ID is identified: coordination of health care visits, explanation of test results, administration of iron and retesting. When the prevalence of ID increases the numbers of individuals proceeding through this second stage also increase. With estimated prevalence in the range of 30 to 40% the economic benefits of screening become supplanted by the efficiency of an alternate program of universal iron-fortification of infant formula. The exact prevalence at which this trade off occurs is difficult to estimate but for the general Canadian population with a prevalence of less than 5% a testing strategy is not recommended.

The estimated costs of laboratory determinations are variable in Manitoba. For example the variable costs of biochemical tests, SF and TS, are estimated to be $3.50 and $7.00 respectively. (Personal communication, Ms. Elizabeth Stockl, supervisor Cadham laboratory, February 28, 1997.) Alternatively the schedule of payments to private laboratories by Manitoba Health Services Commission (MHSC) provide different estimates: $32.65 and $24.00 respectively. FEP determinations are not performed by

Cadham laboratory and MHSC does not have a payment schedule for FEP but I received an estimation that it would be in the range of the other, above mentioned, biochemical tests.

By comparison, for the hematologic assessment of HB, MCV and RDW performed on a Coulter Counter, a lab is paid $5.95 by MHSC. Fluorometric assessment of HB alone is $1.95. Therefore, a screening program using hematololgic tests would be much less expensive. For an isolated setting, one would also need to weigh the cost of a combined approach (HB plus a biochemical test) test with the perceived benefits of a multiple test strategy.

All of the above costs should be considered when a screening strategy is contrasted with one of primary prevention. These costs would include providing human resources to promote better nutrition, include breast-feeding support and the provision of iron-fortified formula or solid foods to all infants in the population. It should also include iron supplementation and nutritional counseling to pregnant women.

Recommendations:

The objectives of this thesis were twofold:

1. To address the question of which test or series of tests can best be employed in the population assessment of ID in young children.
2. To recommend a testing strategy to apply to the problem of screening within aboriginal communities.

A population assessment, or community survey is an important step toward addressing the health of infants within a community. Population health rather than individual diagnosis is a challenging application for a diagnostic test. Many reviews have advocated a battery of tests be applied to population studies in order to maximize the available information. The limitations on financial resources and practical details within urban and remote Manitoba settings dictate a more practical approach. Within many communities the prevalence of ID is unknown, though likely high. As with most of the

cited population studies, the prevalence of significant ID is likely higher than the prevalence of IDA.

In order to maximize detection of those with pre-anemic deficiency a multiple test approach is recommended. In urban settings this would likely mean using a composite hematologic profile such as provided by a coulter counter, (ie. HB <110 g/L or MCV < 74 or RDW > 14%). In more remote settings, a photometrically determined HB could be supplemented by one biochemical test. A capillary sample would suffice for both locally determined HB and a FEP sample transported to a major center. Although the biochemical tests SF and TS could be used, there are problems associated with cost, variability, and in the case of TS, the need for a venous blood sample.

The timing of this test has been conventionally directed at the 9 month infant. This may result in an unnecessary delay in the detection and treatment of ID which appears to be quite prevalent by all markers at the 6 month assessment. Perhaps this is a function of prenatal and intrapartum factors. There is likely a high prevalence of maternal ID and although this was not previously thought to contribute to infant ID,[168-170] it has increasingly been recognized as an important factor.[102,171,172] Intrapartum events such as early cord clamping at delivery may contribute to a reduction in the iron endowment of infants.[173]

An alternate program of primary prevention can be recommended. Since the estimated prevalence is so high it would be advisable that all infants receive breast-milk or iron fortified formula. The breast fed infants could be supplemented with fortified foods after the age of six months. It is also be recommended that pregnant women receive iron fortification. This would likely be more population acceptable and cost-effective than a strategy of testing and supplementing deficient individuals with iron medication and is consistent with published recommendations.[148,149,174] An evaluation of this approach using a composite testing strategy during its implementation is also recommended.

References Cited:

1. Dallman PR. Iron deficiency and the immune response. Am J Clin Nutr 1987; 46:329-34.

2. Thika R. Anemia among children and adolescents in the Keewatin Region of the Northwest Territories. 1992: University of Manitoba.

3. Beaton GH, Corey PN, Steele C. Conceptual and methodological issues regarding the epidemiology of iron deficiency and their implications for studies of the functional consequences of iron deficiency. Am J Clin Nutr 1989; 50:575-588.

4. Heinrich HC. Intestinal iron absorption in man. Methods of measurement, dose relationship, diagnostic and therapeutic applications. In: Hallberg L, Harwerth HJ, Vannotti A, editors. Iron Deficiency. London: Academic Press, 1970.

5. Lozoff B. Methodologic issues in studying behavioral effects of infant iron-deficiency anemia. Am J Clin Nutr 1989; 50:641-54.

6. Lozoff B, Brittenham G, Wolf A, McClish D, Kuhnert P, Jimenez E, et al. Iron Deficiency Anemia and Iron Therapy Effects on Infant Developmental Test Performance. Pediatrics 1987; 79:981-995.

7. Idjradinata P, Pollitt E. Reversal of developmental delays in iron-deficient anemic infants treated with iron. Lancet 1993; 341:1-4.

8. Lozoff B, Brittenham G, Viteri F, Wolf A, Urrutia J. Developmental deficits in iron-deficcient infants: effects of age and severity of iron lack. J Pediatr 1982; 101:948-52.

9. Walter T. Effect of iron-deficiency anaemia on cognitive skills in infancy and childhood. Baillieres Clin Haematol 1994; 7:815-827.

10. Oski F, Honig A, Helu B, Howanitz P. Effect of iron therapy on behavior performance in nonanemic, iron-deficient infants. Pediatrics 1983; 71:877-80.

11. England JM, Ward SM, Down MC. Microcytosis, Anisocytosis and the Red Cell Indices in Iron Deficiency. Br J Haematol 1976; 34:589-597.

12. Finch C. Diagnostic value of different methods to detect iron deficiency. In: Hallberg L, Harwerth H, Vannotti A, editors. Iron Deficiency. London: Academic Press, 1970:

13. Mahu J, Leclercq C, Suquet J. Usefulness of Red Cell Distribution Width in Association with Biological Parameters in an Epidemiological Survey of Iron Deficiency in Children. Int J Epidemiol 1990; 19:646-654.

14. Cook J. Clinical Evaluation of Iron Deficiency. Sem Hematol 1982; 19:6-18.

15. Guyatt GH, Oxman AD, Ali M, Willan A, McIlroy W, Patterson C. Laboratory diagnosis of iron-deficiency anemia: an overview [published erratum appears in J Gen Intern Med 1992 Jul-Aug;7(4):423]. J Gen Int Med 1992; 7:145-153.

16. Bothwell TH. The diagnosis of iron deficiency. N Z Med J 1966; 65(suppl):880-3.

17. Valberg LS, Sorbie J, Corbett WEN, Ludwig J. Cobalt test for the detection of iron deficiency anemia. Ann Intern Med 1972; 77:181.

18. A Dictionary of Epidemiology. 2nd ed. New York: Oxford University Press, 1988.

19. Oxford English Dictionary (OED). London: Oxford University Press, 1971.

20. Garby L, Irnell L, Werner I. Iron Deficiency in Women of Fertile Age in a Swedish Community. Acta Med Scand 1969; 185:107-111.

21. Walter T. Impact of iron deficiency on cognition in infancy and childhood. Eur J Clin Nutr 1993; 47:307-316.

22. Siimes M, Addiego J, Dallman P. Ferritin in serum: diagnosis of iron deficiency and iron overload in infants and children. Blood 1974; 43:581-90.

23. Cook J, Lipschitz D, Miles L, et al. Serum ferritin as a measure of iron stores in normal subjects. Am J Clin Nutr 1974; 27:681

24. Dallman PR, Reeves JD, Driggers DA, Lo EYT. Diagnosis of iron deficiency: the limitations of laboratory tests in predicting response to iron treatment in 1-year-old infants. J Pediatr 1981; 99:376-381.

25. Reeves J, Driggers D, Lo E, et a. Screening for iron deficiency anemia in one-year-old infants: hemoglobin and mean corpuscular volume as predictors of response to iron treatment. J Pediatr 1981; 98:894-898.

26. Burks JM, Siimes MA, Mentzer WC, Dallman PR. Iron deficiency in an Eskimo village. The value of serum ferritin in assessing iron nutrition before and after a three-month period of iron supplementation. J Pediatr 1976; 88:224-228.

27. 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; 34:2158-2168.

28. Statland BE, Winkel P, Harris SC, Burdsall MJ, Sanders AM. Evaluation of biologic sources of variation of leukocyte counts and other hematologic quantities using very precise automated analyzers. Am J Clin Path 1978; 69:48-54.

29. Phillips RS, et al. Should tests for *Chlamydia trachomatis* cervical infection be done during routine gynecologic visits? Ann Intern Med 1987; 107:188-94.

30. Hershko C, Bar-Or D, Gaziel Y, et al. Diagnosis of iron deficiency anaemia in a rural population of children. Relative usefulness of serum ferritin, red cell protoporphyrin, red cell indices, and transferrin saturation determinations. Am J Clin Nutr 1981; 34:1600-10.

31. Expert Scientific Working Group. Summary of a report on assessment of iron nutritional status of the US population. Am J Clin Nutr 1985; 42:1318-30.

32. Valberg LS, Birkett N, Haist J, Zamecnik J, Pelletier O. Evaluation of the body iron status of native Canadians. Can Med Assoc J 1979; 120:285-289.

33. Yip R. Age-related changes in laboratory values used in the diagnosis of anemia and iron deficiency. Am J Clin Nutr 1984; 39:427-436.

34. Cox CJ, Habermann TM, Payne BA, et al. Evaluation of the Coulter counter model S-Plus IV. Am J Clin Path 1985; 84:297-306.

35. Looker A, Sempos C, Liu K, Johnson C, Gunter E. Within person variance in biochemical indicators of iron status: effects on prevalence estimates. Am J Clin Nutr 1990; 52:541-7.

36. Borel M, Smith S, Beard J, Derr J. Day-to-day variation in iron status parameters in healthy men and women. Am J Clin Nutr 1991; 54:729-35.

37. Moe PJ. Hemoglobin, hematocrit and red blood cell count in "capillary" (skin-prick) blood compared to venous blood in children. Acta Paediat Scand 1970; 59:49-51.

38. Young P. Evaluation of the capillary microhematocrit as a screening test for anemia in pediatric practice. Pediatrics 1986; 78:206-209.

39. Health and Welfare Canada. Nutrition Canada: national survey. 1973; Ottawa: Information Canada.

40. Garby L. The Normal Haemoglobin Level. Br J Haematol 1970; 19:429-434.

41. Cook JD, Baynes RD, Skikne BS. Iron deficiency and the measurement of iron status. Nutr Res Rev 1992; 5:189-202.

42. Reeves JD, Yip R, Kiley VA, et.al. Iron deficiency in infants: the influence of mild antecedent infection. J Pediatr 1984; 105:874-9.

43. Koerper MA, Mentzer WC, Brecher G, Dallman PR. Developmental change in RBC: Implication in screening infants and children for iron deficiency and thalassemia trait. J Pediatr 1976; 89:580-83.

44. Dallman P, Siimes M, Stekel A. Iron deficiency in infancy and childhood. Am J Clin Nutr 1980; 33:86-118.

45. Dallman P, Siimes M. Percentile curves for hemoglobin and red cell volume in infancy and childhood. J Pediatr 1979; 94:26-31.

46. CDC criteria for anemia in children and childbearing-aged women. MMWR Morb Mortal Wkly Rep 1989; 38:400-404.

47. Saarinen V, Siimes M. Developmental changes in red blood cell counts and indices of infants after exclusion of iron deficiency by laboratory criteria and continuous iron supplementation. J Pediatr 1978; 92:412-16.

48. Perry G, Byers T, Yip R, Margen S. Iron nutrition does not account for the hemoglobin differences between blacks and whites. J Nutr 1992; 122:1417-24.

49. Owen G, Yanochik-Owen A. Should there be a different definition of anemia in black and white children? Am J Pub Health 1977; 67:865

50. Yip R, Schwartz S, Deinard AS. Hematocrit values in white, black, and American Indian children with comparable iron status. Evidence to support uniform diagnostic criteria for anemia among all races. Am J Dis Child 1984; 138:824-827.

51. McCurdy PR. microcytic hypochromic anemias. Postgrad Med 1977; 61:147-151.

52. Piedras J, Alvarez E, Herrera FM, Cordova MS. Clinical usefulness of mean corpuscular volume and red cell distribution width in iron deficient blood donors. Rev Invest Clin 1993; 45:469-472.

53. Conrad M, Crosby W. The natural history of iron deficiency induced by phlebotomy. Blood 1962; 20:173

54. Langer E, Haining R, Labbe R, et al. Erythrocyte protoporphyrin. Blood 1972; 40:112-27.

55. Lehmann F, Gray Donald K, Mongeon M, Di Tommaso S. Iron deficiency anemia in 1-year-old children of disadvantaged families in Montreal. Can Med Assoc J 1992; 146:1571-1577.

56. Saarinen UM, Siimes MA. Developmental changes in red blood cell counts and indices of infants after exclusion of iron deficiency by laboratory criteria and continuous iron supplementation. J Pediatr 1978; 92:412

57. Grant G. Prevalence of iron deficiency in rural pre-school children in Northern Ireland. Br J Gen Prac 1990; 40:112-3.

58. England JM, Down MC. Red-cell-volume distribution curves and the measurement of anisocytosis. Lancet 1974; 1:701-703.

59. Das Gupta A, Hegde C, Mistri R. Red cell distribution width as a measure of severity of iron deficiency in iron deficiency anaemia. Indian J Med Res 1994; 100:177-183.

60. Labbe RF. History and background of protoporphyrin testing. Clin Chem 1977; 23:256

61. Dagg JH, Goldberg A, Lochhead A. Value of erythrocyte protoporphyrin in the diagnosis of latent iron deficiency (sideropenia). Br J Haematol 1966; 12:326-30.

62. Piomelli S. A micromethod for free erythrocyte porphyrins: The FEP test. J Lab Clin Med 1973; 81:932-40.

63. McLaren GD, Carpenter JT, Nino HV. Erythrocyte protoporphyrin in the detection of iron deficiency. Clin Chem 1975; 21:1121-7.

64. Yip R, Schwartz S, Deinard A. Screening for iron deficiency with the erythrocyte protoporphyrin test. Pediatrics 1983; 72:214-9.

65. Center for Disease Control. Preventing Lead Poisoning in Young Children: A Statement by the Centers for Disease Control. 1978; Department of Health, Education and Welfare. -2629:

66. DeBaun M, Sox H. Setting the Optimal Erythrocyte Protoporphyrin Screening Decision Threshold for Lead Poisoning: A Decision Analytic Approach. Pediatrics 1991; 88:121-131.

67. Brock A, Michaelsen KF. Erythrocyte protoporphyrins in normal infants. Clin Chem 1990; 36:1262.

68. International Nutritional Anemia Consultative Group. Measurement of Iron Status. A report of the INACG 1985. 1985; Washington D.C. The Nutrition Foundation.

69. Cook J, Skikne B. Serum ferritin: a possible model for the assessment of nutrient stores. Am J Clin Nutr 1982; 35:1180-5.

70. Jacobs A, Miller F, Worwood M, Beamish M, Wardrop C. Ferritin in the serum of normal subjects and patients with iron deficiency and iron overload. Br Med J 1972; 4:206

71. Sheehan RG, Newton MJ, Frenkel EP. Evaluation of a packaged kit assay of serum ferritin and application to clinical diagnosis of selected anemias. Am J Clin Path 1978; 70:79-84.

72. Saarinen U, Siimes M. Serum ferritin in assessment of iron nutrition in healthy infants. Acta Paediatr Scand 1978; 67:745-51.

73. Preziosi P, Hercberg S, Galan P, Devanlay M, Cherouvrier F, Dupin H. Iron Status of a Healthy French Population: Factors Determining Biochemical Markers. Ann Nutr Metab 1994; 38:192-202.

74. Derman D, Lynch S, Bothwell T. Serum ferritin as an index of iron nutrition in rural and urban South African children. Br J Nutr 1978; 39:383-9.

75. Nelson M, White J, Rhodes C. Haemoglobin, ferritin, and iron intakes in British children aged 12 - 14 years: a preliminary investigation. Br J Nutr 1993; 70:147-155.

76. Godel JC, Pabst HF, Hodges PE, Johnson KE. Iron status and pregnancy in a northern Canadian population: relationship to diet and iron supplementation. Can J Pub Health 1992; 83:339-343.

77. Valberg LS, Sorbie J, Ludwig J, et.al. Serum ferritin and the iron status of Canadians. Can Med Assoc J 1976; 114:417

78. Thomas W, Koenig H, Lightsey AJ, et al. Free erythrocyte porphyrin: hemoglobin ratios, serum ferritin and transferrin saturation levels during treatment of infants with iron-deficiency anemia. Blood 1977; 49:455-62.

79. Wheby M. Effect of iron therapy on serum ferritin levels in iron-deficiency anemia. Blood 1980; 56:138-40.

80. Carriaga MT, Skikne BS, Finley B, Cutler B, Cook JD. Serum transferrin receptor for the detection of iron deficiency in pregnancy. Am J Clin Nutr 1991; 54:1077-81.

81. Elin RJ, Wolff SM, Finch CA. Effect of induced fever on serum iron and ferritin concentrations in man. Blood 1977; 49:147-53.

82. Ontario Association of Medical Laboratories. Guidelines for the Use of Serum Tests for Iron Deficiency. Website http://tantech.com/oaml/qairdef.html,1995;

83. Wick M, Pinggera W, Lehmann P. Ferritin in Iron Metabolism. Diagnosis of Anemias. 2nd ed. New York: Springer-Verlag, 1995.

84. Michaelsen KF, Milman N, Samuelson G. A longitudinal study of iron status in healthy Danish infants: effects of early iron status, growth velocity and dietary factors. Acta Pediatr 1995; 84:1035-44.

85. Lonnerdal B, Hernell O. Iron, zinc, copper and selenium status of breast-fed infants and infants fed trace element fortified milk-based infant formula. Acta Pediatr 1994; 83:367-73.

86. Emond AM, Hawkins N, Pennock C, Golding J. Haemoglobin and ferritin concentrations in infants at 8 months of age. Arch Dis Child 1996; 74:36-39.

87. Cook J, Finch C, Smith N. Evaluation of the Iron Status of a Population. Blood 1976; 48:449-455.

88. Deinard A, Schwartz S, Yip R. Developmental changes in serum ferritin and erythrocyte protoporphyrin in normal (nonanemic) children. Am J Clin Nutr 1983; 38:71-76.

89. Dawkins S, Cavill I, Ricketts C, et al. Variability of serum ferritin concentration in normal subjects. Clin Lab Haematol 1979; 1:41-46.

90. Lu Y, Lynch S, Cook J, Madan N, Bayer W. Use of Capillary Blood for the Evaluation of Iron Status. Am J Hematol 1987; 24:365-74.

91. Pintar J, Skikne BS, Cook JD. A Screening Test for Assessing Iron Status. Blood 1982; 59:110-113.

92. Bainton D, Finch C. The diagnosis of iron deficiency anemia. Am J Med 1964; 37:62

93. Griner P, Oranburg P. Predictive Values of Erythrocyte Indices for Tests of Iron, Folic Acid, and Vitamin B12 Deficiency. Am J Clin Path 1978; 70:748-52.

94. Owen G, Lubin A, Garry P. Hemoglobin levels according to age, race, and transferrin saturation in pre-school children of comparable socioeconomic status. J Pediatr 1973; 82:850

95. Charache S, Gittelsohn AM, Allen H, et al. Noninvasive assessment of tissue iron stores. Am J Clin Path 1987; 88:333-7.

96. Withold W, Neumayer C, Beyrau R, Heins M, Schauseil S, Rick W. Efficacy of Transferrin Determination in Human Sera in the Diagnosis of Iron Deficiency. Eur J Clin Chem Clin Biochem 1994; 32:19-25.

97. Schwartz E, Bachner R. Diurnal variation of serum iron in infants and children. Acta Pediatr Scand 1968; 57:443-5.

98. Koerper M, Dallman P. Serum iron concentration and transferrin saturation in the diagnosis of iron deficiency in children: normal developmental changes. J Pediatr 1977; 91:870-4.

99. Dallman PR. Review of iron metabolism. In: Filer LJ, Jr. editor. Dietary iron: Birth to two years. New York: Raven Press, 1989:1-17.

100. Stekel A. Iron requirements in infancy and childhood. In: Stekel A, editor. Iron nutrition in infancy and childhood. New York: Raven Press, 1984:

101. Puolakka J, Janne O, Vihko R. Evaluation by serum ferritin assay of the influence of maternal iron stores on the iron status of newborns and infants. Acta Obstet Gynecol Scand Suppl 1980; 95:53-6.

102. Kelly AM, Macdonald DJ, McDougall AN. Observations on maternal and fetal ferritin concentrations at term. Br J Obstet Gynaecol 1978; 85:338-43.

103. Kaneshige E. Serum ferritin as an assessment of iron stores and other hematologic parameters during pregnancy. Obstet Gynecol 1980; 57:238-42.

104. Smith CA, et al. Persistence and Utilization of Maternal Iron for Blood Formulation During Infancy. J Clin Investig 1955; 34:1391.

105. Finch C, Hegsted M, Kinney T, Thomas E, Rath C, Haskins D, et al. Iron metabolism. The pathophysiology of iron storage. Blood 1950; 5:983-1008.

106. Haden RL. Historical aspect of iron therapy in anemia. J A M A 1938; 111:1059

107. Schulman I. Iron Requirements in Infancy. J A M A 1961; 175:118

108. Committee on Nutrition American Academy of Pediatrics. Iron Deficiency. In: Barness LA, editor. Pediatric Nutrition Handbook. 3rd ed. Elk Grove Village, Illinois: American Academy of Pediatrics, 1993:227-36.

109. Fomon SJ, Ziegler EE, Anderson TA. Prevention of iron deficiency anemia. In: Fomon SJ, editor. Nutritional disorders of children. Rockville, MD: Department of Health, Education and Welfare, 1977:96-114.

110. Saarinen UM, Siimes MA, Dallman PR. Iron absorption in infants: high bioavailability of breast milk iron as indicated by the extrinsic tag method of iron absorption and by the concentration of serum ferritin. J Pediatr 1977; 91:36-9.

111. Duncan B, Schifman RB, Corrigan JJ, Schaefer C. Iron and the exclusively breast-fed infant from birth to six months. Journal of Pediatric Gastroenterology and Nutrition 1985; 4:421--5.

112. Siimes MA, Salmenpera L, Perheentupa J. Exclusive breastfeeding for 9 months: risk of iron deficiency. J Pediatr 1984; 104:196-9.

113. Moffatt MEK, Longstaffe S, Besant J, Dureski C. Prevention of iron deficiency and psychomotor decline in high-risk infants through use of iron-fortified infant formula: A randomized clinical trial. J Pediatr 1994; 125:527-34.

114. Marsh A, Long H, Steirwalt E. Comparative hematological response to iron fortification of a milk formula for infants. Pediatrics 1959; 24:404-12.

115. Andelman M, Sered B. Utilization of dietary iron by term infants. Am J Dis Child 1966; 111:45-55.

116. Saarinen U. Need for iron supplementation in infants on prolonged breast feeding. J Pediatr 1978; 93:177-180.

117. Miller V, Swaney S, Deinard A. Impact of the WIC program on the iron status of infants. Pediatrics 1985; 75:100-5.

118. Hertrampf E, Cayazzo M, Pizarro F, Stekel A. Bioavailability of iron in soy-based formula and its effect on iron nutriture in infancy. Pediatrics 1986; 78:640-5.

119. Saarinen UM, Siimes MA. Iron absorption from infant milk formula and the optimal level of iron supplementation. Acta Paediatr Scand 1977; 66:719-22.

120. Cook J, Lynch S. The liabilities of iron deficiency. Blood 1986; 68:803-809.

121. Fleming AF. Iron deficiency in the tropics. Clin Haematol 1982; 11:365

122. MacKay H. Arch Dis Child 1928; 3:117-147.

123. Guest G, Brown E. Erythrocytes and hemoglobin of the blood in infancy and childhood, III. Factors in variability, statistical studies. Am J Dis Child 1957; 93:486-509.

124. Katzman R, Novack A, Pearson H. Nutritional anemia in an inner-city community. J A M A 1972; 222:670-673.

125. Yip R, Binkin N, Fleshood L, et.al. Declining prevalence of anemia among low-income children in the Uited States. J A M A 1987; 258:1619-1623.

126. Haddy T, Jurkowski C, Brody H, et al. Iron deficiency with and without anemia in infants and children. Am J Dis Child 1974; 128:787-792.

127. Pollitt E. Poverty and child development: relevance of research in developing countries to the United States. Child Dev 1994; 65:283-295.

128. Waldram JB, Herring DA, Kue Young T. Aboriginal Health in Canada. Toronto: University of Toronto Press, 1995.

129. Millar J. Some observations on hemoglobin levels of an Indian Population. Can Med Assoc J 1952; 67:414-417.

130. Sackett DL, Haynes RB, Tugwell P. The interpretation of diagnostic data. In: Clinical epidemiology, a basic science for clinicians. 2nd ed. Boston/Toronto: Little, Brown, 1991:69-152.

131. Swets J, Pickett R. Anonymous Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press, 1982:

132. Brownie C, Habicht J, Cogill B. Comparing indicators of health or nutritional status. Am J Epidemiol 1986; 124:1031-1043.

133. Centor RM. Signal Detectability: The Use of ROC Curves and Their Analyses. Medical Decision Making 1991; 11:102-106.

134. Bergus GR. When Is a Test Positive? The Use of Decision Analysis to Optimize Test Interpretation. Family Medicine 1993; 25:656-60.

135. Flanagin A, Lundberg GD. Clinical Decision Making: Promoting the Jump From Theory to Practice. J A M A 1990; 263:279-80.

136. Metz C. Basic Principles of ROC Analysis. Sem Nucl Med 1978; 8:283-98.

137. Hanley J, McNeil B. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 1982; 143:29-36.

138. Delong FR, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837-45.

139. Centor RM, Schwartz JS. An Evaluation of Methods for Estimating the Area Under the Receiver Operating Characteristic (ROC) Curve. Medical Decision Making 1985; 5:149-56.

140. Rockette HE, Obuchowski NA, Gur D. Nonparametric estimation of degenerate ROC data sets used for comparison of imaging systems. Invest Radiol 1990; 25:835-7.

141. Mossman D. Resampling techniques in the analysis of non-binormal ROC data. Medical Decision Making 1995; 15:358-66.

142. Metz C, Wang P, Kronman H. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconick F, editor. Information Processing in Medical Imaging. The Hague: Martinus Nijhof, 1984:432-445.

143. Hanley JA, McNeil B. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. Radiology 1983; 148:839-843.

144. Pollack I, Hsieh R. Sampling variability of the area under the ROC-curve and of d'. Psychol Bull 1969; 71:161-73.

145. Wilson JMG, Jungner G. Definitions and Principles. In: Anonymous Principles and Practice of Screening for Disease. Geneva: World Health Organization, 1968:9-38.

146. Sheard NF. Iron deficiency and infant development. Nut Rev 1994; 52:137-140.

147. Soewondo S. The effect of iron deficiency and mental stimulation on Indonesian children's cognitive performance and development. Kobe J Med Sci 1995; 41:1-17.

148. Feightner JW. Prevention of Iron Deficiency Anemia in Infants. In: Goldbloom RB, Battista RN, Anderson G, Beaulieu MD, Elford RW, Feightner JW, et al, Canadian Task Force on the Periodic Health Exam. Ottawa: Minister of Supply and Services Canada, 1994:244-253.

149. Nutrition Committee of Canadian Paediatric Society. Meeting the iron needs of infants and young children: an update. Can Med Assoc J 1991; 144:1451-4.

150. Screening for Anemia. In: U.S. Preventive Services Task Force: Guide to Clinical Preventive Services: an Assessment of the Effectiveness of 169 Interventions. Baltimore, Md: Williams &Wilkins, 1989:163-7.

151. Iron Deficiency. In: Forbes GB, Woodruff CW, editors. Pediatric Nutrition Handbook. 2nd ed. Elk Grove Village, Illinois: American Academy of Pediatrics, 1985:213-20.

152. Hollingshead B, Reeves JD. Anonymous Social Class and Mental Illness. New York: John wiley and Sons, 1958:389-407.

153. Kim SK, Cheong WS, Jun YH, Choi JW, Son BK. Red blood cell indices and iron status according to feeding practices in infants and young children. Acta Paediatr 1996; 85:139-144.

154. Molla A, Khurshid M, Molla AM. Prevalence of iron deficiency anaemia in children of the urban slums of Karachi. JPMA J Pak Med Assoc 1992; 42:118-121.

155. Vida S. Nonparametric Receiver Operating Characteristic Analysis. Comp Meth Prog Biomed 1993; 40:95-101.

156. Statistics Canada. National Population Health Survey: 1994-1995. 1994; Ottawa: Government of Canada.

157. Arkin CF, Wachtel MS. How many patients are necessary to assess test performance? J A M A 1990; 275-8.

158. Kohgo U, Niitsu Y, Kondo H, Kato J, Tsushima N, Sasahi K, et al. Serum transferrin receptor as a new index of erythropoiesis. Blood 1987; 70:1955-61.

159. Skikne BS, Flowers CH, Cook JD. Serum transferrin receptor: a quantitative measure of tissue iron deficiency. Blood 1990; 75:1870-6.

160. Cook JD, Skikne BS, Baynes RD. Iron deficiency: the global perspective. Adv Exp Med Biol 1994; 356:219-228.

161. Reeves JD, Vichinsky E, Addiego JE, Lubin BH. Iron deficiency in health and disease. Adv Pediatr 1984; 30:281-320.

162. Graitcer P, Goldsby J, Nichaman J. Hemoglobins and hematocrits: are they equally sensitive in detecting anemias? Am J Clin Nutr 1981; 34:61-64.

163. Lipschitz D, Cook J, Finch C. A clinical evaluation of serum ferritin as an index of iron stores. N Engl J Med 1974; 290:1213-1216.

164. Raman L, Pawashe AB, Ramalakshmi BA. Iron nutritional status of preschool children. Indian J Pediatr 1992; 59:209-212.

165. Prual A, Daouda H, Develoux M, Sellin B, Galan P, Hercberg S. Consequences of Schistosoma haematobium infection on the iron status of schoolchildren in Niger. Am J Trop Med Hyg 1992; 47:291-297.

166. Moffatt ME. Current status of nutritional deficiencies in Canadian aboriginal people. Can J Physiol Pharmacol 1995; 73:754-758.

167. Lee M, Reyburn R, Carrow A. Nutritional Status of British Columbia Indians 1. Dietary Studies at Ahousat and Anaham Reserves. Can J Pub Health 1971; 62:285-296.

168. Strauss MB. Anemia of infancy from maternal iron deficiency in pregnancy. J Clin Investig 1933; 12:345-53.

169. Hussain MAM, Gaafar TH, Laulicht M, Hoffbrand AV. Relationship of maternal and cord blood serum ferritin. Arch Dis Child 1977; 55:782-4.

170. Fletcher J, Sutter PEN. The transport of iron by the human placenta. Clinical Science 1969; 36:209-20.

171. Rive S, Blacklock H, Hawkins R, Mitchell EA. Anaemia in children aged 6 to 23 months attending the paediatric acute assessment unit in south Auckland. N Z Med J 1996; 109:58

172. Garn SM, Ridella SA, Tetzold AS, Falkner F. Maternal hematological levels and pregnancy outcomes. Seminars in Perinatology 1981; 5:155-62.

173. Usher R, Shepherd M, Lind J. The blood volume of a new-born infant and placental transfusion. Acta Pediatr Scand 1963; 52:497-512.

174. Pizarro F, Yip R, Dallman PR, Olivares M, Hertrampf E, Walter T. Iron status with different infant feeding regimens: relevance to screening and prevention of iron deficiency. J Pediatr 1991; 118:687-692.

175. Dallman PR, Barr GD, Allen CM, Shinefield R. Hemoglobin concentration in white, black, and Oriental children: is there a need for separate criteria in screening for anemia? Am J Clin Nutr 1978; 31:377.

176. Habicht JP, Mayers LD, Brownie C. Indicators for identifying and counting the improperly nourished. Am J Clin Nutr 1982; 35:1241-1254.

177. Braveman PA, Tarimo E. Screening in Child Health Care. In: Screening in primary health care. Setting priorities with limited resources. Geneva: World Health Organization, 1994:86-109.

178. Dallman P. New approaches to screening iron deficiency. J Pediatr 1977; 90:678.

179. Seargeant L. Diagnostic Laboratory and Radiology Services Manual, Health Sciences Centre 820 Sherbrook Street Winnipeg, Manitoba. 18th ed. Winnipeg: HSC, 1996.

180. Bonnet JD. normocytic normochromic anemia. Postgrad Med 1977; 61:139-142.

181. Duggan MB, Steel G, Elwys G, Harbottle L, Noble C. Iron status, energy intake and nutritional status of healthy young Asian children. Arch Dis Child 1991; 66:1386-9.

182. Scott EM, Wright RC, Hanan BT. Anemia in Alaskan Eskimos. J Nutr 1955; 55:137-148.

183. Scott EM, Heller CA. Iron deficiency in Alaskan Eskimos. Am J Clin Nutr 1964; 15:282-286.

184. Cruz A, Parkinson AJ, Hall D, Bulkow L, Heyward W. Associations of early childhood infections and reduced hemoglobin levels in a historic cohort of alaska native infants. Arctic Med Res 1990; 49:175-9.

185. Maynard JE, Hammes LM. A study of growth, morbidity and mortality among Eskimo infants of western Alaska. Bull World Health 1970; 42:613-622.

186. Hoffer J, Ruedy J, Verdier P. Nutritional status of Quebec Indians. Am J Clin Nutr 1981; 34:2784-9.

187. High prevalence of iron deficiency anemia among Alaskan native children. MMWR Morb Mortal Wkly Rep 1988; 37:200-202.

188. Sellers FJ, Wood WJ, Hildes JA. The incidence of anemia in infancy and early childhood among central arctic eskimos. Can Med Assoc J 1959; 81:656-657.

189. Moffatt MEK, Young TK, O'Neil JD, Eidelheit S, Fish I, Mollins J. The Keewatin Health Assessment Study, NWT, Canada. Arctic Med Res 1993; 52:17-23.

190. Plan and operation of the second National Health and Nutrition Examination Survey, 1976-1980. Vital and health statistics. Series 1, no 232. In: Anonymous DHHS publicaiton no (PHS) 81-1317. Progress and collection procedures; no 15. US Government Printing Office. Washington DC: National Center for Health Statistics, 1981:

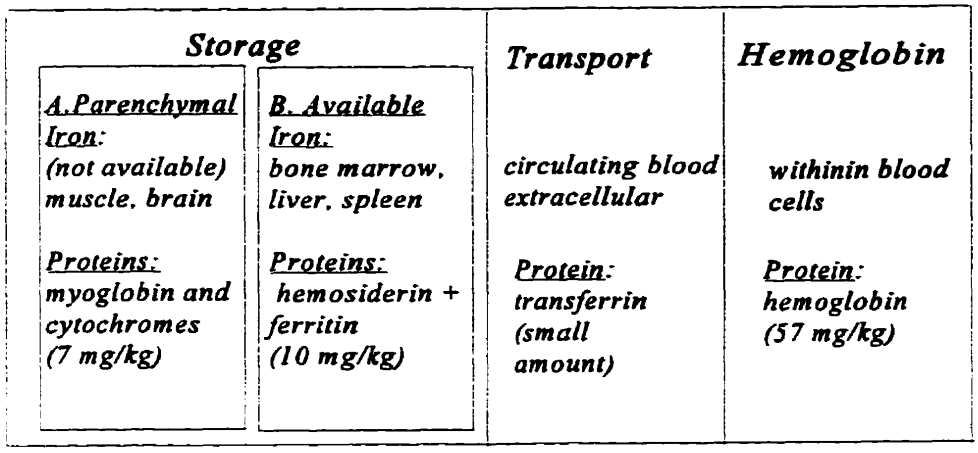


Figure 1.
Three Body Compartments of Iron.[107]

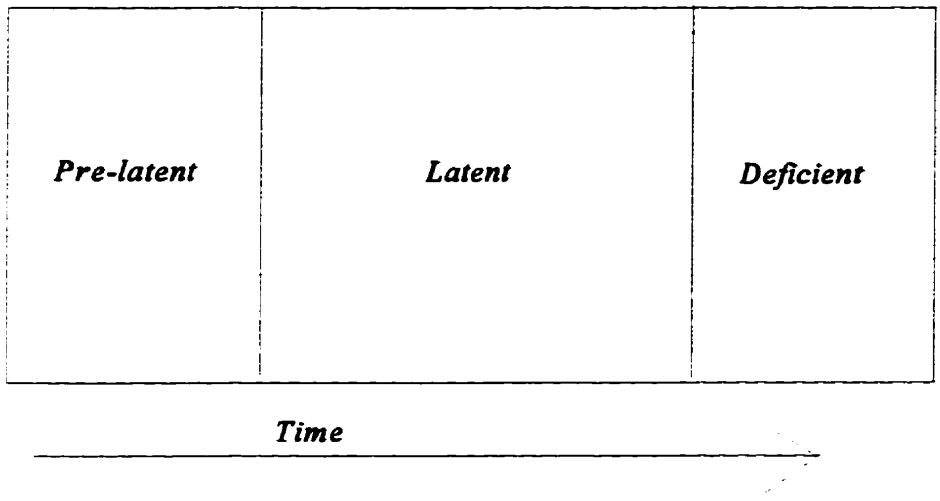


Figure 2.
Three Stages of Iron Deficiency
Adapted from Heinrich, 1970.[4]

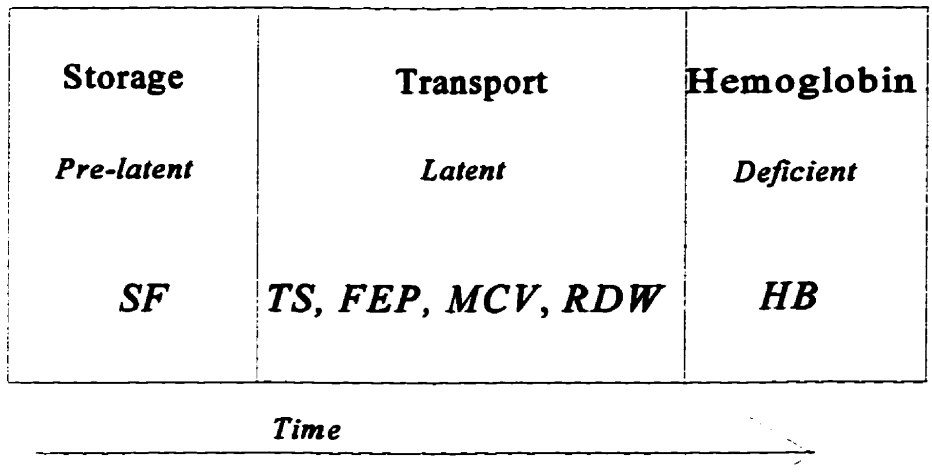


Figure 3.
Three Compartments, Stages and Tests of Iron Deficiency.[11-14]

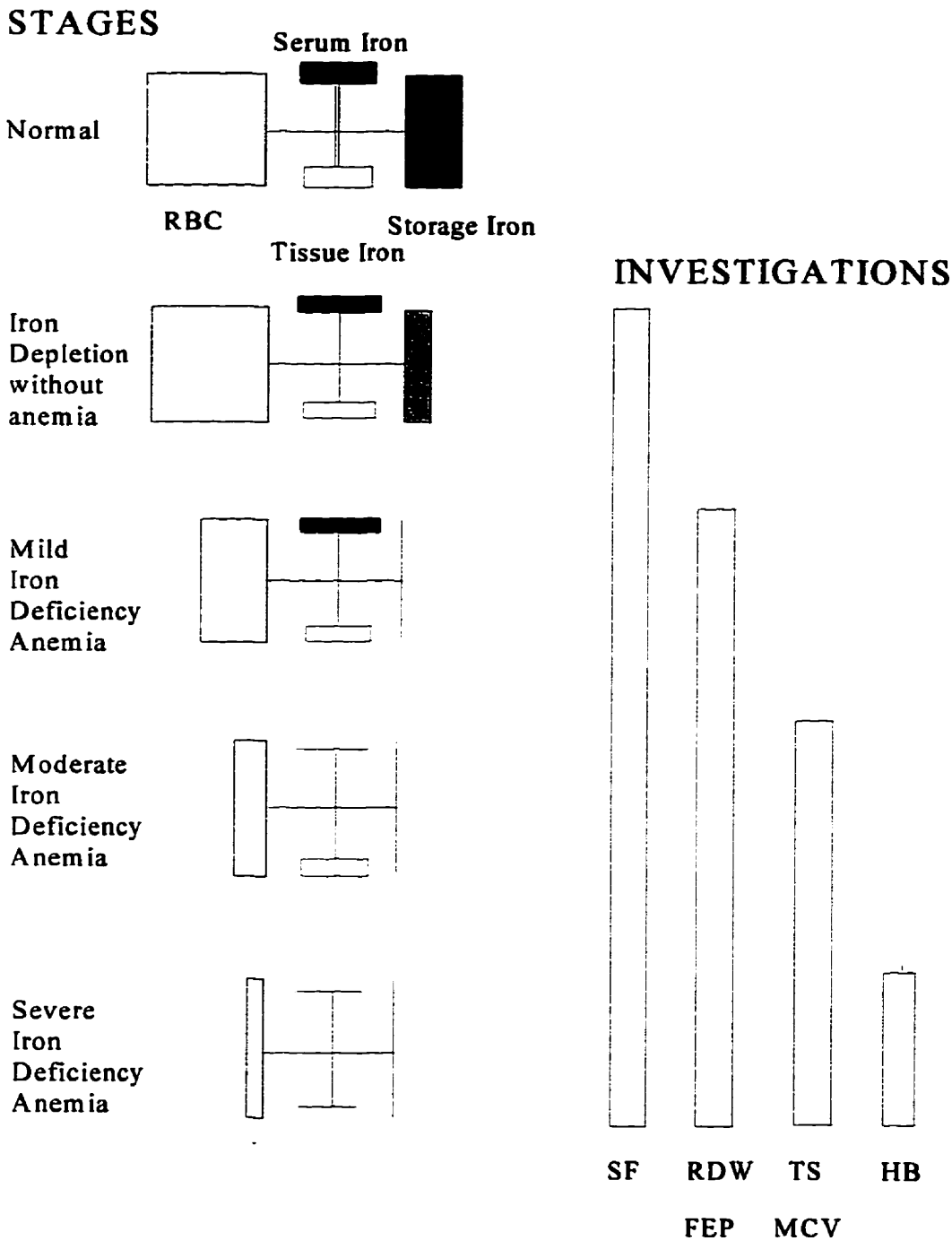
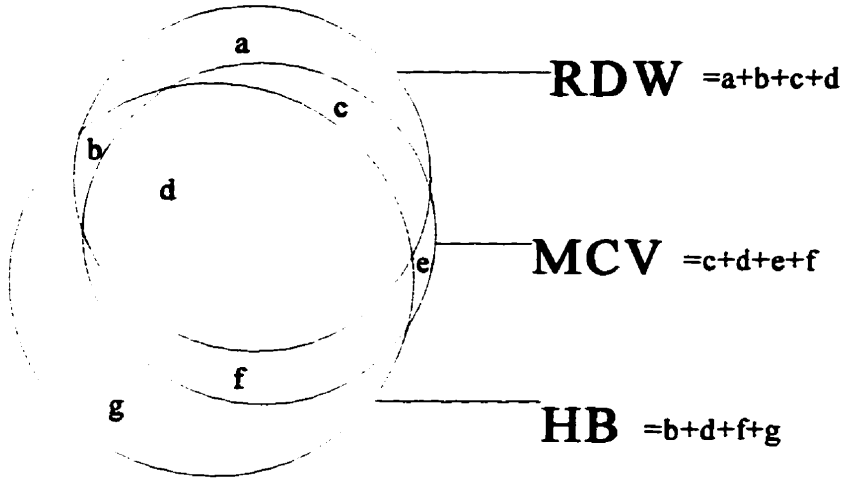


Figure 4.
Schematic representation of the relationship of iron tests to the different body iron pools, (adapted from Bothwell, 1966).[16,17]



Tests in Parallel (inclusive combination)
ie. HB **or** RDW = b+d+f+g+a+c

or "Maximal hematologic deficiency definition",
HB or MCV or RDW = a+b+c+d+e+f+g

Tests in Series (restrictive combination)
ie. HB **and** RDW = b+d

Figure 5.
Venn Diagram illustrating which groups of individuals are identified by tests used in parallel or series.

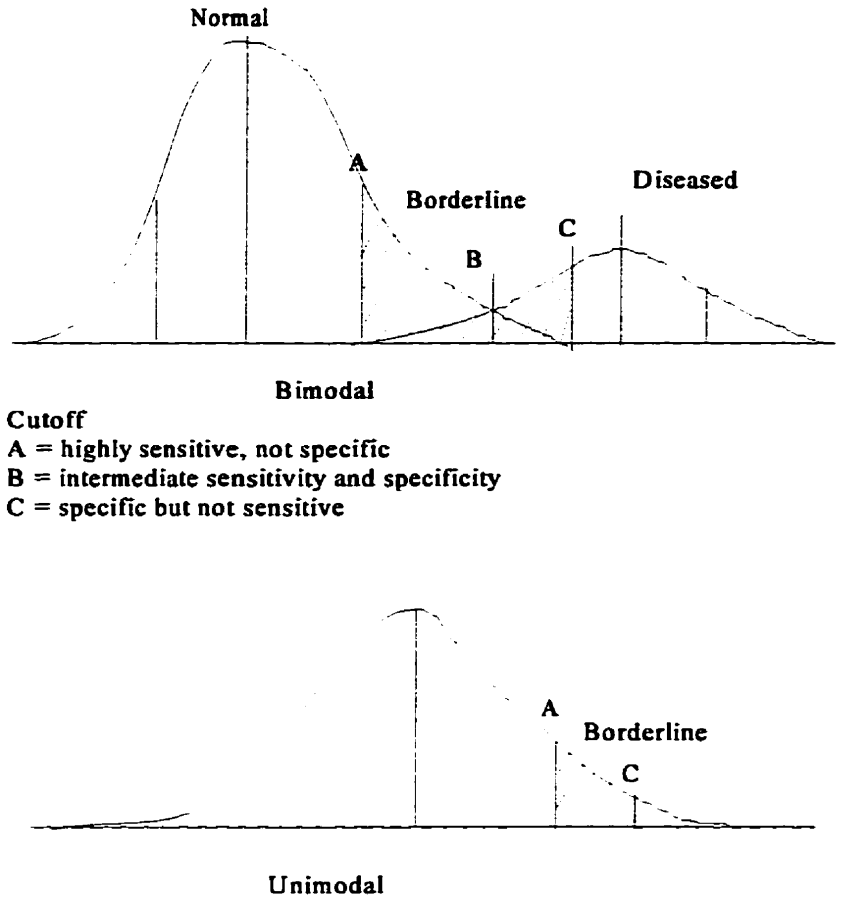


Figure 6.
Selecting a cut-off point from a distribution.
Adapted from Wilson and Jungner, 1968.[145]

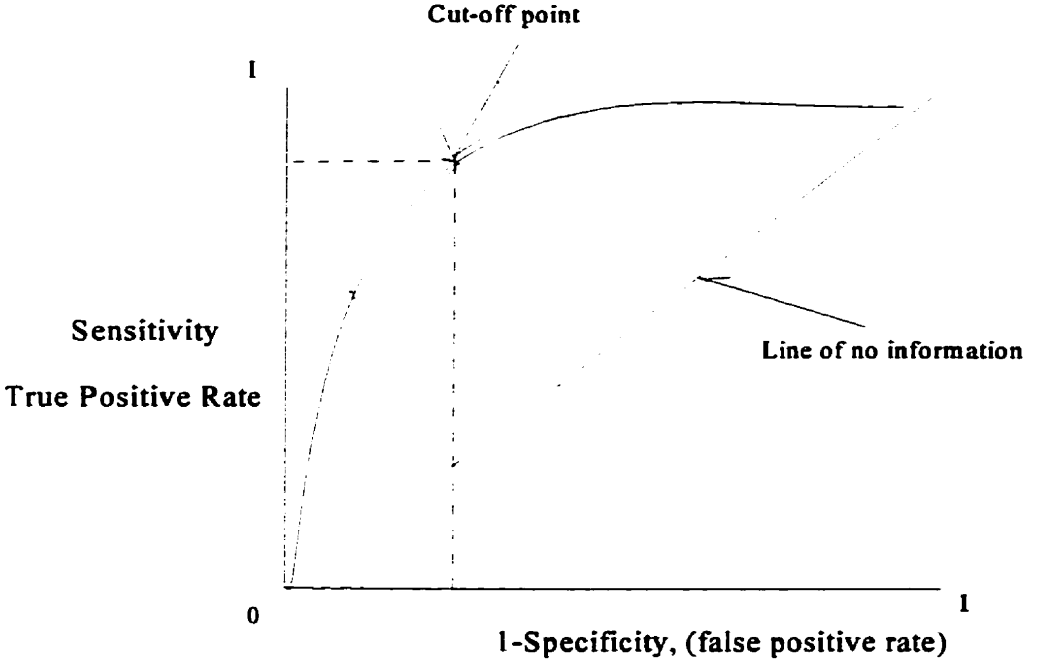


Figure 7.
Sample Receiver Operator Characteristic Curve, (ROC).
Adapted from Sackett, 1991.[130]

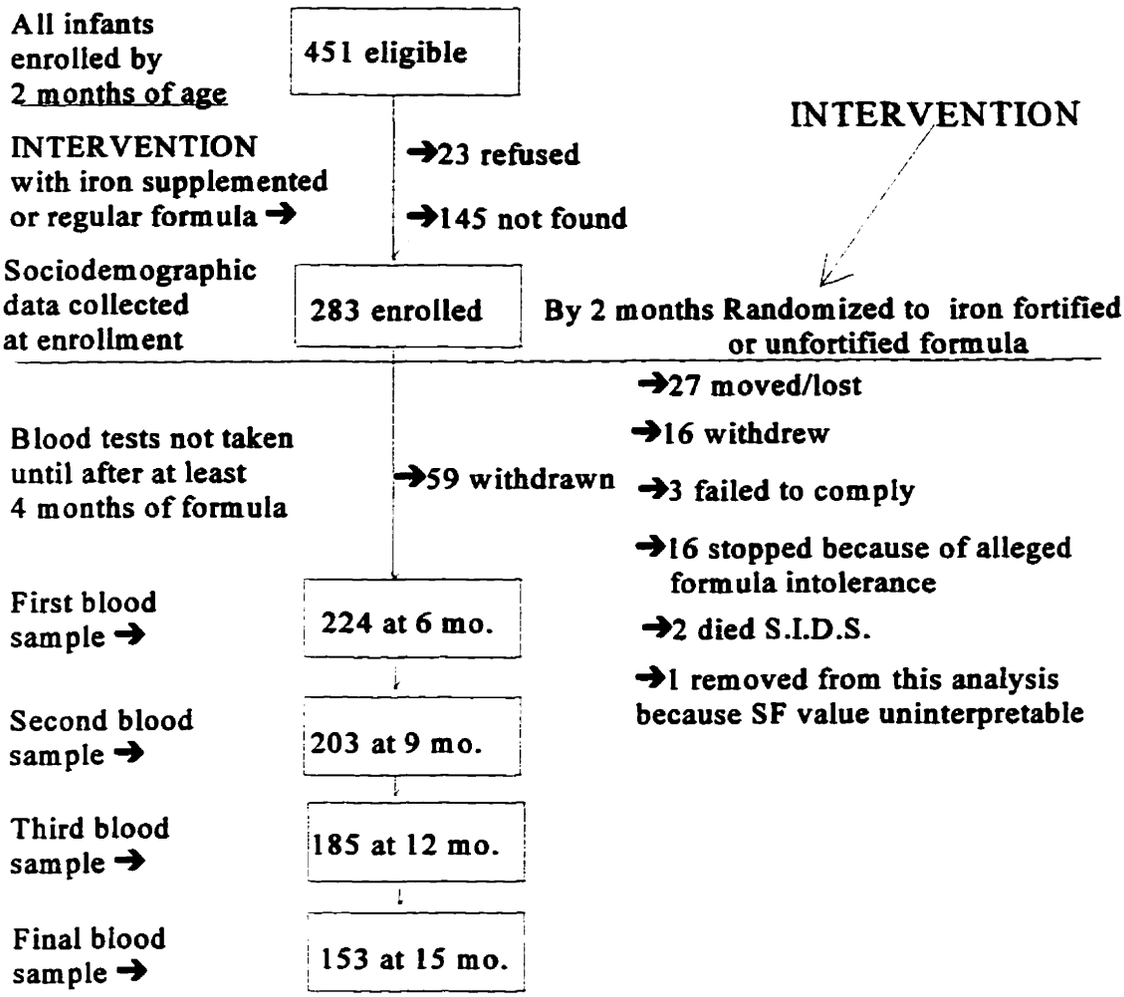


Figure 8.
Flow Chart of Data Collection
Adapted from Randomized Controlled Trial, Moffatt 1994.
(Progress of time measured from top to bottom of figure)[113]

Iron Test	Assessment Time			
	6 month	9 month	12 month	15 month
	HB1	**HB2**	HB3	HB4
	MCV1	MCV2	MCV3	MCV4
	RDW1	RDW2	RDW3	RDW4
	FEP1	FEP2	FEP3	FEP4
	SF1	SF2	SF3	SF4
	TS1	TS2	TS3	TS4

1. Italic — Cross sectional analysis at the 6 month assessment

2. Underline — Examination of trends in the group's iron status over time

3. Bold — Relationship between an individual's test at the 6 month assessment and another iron test(s) at a subsequent assessment

4. Any of the above analysis performed separately according to treatment group

Figure 9
Data structure perspectives.

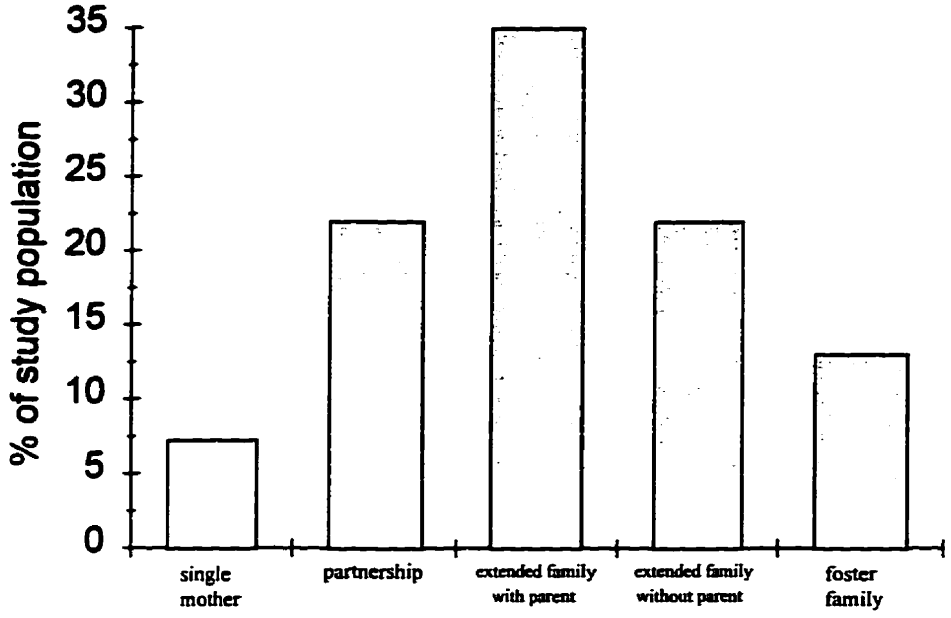


Figure 10.
Family status, from total sample.

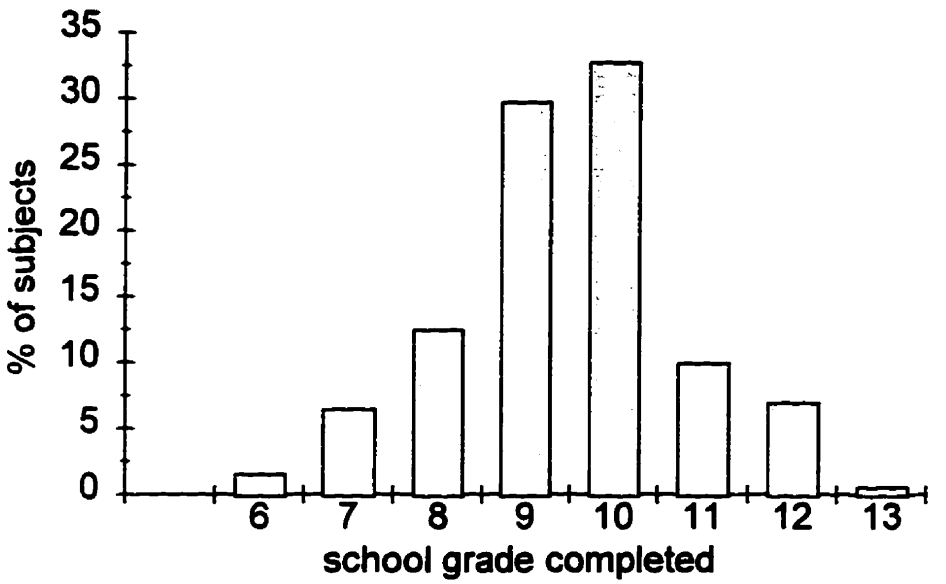


Figure 11.
Mothers' educational level, from total sample..

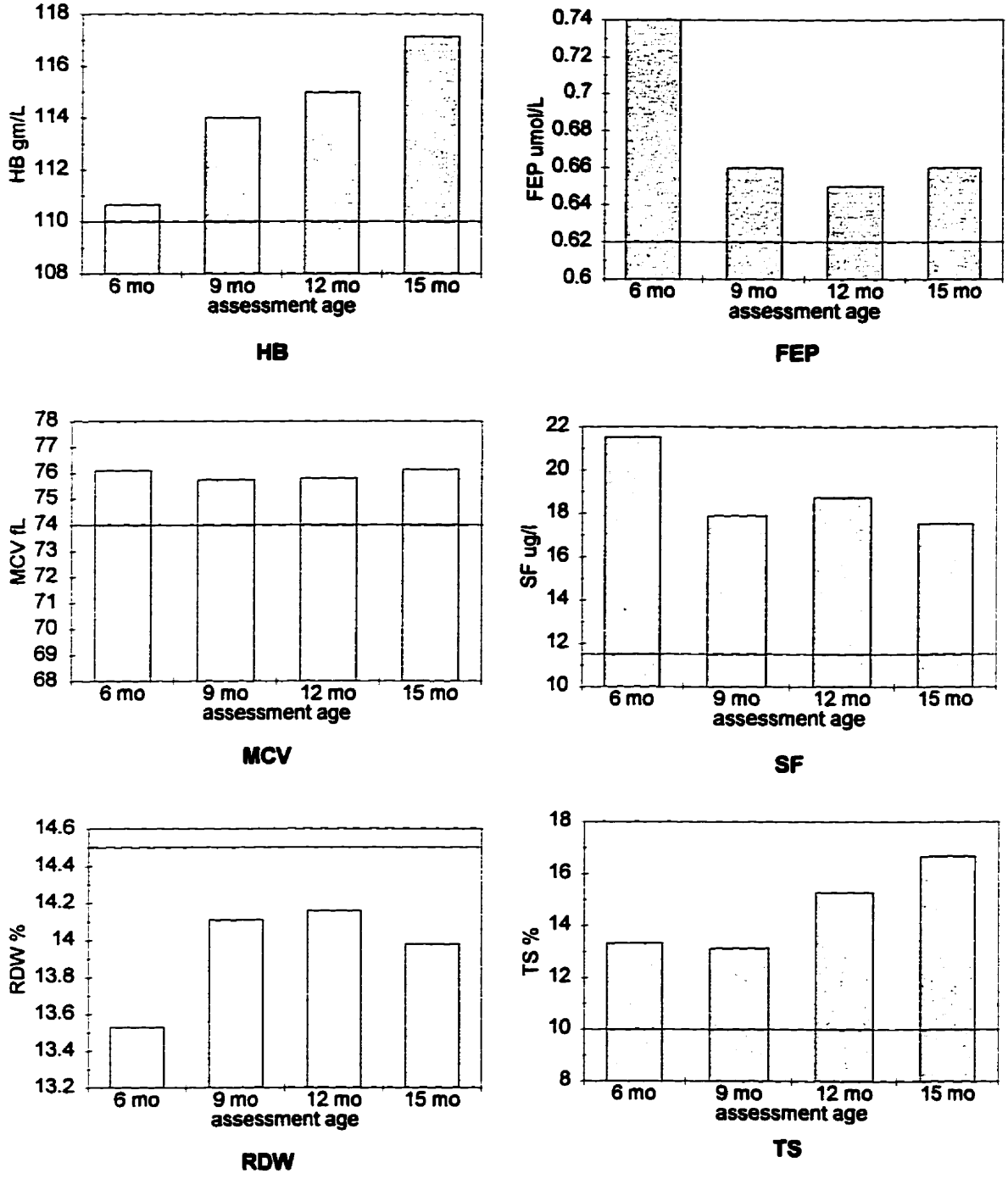


Figure **12.**
Mean test values for total group.
Horizontal line = chosen "cut-off" value used for subsequent analysis

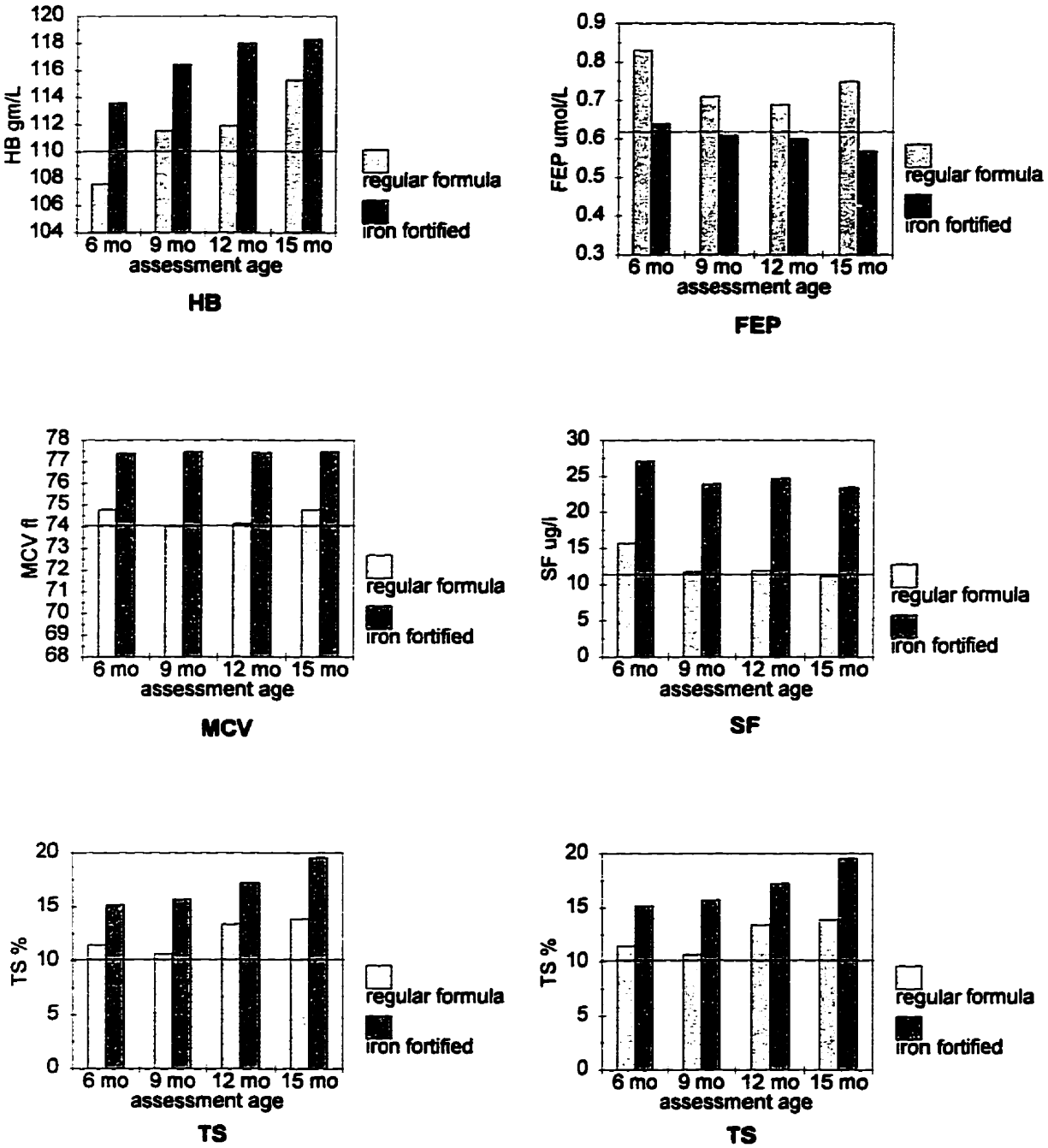


Figure 13.
Mean test values by treatment group.
horizontal line = "cut-off" value chosen from literature
for subsequent analysis

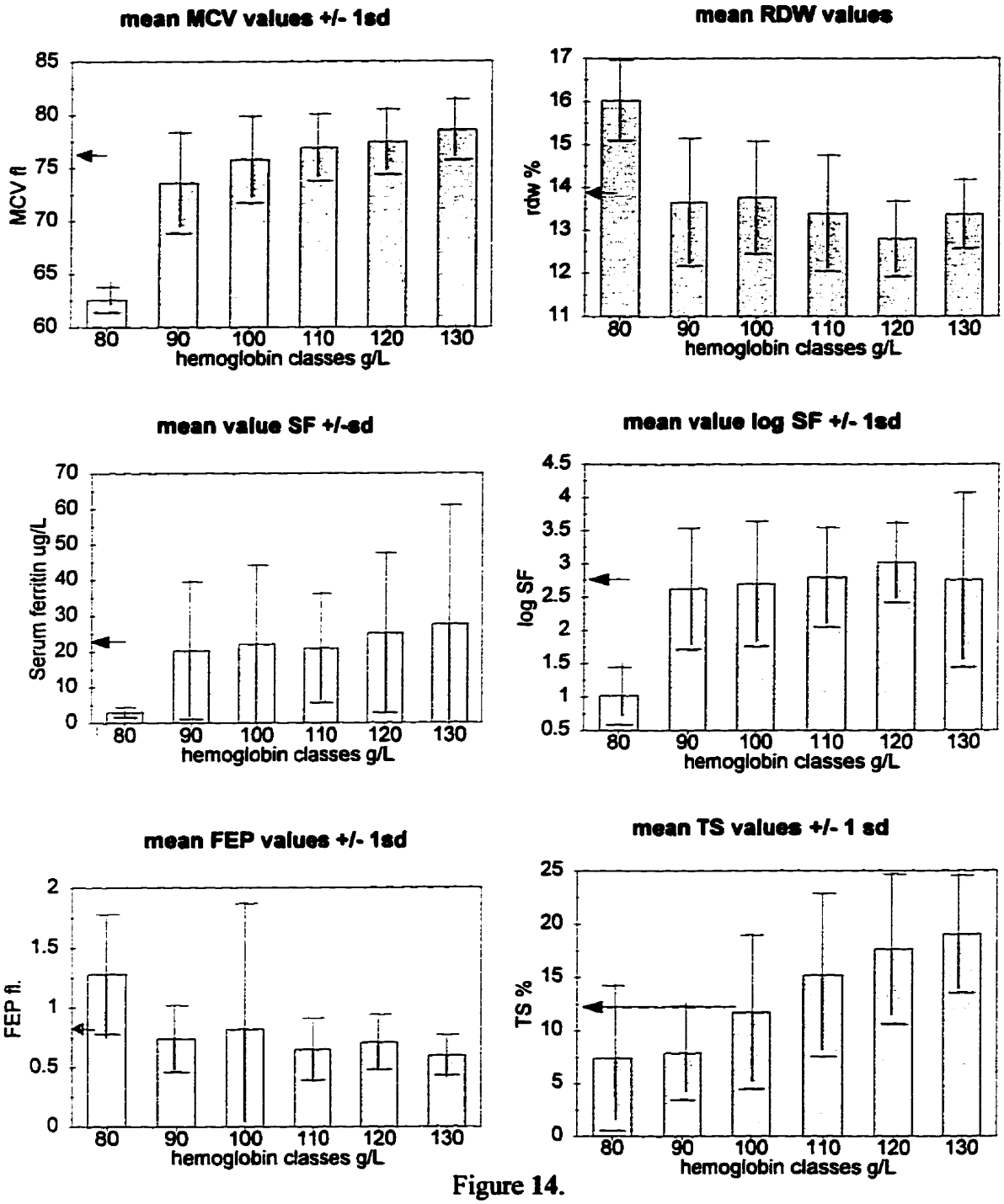


Figure 14.

Mean test scores ± 1 SD within HB clusters,
(understanding tests in relation to HB)
*80 = <90, 90 = <100, 100 = <110, 110 = <120, 120 = <130 and 130 = >130
horizontal arrow indicates "cut-off" point suggested from this analysis
(from top of HB = 110 bar)

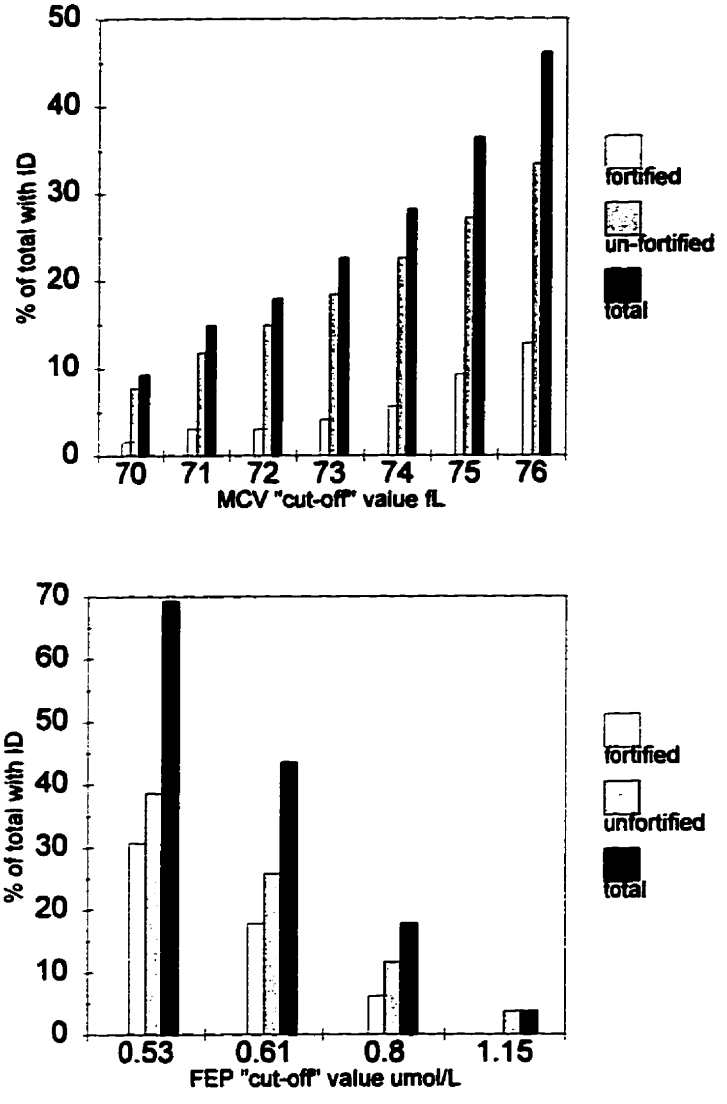


Figure 15.
Effect on prevalence estimates of varying the "cut-off points for MCV and FEP displayed for total and by treatment groups for the 9 month assessment age.

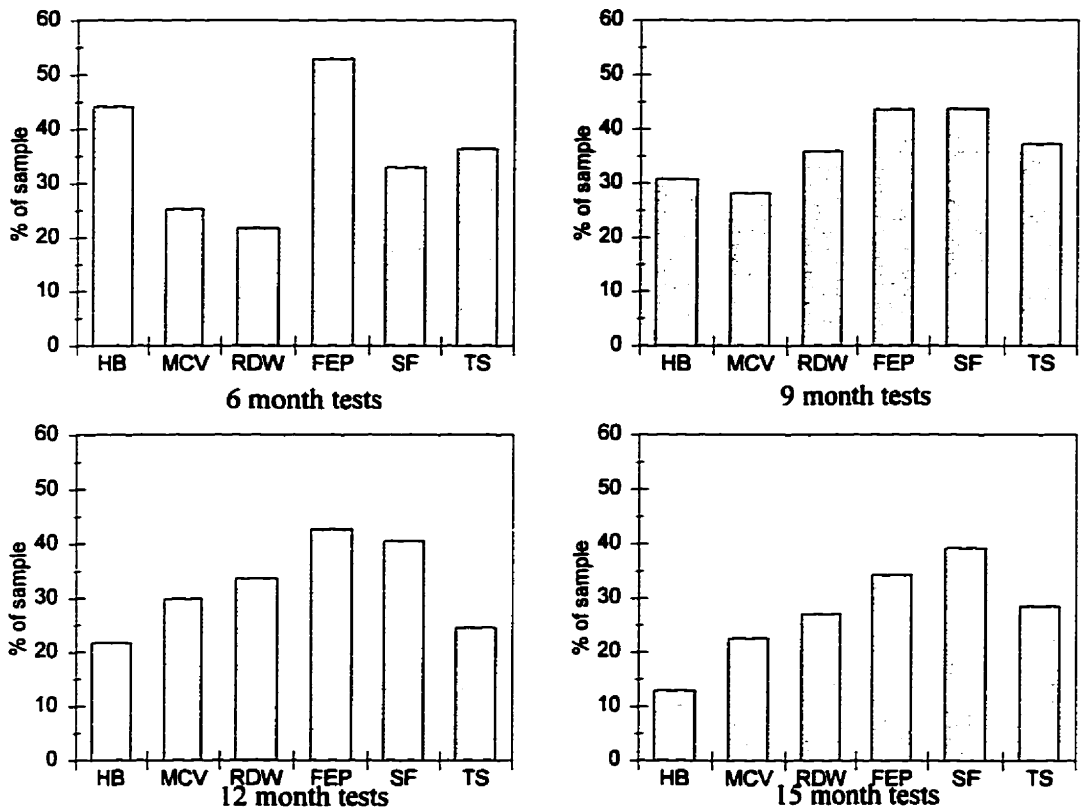


Figure 16.
Estimated prevalence of ID as determined by each iron test
according to published "cut-off" values,
grouped by assessment age.
HB <110 g/L, MCV <74 fL, RDW > 14.5 %,
FEP > 0.62 µmol/L, SF < 12 µg/L, TS < 10%

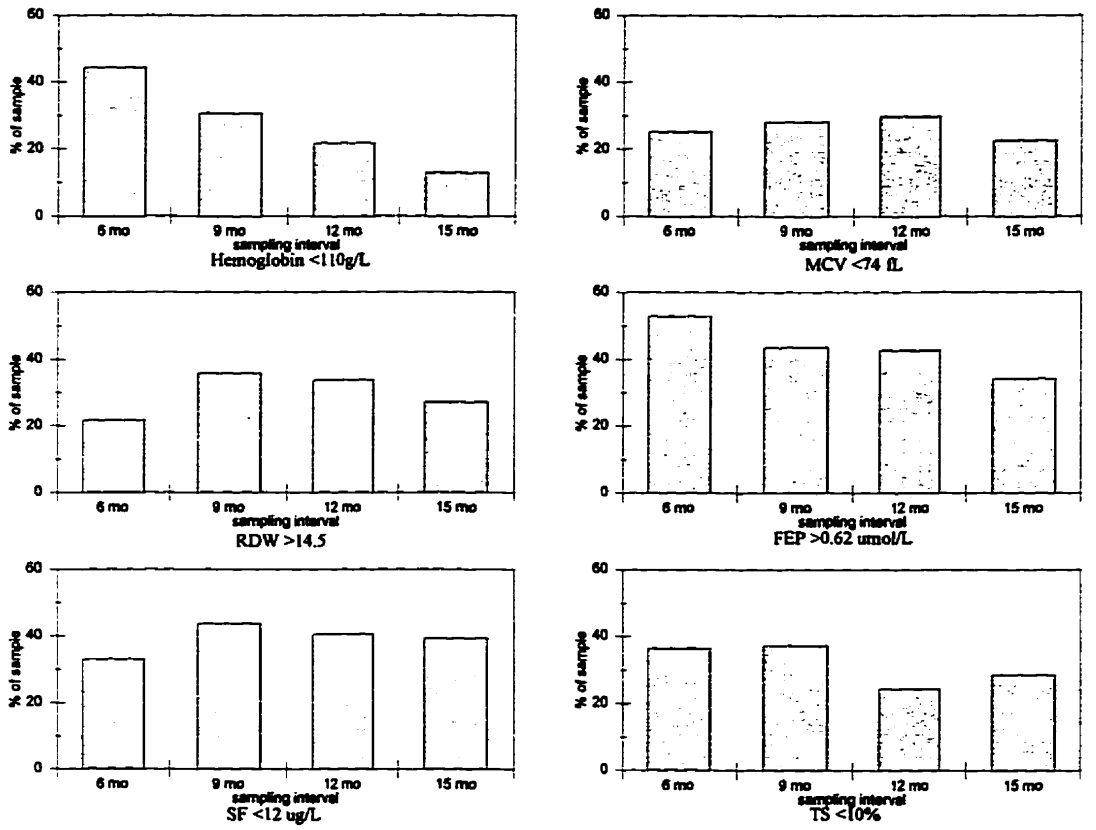


Figure 17.
Estimated prevalence of ID as determined by each iron test
according to published "cut-off" values,
grouped by test.
HB <110 g/L, MCV <74 fL, RDW > 14.5 %,
FEP > 0.62 µmol/L, SF < 12 µg/L, TS < 10%

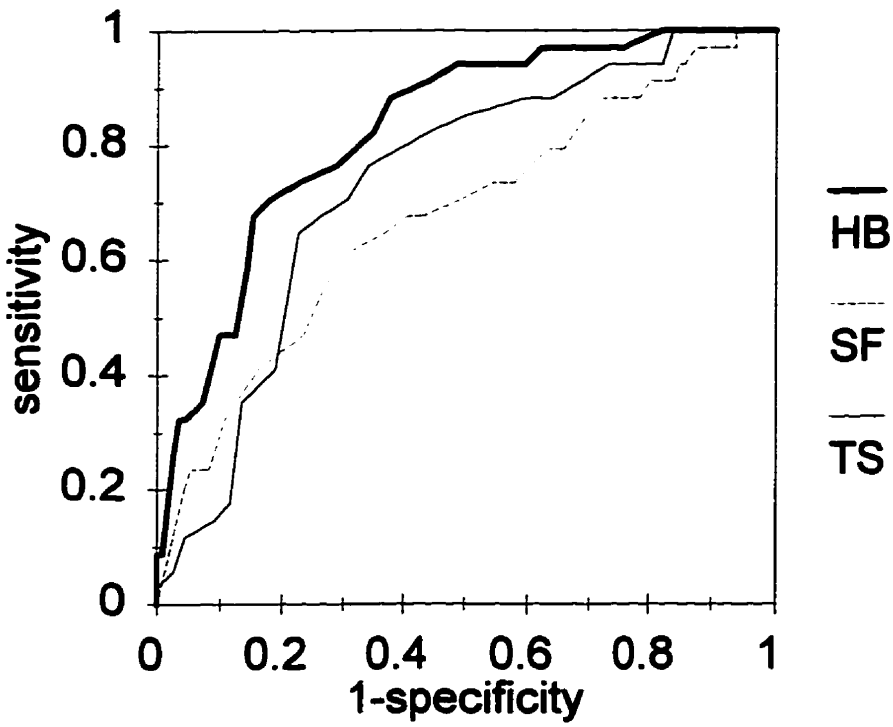


Figure 18.
Receiver Operator Curve, (ROC), of 9 month tests' capacity to predict HB < 110 g/L at 12 months: HB, SF and TS.
(Analysis conducted on entire group.)

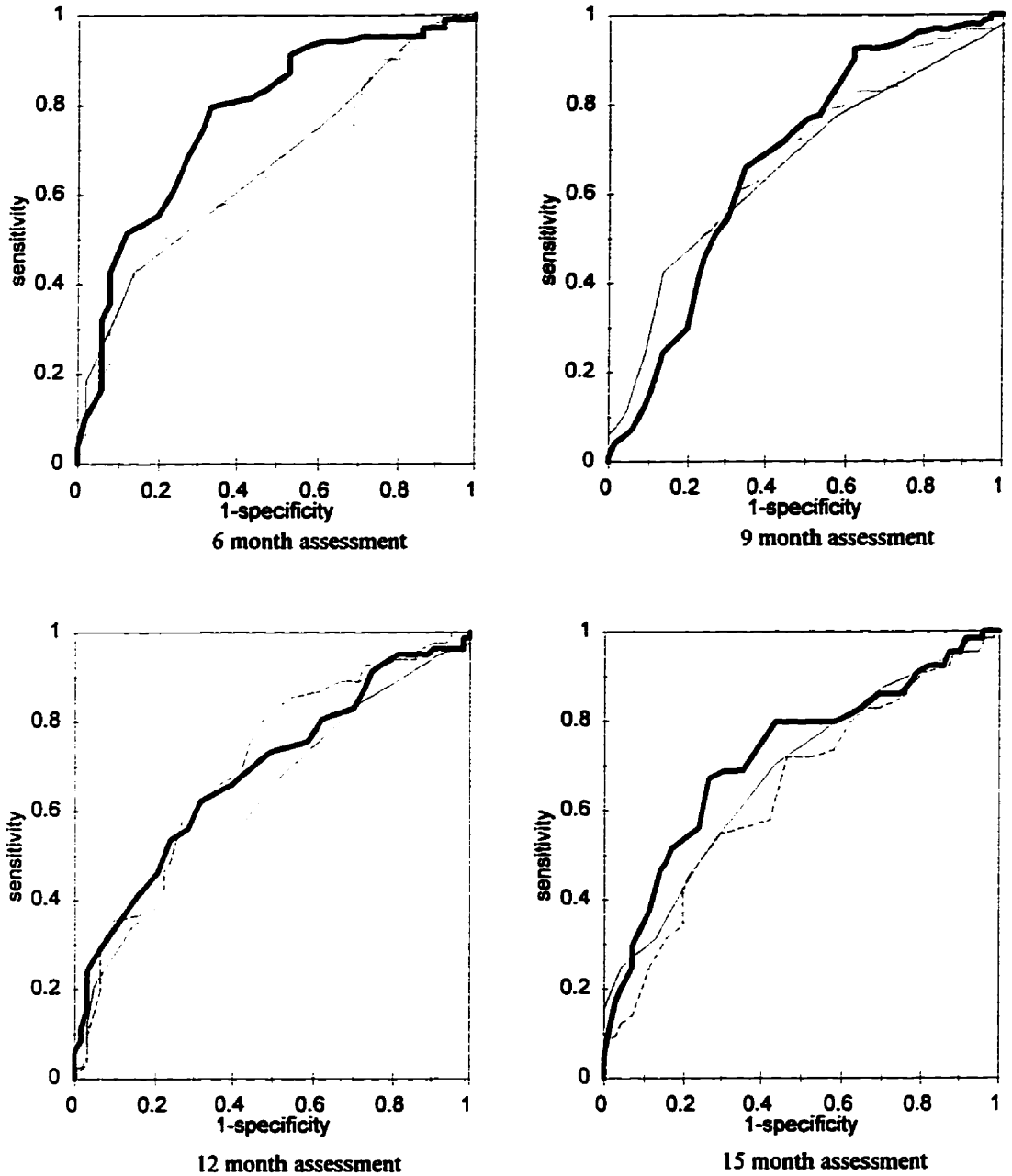


Figure 19
Receiver Operator Curve, (ROC) of biochemical tests with "maximal" hematologic reference (HB <110 g/L, MCV < 74 fL and RDW > 14%) at all four assessment ages

	HB	MCV	RDW	FEP	SF	TS
Biology	-Principle O_2 Carrying Protein in red blood cells	-cells decrease in size with ID[51]	variation in cell size with latent ID	-HB precursor -surplus FEP in ID	-storage iron 1 µg/L SF = 10 mg[69]	-only iron transport protein
Method	-1.Electronic particle counter (coulter) -2. Photometric (portable)	-1 Electronic counter - 2. Indirect MCV=HCT/RBC (centrifuged tubes + optically read RBC)	- volume distribution curve of the RBC through a range of microapertures	-spectrofluorometer -micromethod of Piomelli[62]	-radioimmunoassay 1972[70]	-automated biochemical analyzer =Serum Iron + Total Iron Binding Capacity (TIBC) x 100
Advantages	-inexpensive -has been the defacto standard -neurodevelopmental delay only correlated with low HB	-earlier than HB -available from multichannel electronic analyzer with HB	-earlier than HB -available from multichannel electronic analyzer with HB - population with β Thalassemia[59]	-capillary sample -reliable -reproducable -relatively inexpensive -established in pop screening of lead	-reflect iron stores throughout a wide range of iron status -trends away from the direction of deficiency with inflammation/ infection[42]	-well studied -transports wellron status and ages[22] -responds immediately[78,79]
Limitations	-HB decline is a late finding in ID -prevalence of hemoglobinopathies limit specificity -normal values change with age -ethnic differences in black infants[48,49,175] -(no aboriginal diff)[50]	-elevated in presence of B_{12} or Folate, therefore not specific -normal values change with age	-correlates inversely with body iron status[13] -slow to reflect developing ID[41]	-elevated in anemia of chronic illness and infections [42,54] -elevated in hemolytic anemia, sideroblastic anemia, protoporphyria[64] -elevated in	-marked elevations (>50µg/L with inflammation)[12,31,176]	-TIBC elevated in fever[81] -variableless accurate in pregnancy [80] -elevated in fever [81] -some doubt its accuracy in first months of life[84] lead poisoning[63]

	HB	MCV	RDW	FEP	SF	TS
Published estimates of sensitivity / specificity	-usually this is the reference test	-50% detection of blood donors with low TS [52] -low compared to phlebotomy [53] (too slow?)[54] -90 and 53.8[153]	-83.3, (specif=57.7)[153]	-4,160 children at cutoff 35 µg/dl compared with SF <15 µg/L = 88% (incl 10% aboriginal)	-validated in hospitalized patients[23,71] -more sens than MCV in NHANES II [31] -ROC AUC	=0.95 compared to BM
Recommended as Primary Screening by	[25,150,177]	-main screening test[51,178]	-Mahu[13] -benefit over SF[95]	-popular screening test [64] (especially when it used to also be used to screen for Lead [65]	-meta-analysis of 1,179 articles (55 with Bone Marrow comparison)[15] -Ontario Assc Med Lab[82] -superior to TS[26,32,83,133]	-superior to SF[30,95].
Stage of Deficiency	-last	-middle (latent)	-preanemic	-pre-anemic [54]	-iron stores, earliest measure	-preanemic
First Described	-last century	-early this century	-1974[11,58]	-role in heme synthesis 1930[60] -accurate testing 1970[61]	1972[23,70,71]	mid century

	HB	MCV	RDW	FEP	SF	TS
Variability	-insignificant day to day or within day[28,36] -diurnal variation 3-5% of morning values[3] -less with coulter counter[28,34] -with photometer CV=1.3 - 5.28[2,27] (2.5 g/L)	-Direct, quite repeatable -Indirect (unacceptable optically read RBC CV =19[2]	-as reliable as coulter counter	-micromethod correlation (r=0.98) with classical method[62] -lease amount of within-person-variance of all tests (NHANES)	-unacceptable day-to-day variation[89] -3-10 independen	-diurnal var. 20%[3] -no diurnal var. prior to age 3 years[97] -unacceptably high variance[35,84] -3-10 measurements req. To accurately determine values[36] -Capillary var. only slightly higher than venous[35,90,91]
Limitations	-HB, only late in ID -hemoglobinopathies limit specificity -normal values change with age -ethnic differences in black infants ?[48,49,175], not aboriginals [50]	-elevated in presence of B_{12} or Folate, therefore not specific -normal values change with age -cost and size of multichannel coulter -too variable with other methods	-correlates inversely with body iron status [13] -slow to reflect developing ID [41]	-1 in chronic illness and infections [42,54] -also hemolytic anemia, sidero-blastic anem, protoporphyria[64]	-marked elevations (>50µg/L with inflammation)[12,31,176]	-TIBC elevated in fever [81] -less accurate in pregnancy [80] -some doubt its accuracy in first months of life[84]
Cost	~$3.00	Comes with HB	comes with HB	$3.00-$15.00	$4.00 - $32.00	$7.00-17.00
units	-gram per Liter (g/L) -some report as number of SD from age appropriate mean value (to allow comparison across ages)[27,33]	-femtoliters (fL) = 10^{-15} Liters	%	-30 µg /dl of blood -0.8 µmol/L blood (all ages)[179] -0.62 µmol/L blood -75 µ/dl rbc -3µg/g HB[27,68]	-µg /L	%

	HB	MCV	RDW	FEP	SF	TS
Population Studies				-4,039 Mpls. Children -26 Danish infants[67]	-many studies positive[32,57,72,73,75-77,87,18]	some studies where it failed to identify a significant number with anemia[27,30,85]
Cut-off Values 6 months to 2 years	- NHANES II 110-130 g/L [33,45-47] -97 g/L[86]	-NHANES II, 79 (67-88) -(10th %ile),72 fL [55] -75 fL[57] -70 fL[24,44,45]		-highest at 1-2 yrs -mean +2SD 65 µg/dl (highest at this age) [88]	-meta-analysis of 55 adult studies, bone marrow comparison, adults with inflammation <50 µg/l[15] -range of acceptable cut offs [87] -8 months of age 16 µg/l[64,86] -age dependent -2SD value 12-21 µg/l [88] -most widely accepted, INACG	-359 children with deficient subjects excluded <7%[98] -adults (INACG) <16%[68] -NHANES II age 1-2 years 10%ile <12% [31,33] -<10% [13] <12 µg/l [57,68] - <10 µg/l[55]
Distribution	-gaussian unless high prevalence ID[39] -no clear separation of normal and abnormals[20,40]	-gaussian unless high prevalence ID[39]	-normal	-normal	-log-normal[86]	normal

Table 1.
Summary table comparing iron tests

Authors	Location	Subjects	n	Test	Prevalence
Mackay '28[122]	London	poor children		HB	"near universal"
Guest 1930s	Cincinnati	poor children	1735	HB	30%
Guest 1950s[123]	Cincinnati	poor children			30%
Katzman '68[124]	New Haven	ethnically diverse 10-36 month		HB	12.5%
Yip '75[125]	US	low income		HB	7.8%
Yip '85	US	low income		HB	2.9%
Dallman et al. '81[24]	San Francisco	children of US army		HB	25%
Haddy '74[126]	Michigan	low income		TS<16 + HB<110	50%
Lehmann et al '92[55]	Montreal	low income	220	SF<10 and HB <115 or MCV <72	25% (SF <10 in 37%)
Derman et al '78[74]	South Africa			HB (57%) SF	70%
Grant, '90[57]	Ireland	rural preschool		SF <12	17%
Duggan '91[181]	Sheffield	Asian babies		SF <10	35%

Table 2.
Summary of prevalence studies of infants with low socioeconomic status .

Authors	Location	Subjects	n	Test	Prevalence
Scott et al '55[182]	Arctic	Adult Inuit	715	HB	mean HB <white contrl
Scott and Heller '64[183]	Arctic	men and women	422	HB	24%M 26%F
Cruz '90[184]		infants 6-23 mo (1960)		HB <110	40-50%
Maynard & Hammes '60-'62[185]		Children from 43 Inuit villag.	643	HB	low HB compared to controls
Valberg '79[32]	Yukon and NWT	children 1-4		TS <16 SF <10	47% 21%
Hoffer '78[186]	James Bay Cree	adults		HCT and TS	lower than the rest of Canada
Margolis et al '81[27]	Bristol Bay SW Alaska	6 mo-17 yrs Inuit	344	HB↑ 5 g/L HB↑ 10 g/L SF<10 TS, FEP	69% 43% 45% 35%, 70%
MMRW '88[187]	Yukon-Kuskokwim Delta	children	83 352 age 6-11	SF <10 HB <115	65% 4.3%
Sellers et al '58[188]	Keewatin, Cen Arctic	children	331	HB <100	40%
Moffatt et al '90-91[189]	Keewatin			HB	high prevalence
Millar '51[129]		infants 6-24 mo		HB	73-85%

Table 3.
Summary of prevalence studies of aboriginal infants.

age †	HB	‡	MCV		RDW		FEP		SF		TS	
6.3 (5.5-8.1)	217	110	217	110	211	107	170	84	203	104	184	93
		107		107		104		86		99		91
9.5 (8.6-12.6)	195	98	195	98	192	96	163	80	194	98	172	84
		97		97		96		83		96		88
12.7 (11.7-14.7)	174	88	174	88	172	88	150	75	173	87	155	76
		86		86		84		75		86		79
15.8 (14.6-20.5)	155	78	155	78	155	78	143	70	153	79	144	71
		77		77		77		73		74		73

† Mean age in months (range). Text of thesis will approximate these to 6, 9, 12, and 15 months.
‡ Number who received iron-fortified formula over the number who received regular formula.

Table 4.
Numbers of lab values available for analysis (total and by treatment group) at each assessment age.

	HB	MCV	RDW	FEP	SFL	TS
HB	1	0.41018	-0.47401	-0.19204	0.21177	0.3791
"p' value	0	0	0	0.0141	0.0032	0
n	195	195	192	163	192	172
MCV	0.41018	1	-0.62236	-0.49391	0.46964	0.42619
"p' value	0	0	0	0	0	0
n	195	195	192	163	192	172
RDW	-0.47401	-0.62236	1	0.54268	-0.41718	-0.40667
"p" value	0	0	0	0	0	0
n	192	192	192	161	189	170
FEP	-0.19204	-0.49391	-0.41718	-0.32324	1	0.21023
"p" value	0.0141	0	0	0	0	0.0056
n	163	163	161	163	163	162
SFL*	0.21177	0.46964	-0.41718	-0.32324	1	0.21023
"p" value	0.0032	0	0	0	0	0.0056
n	192	192	189	163	194	172
TS	0.3791	0.42619	-0.40667	-0.3225	0.21023	1
"p" value	0	0	0	0	0.0056	0
n	172	172	170	162	172	172

*Log of Serum Ferritin

Table 5.
Pearson correlation coefficients for iron test values at the
9 month assessment

	ID present	ID absent	
Test +	a	b	a +b
Test -	c	d	c+d
	a+c	b+d	total a+b+c+d

Sensitivity[1] = a/a+c
Specificity[2] = d/b+d

	MCV <74 (ID)	MCV≥74 (no ID)	
HB <110	29	31	60
HB ≥110	26	109	135
(29 missing)	55	140	195

sensitivity = 29/55 = 52.7%
specificity = 109/140 = 77.9%

	HB <110 (ID)	HB ≥110 (no ID)	
MCV <70	29	26	55
MCV≥70	31	109	140
(7 missing)	60	135	195

sensitivity = 29/60 = 48.3%
specificity = 109/135 = 80.7%

Table 6.
Outline of estimated sensitivity and specificity calculations.
Example: HB and MCV at the 9 month assessment

0*	HB	MCV	RDW	FEP	SF	TS
# with ID→	60	55	69	71	85	64
HB		52.7% (29)	53.6% (37)	40.9% (29)	36.1% (30)	50.0% (32)
MCV	48.3% (29)		55.1% (38)	52.1% (37)	42.2% (35)	56.3% (36)
RDW	61.7% (37)	69.1% (38)		58.6% (41)	54.9% (45)	59.4% (38)
FEP	55.8% (29)	75.5% (37)	65.1% (41)		52.1% (38)	60.7% (37)
SF	50.9% (30)	63.6% (35)	65.2% (45)	53.5% (38)		54.7% (35)
TS	58.2% (32)	70.6 (36)	57.6% (38)	52.1% (37)	44.9% (35)	

% = sensitivity from 2 x 2 table
( ) = number of infants identified as deficient by both criteria

* thresholds used: HB <110 g/L, MCV<74 fL, RDW >14.5%, FEP >0.62 µmol/L, SF<12 µg/L, TS<10%

Top row designates the "gold standard" for comparison
Second row lists number identified with ID using that "gold standard"

Table 7.
Estimated sensitivity table
from the nine month assessment.

0	HB	MCV	RDW	FEP	SF	TS
# without ID⇒	135	140	123	92	109	108
HB*		77.9% (109)	81.3% (100)	75% (69)	73.4% (80)	78.7% (85)
MCV	80.7% (109)		86.2% (106)	87.0% (80)	81.7% (89)	86.1% (93)
RDW	75.8% (100)	77.4% (106)		75.8% (69)	77.6% (83)	73.6% (78)
FEP	62.2% (69)	70.2% (80)	70.4% (69)		63.3% (57)	66.3% (67)
SF	60.2% (80)	64.9% (89)	69.2% (83)	62.0% (57)		60.2% (65)
TS	72.7% (85)	76.9% (93)	75.0% (78)	73.6% (67)	69.2% (65)	

% = specificity from 2 x 2 table
( ) = number of infants identified as deficient by both criteria

* thresholds used: HB <110 g/L, MCV<74 fL, RDW >14.5%, FEP >0.62 µmol/L, SF<12 µg/L, TS<10%.

Top row designates the "gold standard" for comparison
Second row lists number identified with ID using that "gold standard"

Table 8.
Estimated specificity table
from the nine month assessment.

	n	ID +	ID-	AUC	SE of AUC	95% CI	Z[§]	P2
HB 6	110	36	74	0.6845*	0.0517	0.583I-0.7859	3.1354	0.0017
HB 9	98	25	73	0.6345	0.0661	0.5051-0.7640	2.0028	0.0452
MCV 6	110	36	74	0.4747	0.0588	0.3593-0.5900	0.4321	0.6656
MCV 9	98	25	73	0.5416	0.0653	0.4136-0.6697	0.6229	0.5334
RDW 6	107	36	71	0.5704	0.0593	0.4541-0.6867	1.2185	0.223
RDW 9	96	25	71	0.5389	0.0639	0.4136-0.6641	0.5954	0.5516
FEP 6	84	28	56	0.5003	0.0715	0.3601-0.6405	0.0048	0.9962
FEP 9	80	22	58	0.5141	0.0718	0.3734-0.6548	0.1971	0.8438
SF 6	104	35	69	0.4126	0.0602	0.2946-0.5306	1.4526	0.1463
SF 9	98	25	73	0.4118	0.0682	0.2782-0.5454	1.313	0.1892
TS 6	93	32	61	0.5914	0.0645	0.4650-0.7179	1.4451	0.1484
TS 9	84	22	62	0.5799	0.0702	0.4424-0.7175	1.111	0.2666

*Largest AUC, (TS next largest)

§ Z score is the difference between the AUC and the line of "no information"

Table 9.
Results of ROC analysis: AUC of Iron Tests at 6 and 9 months.
Ability to predict a rise in HB of 5 g/L at the next assessment interval
in response to the iron-fortification of formula

	n	ID +	ID -	AUC	SE of AUC	95% CI	Z	P2
HB	170	38	132	0.8123*	0.0423	0.7294-0.8952	5.8637	0
MCV	170	38	132	0.7635	0.0456	0.6742-0.8527	4.9586	0
RDW	167	38	129	0.7028	0.0476	0.6095-0.7961	3.875	0
FEP	141	32	109	0.6319	0.0568	0.5206-0.7432	2.2949	0.0217
SF	169	38	131	0.7056	0.0498	0.6081-0.8031	3.8577	0
TS	150	35	115	0.7472	0.0446	0.6598	0.8346	0

*AUC of HB at 9 months significantly different from TS or SF, chi square = 6.6430 2df p_2=0.0361

§ Z score is the difference between the AUC and the line of "no information"

† "anemia" defined as HB value at 12 months less than 110 g/L

‡ 5 children treated with extra iron included in this analysis

Table 10.
Iron tests at 9 month visit; Ability to "predict" anemia at 12 months.†‡

test, mo.	# ID+	#ID -	total	AUC	se	Z[5]	p_2
HB 6	66	135	201	0.6899	0.0408	4.3741	0
HB 9	83	109	192	0.6232	0.0404	2.9238	0.004
HB12	70	102	172	0.5614	0.0456	1.3686	0.1711
HB 15	59	91	150	0.658	0.0449	3.268	0.001
MCV 6	66	135	201	0.7167	0.0394	5.0053	0
MCV 9	83	109	192	0.7149	0.0379	5.1127	0
MCV 12	70	102	172	0.6263	0.0432	2.8241	0.005
MCV 15	59	91	150	0.6739	0.0437	3.6047	0
RDW 6	64	131	195	0.6449	0.0423	3.363	0
RDW 9	82	107	189	0.6765	0.0403	4.2458	0
RDW 12	69	101	170	0.5915	0.0448	2.0696	0.0385
RDW 15	59	91	150	0.6162	0.0469	2.4636	0.0138
FEP 6	59	108	167	0.6877	0.0431	4.0471	0
FEP 9	73	90	163	0.6016	0.0449	2.2575	0.024
FEP 12	58	92	150	0.5956	0.048	1.9901	0.0466
FEP 15	59	83	142	0.622	0.0479	2.5072	0.0122
TS 6	60	121	181	0.6456	0.0418	3.1888	0.001
TS 9	78	94	172	0.5996	0.0432	2.2487	0.0245
TS 12	60	95	155	0.5936	0.0472	1.962	0.0498
TS 15	57	86	143	0.6382	0.0466	2.7963	0.005

[5]Z statistic: the measurement of the difference between the AUC and the null hypothesis AUC=0.5, "line of no information"

Table 11.
ROC calculations of AUC for iron tests with ferritin below 12μg/L as a "gold standard".

	FEP		SF		TS		3 way comparison	
age mo.	AUC	SE	AUC	SE	AUC	SE	χ^2 2df	2 tail p
6	0.6712	0.0436	0.6567	0.0454	0.7764†	0.04	5.5113	0.064
9	0.6623	0.0425	0.7009	0.0407	0.6765	0.045	0.4985	0.7794
12	0.6241	0.0458	0.7047	0.0438	0.688	0.044	1.9059	0.3856
15	0.684	0.0451	0.6437	0.0478	0.7208	0.045	1.7585	0.4151

†two way comparison TS vs. FEP $\chi^2 = 4.1285$ 1df 2 tailed p = **0.0422**
TS vs. SF $\chi^2 = 4.5231$ 1 df 2 tailed p = **0.0334**

Table **12.**
Summary of areas under the ROC curve.
Three biochemical tests with reference to a maximal hematological definition of ID:
HB < 110 g/L, MCV < 74 fL and RDW > 14 %

"definitional statement"[§]	number deficient	fep auc and se	sf auc and se	ts auc and se	chi sq df=2	2 tail p_value
HB <110	75/158	0.5801 0.0450	0.5962 0.0468	0.7238 0.0405	8.2931	0.0158
HB<110 MCV<70	77/158	0.5996 0.0447	0.6039 0.0463	0.7181 0.0405	6.002	0.0497
HB<110 MCV<70 RDW>18	77/154	0.5955 0.0453	0.6100 0.0465	0.7213 0.0408	6.3269	0.0423
HB<110 MCV<70 RDW>14	97/154	0.6521 0.0446	0.6113 00468	0.7468 0.0409	6.3547	0.0417
HB < 110 MCV < 74 RDW>14	103/154	0.6712 0.0436	0.6567 0.0454	0.7764 0.0404	5.5113	0.0636

[§]"definitional statement" using a combination of reference standards with differing cutoff points

Table 13.
Three way comparison of correlated ROC curves of biochemical tests
with reference to a hematologic definition of deficiency.
Effect of varying the hematologic definitional statement of deficiency.
All calculations on the 6 month assessment lab results

2-tailed alpha	Stastical Power	Beta	Normal subjects	Abnormal subjects
5%	80%	20%	154	154
5%	90%	10%	204	204
5%	95%	5%	250	250
2.5%	80%	20%	188	188
2.5%	90%	10%	242	242
2.5%	95%	5%	292	292
1%	80%	20%	231	231
1%	90%	10%	291	291
1%	95%	5%	346	346

Deficiency defined by the most inclusive hematologic definition of deficiency
HB <110 gm/L, MCV <74 fL and RDW >14%.

Table 14.
Power comparison of AUC difference between SF and TS at 6 months ability to diagnose ID compared with a maximal hematologic definition (AUC = 0.6567 and 0.7764 respectively).

Appendix 1: Coulter Counter methodology:

These multichannel analyzers combine several different types of tests within a specific machine. The hemoglobin is determined by the cyanmethemoglobin method. The addition of cyanide results in a compound which has a wider absorption spectrum at 540 nm and the hemoglobin measurement can be reliably made by a photometric method. The determination of Mean Corpuscular Volume (MCV) is determined directly using the capacity of the machine to individually count the cells of different sizes with a 256 channel pulse height discriminator as they pass through 3 separate microapertures from 36 to 360 fL (femtoliters= 10^{-15} liters). In the same way the red blood cells are counted directly to produce the Red Blood Cell Count (RBC). The internal computer then generates a red cell histogram, (frequency of different sizes of cells), from which the Red Blood Cell Distribution Width (RDW) is calculated (coefficient of variation of the central portion of the histogram). The Hematocrit is then calculated by multiplying the RBC by the MCV.

This results in a much more reliable determination than the older method where MCV, rather than being measured, is calculated from the RBC and Hematocrit. The hematocrit (Hct) is the height of the column of cells measured directly after centrifugation in Wintrobe tubes or smaller glass columns. RBC can be measured using the turbidity of the blood sample measured with a photometer. The MCV is then calculated using the formula MCV = HCT x 1000/RBC. RDW is not available. The

alternate method for HB measurement is through a portable photometer with acceptable though less variability than an electronic particle counter.

Appendix II

Methodologic and Practical Problems with Portable Hematologic Equipment:

Coulter counter methodology has been shown to be reliable. Alternate methodologies can involve substantially more variation.[2] In the Keewatin Area Health Study it was decided that an electronic particle counter would be too expensive and would not travel well. In its place they used a portable machine called "AMES™ MINILAB". This micro processor-controlled photometer weighed only 450g. and functioned at a wide range of ambient temperatures. It used the Cyanmethemoglobin method on capillary samples for determination of HB and when compared with the Coulter STKR had a satisfactory coefficient of variation (CV=) 5.28.% There was some question of the accuracy of hemoglobin measurement with capillary samples which some have reported as producing lower values than venous samples.[37] In a comparable population survey conducted in Alaska, Margolis et al. used venous samples and a similar portable photometer arriving at a CV=1.3% and a variation of 2.5 g/L. This was cited as similar to values obtained in two other calibration experiments.[27]

The MINILAB could not perform the other determinations directly and so the calculation of MCV was measured using the formula: MCV = Hematocrit divided by red blood cell count. Hematocrit was measured using the Ames MICROSPIN and was determined to be accurate in an evaluation conducted by the Winnipeg Health Sciences Centre Laboratory.[2]The measurement of the red blood cell count was conducted using an

optical reader but resulted in a CV = 19% which adversely affected the reliability of the MCV values. Application of this methodology to other than HB testing in remote locations has limited reliability.

Appendix III: Glossary of Abbreviations

AUC	area under the curve
FEP	free erythrocyte protoporphyrin or protoporphyrin
fL	femtoliters $=10^{-8}$ Liters
HB	hemoglobin
HCT	hematocrit
HSC	Health Science Center
ID	iron deficiency
IDA	iron deficiency anemia
INACG	international anemia consulting group
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular hemoglobin
NHANES	national health and nutrition examination survey
RCT	randomized controlled trial
RCT	randomized controlled clinical trial
RDW	red blood cell distribution width
ROC	receiver operator curve
SD	standard deviation
SF	serum ferritin
TS	transferrin saturation
Z score	significance of difference between AUC and "line of no information"

Appendix IV Calculation of Normal from NHANES II

Between 1976 and 1980 the Second National Health and Nutrition Survey (NHANES II) was conducted in the USA.[190] It was a broad based sampling program of 27,801 individuals from 64 sampling areas using a probability sample which allowed extrapolation to the entire US population. The sample design was a stratified, multistage probability cluster of households throughout the US.[31] Unfortunately because only 22 infants between 6 and 11 months were able to give venous blood samples and did not have any exclusions based on illness or hemoglobinopathies these 22 were excluded.[44]

Individuals were excluded from the subject pool that was used to determine "normal" if they had evidence of deficiency as defined by three abnormal values of TS, FEP, MCV or Lead. After these exclusions, median and 95% limits were calculated. (Mean and SD were not used because the values of Fe and TS deviated markedly from a Gaussian distribution even after logarithmic transformation). There were some limitations of this approach for infants between 1 an 2 years of age because the small sample size made for increased uncertainty about the 95% ranges. Of 122 subjects the median Hb was 123 g/dl and 95% range (107 -138), for MCV 79 (67-88).