Laboratory Diagnosis of Iron-deficiency Anemia:

An Overview

GORDON H. GUYATT, MD, ANDREW D. OXMAN, MD, MAHMOUD ALI, MD, ANDREW WILLAN, PhD, WILLIAM McILROY, PhD, CHRISTOPHER PATTERSON, MD

Background and methods: ***To determine the diagnostic values of laboratory tests used in the diagnosis of iron-deficiency anemia, the authors conducted a systematic overview of the relevant literature. Computerized searches of the MEDLINE database yielded 1,179 potentially relevant citations. Fifty-five studies included the results of laboratory tests and histologic examination of the bone marrow for at least 50% of an identifiable patient group. In these 55 studies, quality was assessed and descriptive information concerning the study populations, the tests conducted, and the results was extracted, all in duplicate.***

Results: ***Serum ferritin radioimmunoassay was by far the most powerful test, with an area under the receiver operating characteristic curve of 0.95. Test properties differed for populations of patients with inflammatory, liver, or neoplastic disease and patients without these conditions. Likelihood ratio lines, which allow precise interpretation of results across the entire range of ferritin concentration values, were constructed for the individual populations.***

Conclusion: ***Serum ferritin radioimmunoassay is an extremely powerful test for the diagnosis of iron-deficiency anemia and, appropriately interpreted, can be applied to the complete range of patients.***

Key words: ***anemia; iron deficiency; likelihood ratios; meta-analysis; serum ferritin radioimmunoassay; diagnostic tests.*** J GEN INTERN MED 1992;7:145–153.

ANEMIA is a common problem in all age groups and populations, and iron deficiency a common cause. While the definitive diagnosis of iron deficiency is made by examination of the bone marrow obtained by aspiration, laboratory tests are often used to aid diagnosis. Mean cell volume (MCV) determination, transferrin saturation (TS) testing, and serum ferritin radioimmunoassay are used most commonly; determinations of red cell protoporphyrin (RCP) and red cell volume distribution (RDW) have also been suggested. Most recently, measurement of the red cell ferritin (RCF) has been suggested as being particularly helpful in evaluating patients who have inflammatory and liver disease, conditions in which serum ferritin testing is thought to be unreliable.

Received from the Departments of Medicine (GHG, MA, CP), Clinical Epidemiology and Biostatistics (GHG, ADO, AW, WM), and Family Medicine (ADO), Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.

Supported in part by the Ontario Ministry of Health. Drs. Guyatt and Oxman are Career Scientists of the Ontario Ministry of Health.

Address correspondence and reprint requests to Dr. Guyatt: Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre, Room 2C12, 1200 Main Street West, Hamilton, Ontario, Canada, L8N 3Z5.

Optimal use of a diagnostic test in clinical practice requires an accurate estimate of pretest probability and a knowledge of the test characteristics: sensitivity, specificity, and the likelihood ratios associated with various test results.[1] Any single study may provide estimates of these test characteristics, but these estimates may be distorted by a number of factors, including the strategy of sampling patients and the play of chance. Precise estimates of test characteristics cannot be achieved using sample sizes of most single studies.[2,3] We therefore conducted a comprehensive review of the literature concerning the laboratory diagnosis of iron-deficiency anemia. Specifically, we wished to ascertain, in patients with clinically significant anemia who are suspected of being iron-deficient, the likelihood ratios associated with MCV, TS, serum ferritin, RCP, RDW, and RCF with respect to the diagnosis of iron deficiency. In addition, we wished to determine what characteristics of the population (particularly age and presence of acute or chronic disease) influence performances of the tests.

PRIOR REVIEWS

Using search strategies described in detail below, we identified a number of reviews related to the diagnosis of iron deficiency.[4-8] These reviews, however, failed to meet many of the criteria of a scientific overview[9]: none reported a systematic search of the literature, provided a methodologic assessment of the evidence, or attempted a quantitative analysis. Furthermore, clinicians remain confused about the role of laboratory tests in the diagnosis of iron deficiency, and no prior review has clarified this role. We therefore concluded that a scientific overview was warranted.

METHODS

Literature Search

Two MEDLINE searches were done. The first was as follows: [iron or iron (tw)*] and [anemia/diagnosis or bone marrow/analysis or bone marrow/metabolism]. The second was: [iron (tw) or anemia or anemia (tw)] and [erythrocytes/analysis or erythrocytes/pathology

*Text word.

or erythrocyte count]. In the second search, if more than 100 articles were obtained from any particular MEDLINE file, "diagnosis" was added as a subheading. These searches were repeated for all MEDLINE files between 1966 and the time the final search was conducted (January 1990). The citation lists were reviewed by two of us (GHG and ADO), and any articles that either person thought were relevant were obtained. The citation lists of all reviews and primary articles that were obtained were examined, and any possibly relevant article was noted and obtained. For all abstracts, the first author was contacted and a full text of the study was requested.

Selection Criteria

The following criteria were used to select research:

1. The target population: patients over 18 years old with low levels of hemoglobin (<13.0 g/dL for men, <11.0 g/dL for women). In many studies, only a proportion of patients met these criteria and it was impossible to separate the patients who met the criteria from those who did not. So long as it appeared that 10% or more of the patients met the criteria, the studies were included.
2. The diagnostic intervention: quantitation of MCV, TS, serum ferritin, RCP, RDW, or RCF.
3. Outcomes: the relation between test results and findings on histologic examination of aspirated bone marrow. Studies were included only if data were presented in a manner that allowed the calculation of, at least, test sensitivity. Studies in which it was not possible to separate individual patients who had had bone marrow aspiration from those who had not had the procedure were included only if at least 50% of an identifiable subgroup of patients had had bone marrow aspiration.

Initially, study titles were examined for eligibility by two of us (GHG and ADO), and articles that either one thought might be relevant were obtained. Subsequently, full papers were reviewed by two of us (GHG and either ADO or MA), and papers judged relevant were included in the overview. Disagreements were resolved by conference.

Foreign-language papers that were thought potentially eligible were reviewed by a single reviewer who both understood the language and had medical training. Foreign-language papers for which we did not have easy access to a translator with medical training were excluded. One paper in Japanese and one in Chinese was excluded on this basis.[10]

Assessments of Methodologic Quality

The following criteria were used to assess methodologic quality.

1. The populations:
   - Ideal: consecutive anemic patients (with explicit definition of anemia) who consented to have bone marrow aspiration for histologic examination.
   - Second best: consecutive anemic patients who underwent bone marrow aspiration.
   - Worst: anything else.
2. The interventions:
   - Ideal: specified method of testing (i.e., how laboratory tests were done). If reference was made to an article that apparently provided a detailed method, this was considered adequate.
   - Anything else.
3. Outcome measures:
   - Bone marrow examined by two or more readers blinded to the results of other tests.
   - Either blinding or two or more readers, but not both.
   - Neither blinding nor two or more readers.

Papers were once again reviewed twice, with disagreement resolved by conference.

Data Collection

Two reviewers (GHG and, for each paper, one of ADO and MA) abstracted information from all papers. Information abstracted included: the numbers of patients in all age, gender, and disease categories, the process of patient selection ("consecutive," "approximating consecutive," or "arbitrary, probably unrepresentative"), whether the patients were anemic by our definition (versus anemic by authors' definitions, or impossible to separate anemic population), the proportion of patients having undergone bone marrow aspiration, and the results of the investigations in iron-deficient and non–iron-deficient subjects. Discrepancies were resolved by a single reviewer (GHG), who reviewed the original paper to discern the reason for the disagreement.

We found one instance in which several reports had been published with apparently overlapping data.[11-13] In this case, we included only the report with the most complete data set.

Analysis

For the initial relevance and methodologic quality ratings, agreement between two observers was calculated using a weighted kappa statistic[14] with quadratic

weights.[15] The principal analysis was conducted for articles in which dot plots or tables of data allowed test results for each individual subject to be abstracted. In the initial analysis, receiver operating characteristic (ROC) curves were generated, the area under each curve was calculated, and the areas under the curves were compared.[16] The data from the studies were combined to estimate the typical likelihood ratios across studies and associated 95% confidence intervals.[17]

Subsequent analyses were restricted to serum ferritin radioimmunoassay. Initially, we wished to see whether the results for individual studies were consistent. Using cutoff points of 18 or 45 μg/dL, homogeneity of the proportions of patients with iron deficiency above and below each cutoff point was formally tested using the Breslow-Day test.[18] This was done by comparing the odds ratios for each study (that is, the ratio of the odds of being above the cutoff point and the odds of being below the cutoff point).

In addition to random error, we identified the following potential sources of variability among the relevant studies.

1. The target population: age, gender, health state. The following health-state-related categories of patients were identified: patients who had chronic renal failure, those who had miscellaneous inflammatory disease, those who had infection, those who had rheumatoid arthritis, those who had liver disease, those who had inflammatory bowel disease, those who had hematologic malignancy, those who had nonhematologic malignancy, those without any other underlying disease, and a miscellaneous population of patients who did not fit into any one of these categories.
2. The methodologic quality of the studies (relevant factors included strategies for patient selection, the proportion of subjects for whom bone marrow aspiration was done, and blinded interpretation of the marrow).

When heterogeneity was found, each variable was tested in a univariate analysis to see whether some of the heterogeneity could be explained by that variable. For example, patients were divided into categories according to underlying health state and a test was conducted to see whether there was heterogeneity across the health states.

Models of the distribution of values in iron deficiency and non–iron deficiency were generated, and likelihood ratio lines calculated on the basis of the best-fit models.[19] The data proved to be skewed but were normalized by a natural logarithmic transformation, and the transformed data were used to generate the likelihood functions.

TABLE 1
Methodologic Characteristics of the Studies

	No. Studies (%)
Population	
Consecutive patients	40 (75.5%)
Sample approximates consecutive patients	5 (9.4%)
Arbitrary sample	8 (15.1%)
All anemic patients	25 (47.2%)
Can't separate anemic population	28 (52.8%)
Intervention	
Laboratory methods specified	50 (94.3%)
Laboratory methods not specified	3 (5.7%)
Outcome	
>80% of patients had bone marrow aspiration	50 (94.3%)
50% to 80% of patients had bone marrow aspiration	2 (3.8%)
Proportion of patients having bone marrow aspiration not clear	1 (1.9%)
Bone marrow examination results read by 2 or more blinded observers	4 (7.5%)
Bone marrow examination results read by 2 or more observers *or* blinded observer	18 (34.0%)
Bone marrow examination results read by one unblinded observer	31 (58.5%)

RESULTS

Agreement Studies

Initial literature searches generated 1,179 titles. After the citation lists of possible relevant articles were reviewed, 132 articles for which at least one of the two observers felt the article might be relevant were identified. Agreement concerning possible relevance was obtained for 1,035 of 1,080 (96%) of the titles included in the agreement study (weighted kappa 0.82). Of the 135 articles retrieved, 127 were evaluated by two independent reviewers. Two German articles were translated but not reviewed independently, and three others were inadvertently not reviewed independently. Fifty-five[11-13, 19-70] were ultimately judged relevant. Of the 127 reviewed independently, two reviewers agreed regarding relevance 103 times (weighted kappa 0.64).

For the validity criteria concerning population, intervention, and outcome, absolute agreements and weighted kappas were 0.72 and 0.40, 0.86 and 0.49, and 0.84 and 0.63, respectively. In most cases, oversight on the part of one or other of the reviewers was responsible for the disagreement. In the case of the population criterion, a problem arose from judgments combining bone marrow aspiration and patient selection in a single question. These were subsequently separated in the data extraction process, and the results from the data abstraction process were those used in the analysis. A summary of the methodologic characteristics of the studies is presented in Table 1.

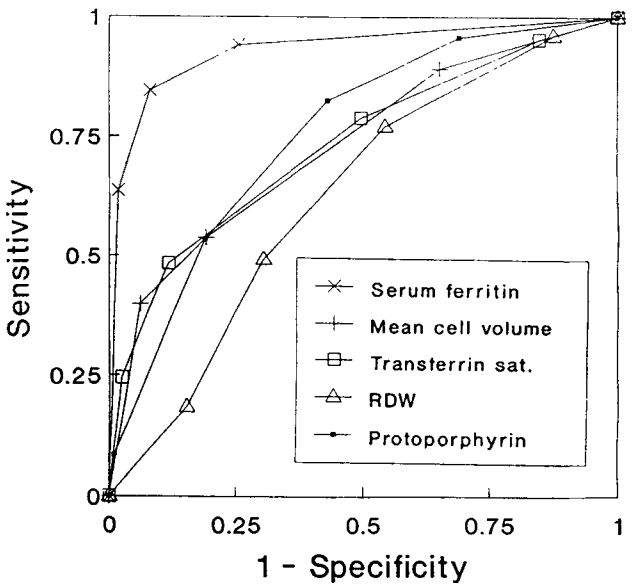


FIGURE 1. Receiver operating characteristic curves for serum ferritin radioimmunoassay, red cell protoporphyrin determination, transferrin saturation test, mean cell volume determination, and red cell volume distribution (RDW). For each value of each test, the y-axis represents the sensitivity of the test (the proportion of patients with iron deficiency correctly identified by the test) and the x-axis (1 — specificity) of the test (the proportion of patients without iron deficiency who are falsely classified as having iron deficiency).

TABLE 2

Area under the Receiver Operating Characteristic (ROC) Curve for Each Test

	Total Number of Subjects	Mean Area Under ROC Curve	95% Confidence Interval
All subjects from all studies			
Serum ferritin	2,579	0.95	0.94–0.96
Red cell protoporphyrin	288	0.77	0.71–0.83
Mean cell volume	436	0.76	0.72–0.80
Transferrin saturation	764	0.74	0.70–0.78
Red cell volume distribution	273	0.62	0.55–0.69
Only subjects with inflammatory, liver, or neoplastic disease			
Red cell ferritin	101	0.76	0.70–0.82
Serum ferritin	919	0.93	0.91–0.95

All Tests

Figure 1 shows the ROC curves associated with the tests. Data from all subjects were used in calculating these curves. The greater the area under the ROC curve, the more powerful the test. Visual inspection of the ROC curves demonstrates that radioimmunoassay for serum ferritin is by far the most powerful test. MCV, RCP, and TS determinations are comparable; the RDW test is the least useful. These findings are reflected in the calculated areas under the ROC curves, which are presented in Table 2. The area under the curve is a measure of the predictive value of the test; the greater the area, the greater the predictive value. Because of the large sample size, confidence intervals around the estimates of area under the curves are narrow, and power to differentiate between the curves is high. The difference between the area under the ferritin curve and the areas under the other curves is highly significant ($p<0.001$). The MCV determination proved significantly more powerful than all others ($p<0.001$). Differences between RCP, TS, and MCV results all could have occurred by chance (i.e., $p>0.05$).

The likelihood ratios for the five tests are presented in Table 3. The extreme likelihood ratios seen with high and low levels of serum ferritin reflect the power of the test. Likelihood ratios are considerably less extreme for the other tests.

Combinations of Tests

Five groups of researchers used regression analysis to examine the independent contributions of different tests to diagnosis.[24, 34, 39, 53, 71] Radioimmunoassay for ferritin proved the most powerful predictor in all analyses in which it was one of the tests carried out.[24, 39, 53, 71] One set of investigators found the erythrocyte sedimentation rate the only variable with additional predictive power.[39] In a second instance, knowledge of RDW and, to a lesser extent, TS improved the predictive model.[53] In a third study, TS was statistically significant, but made only a trivial contribution to the model's predictive power.[71] In a fourth study, MCV was the only test that improved the predictive model once serum ferritin concentration had been entered.[24] In all cases the additional contributions of other tests after serum ferritin assay were small.

Serum Ferritin

Because serum ferritin assay was so much more powerful than the other tests, and because results of regression analyses suggested that other tests provided little, if any, additional information, subsequent analyses were restricted to this test. The Breslow-Day test for the homogeneity of the likelihood ratios[18] revealed significant heterogeneity across studies ($p<0.001$). This suggested that test properties varied across studies. Our first hypothesis was that this heterogeneity could be explained by differences in underlying health states. The test for homogeneity across populations with differing health states was also positive ($p = 0.001$), suggesting that test properties do vary across populations. Because there was an inadequate number of subjects for

precise ascertainment of test properties in every health state, we selected two clinically relevant populations for further study. The first included patients who had "inflammatory disease," including any of the following conditions: chronic renal failure, miscellaneous inflammatory disease, infection, rheumatoid arthritis, liver disease, inflammatory bowel disease, hematologic malignancy, and nonhematologic malignancy. The second group was a "mixed population" including miscellaneous inpatients and outpatients for whom investigators did not identify any of the conditions listed above, but who had other health problems in addition to their suspected iron-deficiency anemias. We refer to this group as our mixed population.

We then looked to see whether there was remaining heterogeneity within the mixed group of patients. Heterogeneity remained ($p<0.001$). We tested to see whether age, gender, or methodologic quality of the studies explained the remaining variability. There was no heterogeneity across any of these variables, suggesting that they could not explain any of the residual variability.

Subsequent analysis was therefore done for both the inflammatory disease group and the mixed population. Likelihood ratio lines and their associated 95% confidence intervals for the two populations were constructed, and are presented in Figure 2. The properties of serum ferritin assay, reflected in the likelihood ratio lines, clearly differ for the two populations. The equation for calculating the likelihood ratio associated with serum ferritin values in the mixed population is as follows:

$$L = e^{(0.65429 - 1.6985 \cdot \ln(x))}$$

where L is the likelihood ratio and x is the serum ferritin value.

Using the same notation, the equation for calculating the likelihood ratio line associated with serum ferritin values in the inflammatory population is as follows:

$$L = e^{7.4793 - 1.7807 \cdot \ln(x))}$$

TABLE 3

Likelihood Ratios of the Tests

Interval	Number Iron Deficient	Number Not Iron Deficient	Likelihood Ratio	95% Confidence Interval*
Serum ferritin				
≥100 μg/L	48	1,320	0.08	0.07–0.09
45 < 100 μg/L	76	398	0.54	0.48–0.60
35 < 45 μg/L	36	43	1.83	1.47–2.19
25 < 35 μg/L	58	50	2.54	2.11–2.97
15 < 25 μg/L	117	29	8.83	7.22–10.44
≤15 μg/L	474	20	51.85	41.53–62.27
Red cell protoporphyrin				
≤50 μg/dL	1	15	0.12	0.00–0.25
50 < 150 μg/dL	42	132	0.56	0.48–0.64
150 < 250 μg/dL	26	23	2.01	1.44–2.58
250 < 350 μg/dL	17	5	6.05	2.76–9.34
≥350 μg/dL	14	3	8.31	2.60–14.02
Mean cell volume				
≥90 μm^3	24	128	0.29	0.21–0.37
85 < 90 μm^3	32	63	0.76	0.56–0.96
80 < 85 μm^3	43	71	0.91	0.71–1.11
75 < 80 μm^3	26	39	1.00	0.69–1.31
70 < 75 μm^3	31	14	3.33	1.99–4.67
≤70 μm^3	58	7	12.47	6.13–18.81
Transferrin saturation				
≥50%	4	44	0.15	0.06–0.24
30 < 50%	22	82	0.43	0.31–0.55
20 < 30%	36	111	0.52	0.41–0.63
10 < 20%	90	178	0.81	0.70–0.92
5 < 10%	70	44	2.54	1.99–3.09
≤5%	72	11	10.46	6.42–14.50
Red cell volume distribution				
≤15	29	80	0.61	0.48–0.74
15 < 17	25	50	0.84	0.63–1.05
17 < 21	35	33	1.78	1.35–2.21
≥21	13	8	2.72	1.34–4.10

*95% confidence interval around likelihood ratio for serum ferritin assay, red cell protoporphyrin determination, and transferrin saturation test.

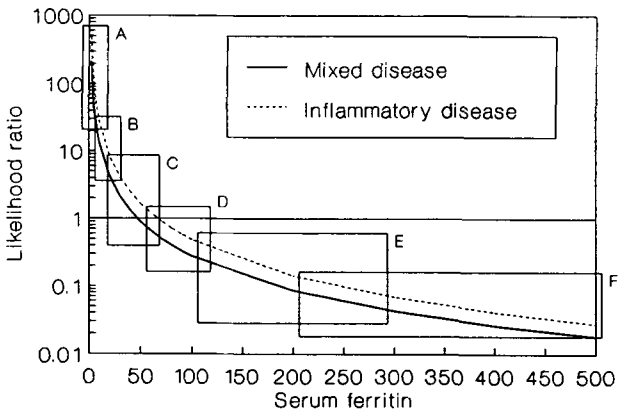
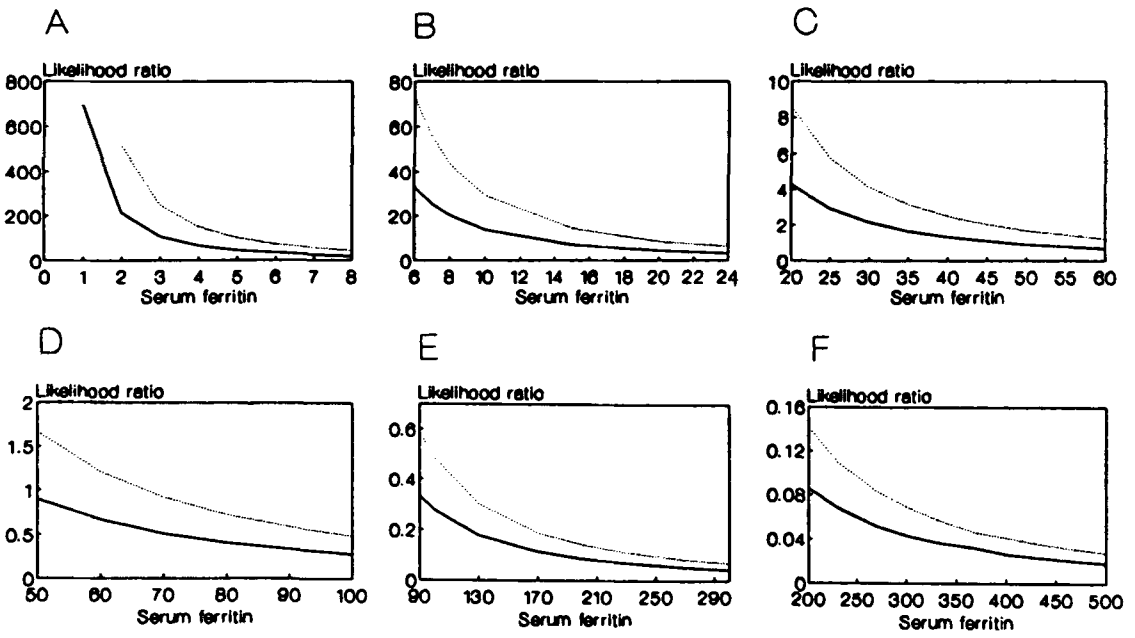


FIGURE 2. Likelihood ratio lines of serum ferritin radioimmunoassay for the "mixed" and "inflammatory disease" populations. The top figure is a plot of the serum ferritin in natural units (μg/L) against the logarithm of the likelihood ratio. Subfigures A to F correspond to blocks A to F in the large figure and are included to give a better sense of the curves at different sections. For each of subfigures A to F, both the serum ferritin concentrations and the likelihood ratios are plotted in natural units (μg/L). For each of subfigures A to F, the scales of both serum ferritin concentrations and the likelihood ratios change to best depict the significance of the curve for those values of serum ferritin.



In our final analysis we examined in the population of patients with inflammatory disease the issue of the absolute usefulness of determining serum ferritin concentrations in comparison with the usefulness of RCF testing, proposed to take its place in this population. Only two studies examined the properties of RCF.[30, 31] The area under the ROC curves for RCF testing in patients with inflammatory disease was substantially less than that for serum ferritin determination (Table 2).

DISCUSSION

The current study meets most of the methodologic criteria developed for the conduct of scientific overviews.[9, 72] The results are therefore likely to present a valid summary of the usefulnesses of laboratory tests in the diagnosis of iron-deficiency anemia.

Observer agreement for deciding the relevances of the individual studies was not perfect, but was very good. Observer agreement regarding the validity was not nearly so good. Lack of precision in judgments of methodologic quality may have contributed to the fact that differences in levels of methodologic quality did not explain any of the differences in study results. The difficulty we had agreeing emphasizes the need for more than one independent evaluation in deciding study relevance and validity when performing scientific overviews.

The heterogeneity found across studies, even within populations, raises questions about the appropriateness of aggregating the data. The most likely explanation for the remaining heterogeneity is that even our mixed population was contaminated, to varying degrees, with patients who had underlying inflammatory conditions. Nevertheless, the test properties derived from the overview represent the best available estimate and are likely to be a more accurate guide for practice than are the results of any individual study.

Another limitation of the data is that results of MCV and RDW determinations are available to physicians earlier than the results of other tests. This may bias the patients who enter the studies. For instance, patients may be less likely to be included if their MCVs are normal. Alternatively, patients whose MCVs are very low in relation to their levels of hemoglobin may have

laboratory studies to rule in thalassemia, never have histologic examination of bone marrow, and thus be excluded from the studies. Despite this limitation, the studies enrolled heterogeneous populations that provided a reasonably fair representation of patients whom physicians suspect of having iron deficiency.

An associated limitation of the data is that few studies have formally looked at whether radioimmunoassay for serum ferritin adds important information to that obtained by the routinely available MCV test. This could be done using, for instance, a regression analysis. Fortunately, ferritin determination is so much more powerful than MCV testing that the conclusion that it adds diagnostic power remains secure. A more difficult issue is that there may be subgroups of patients with specific patterns of MCV or RDW results in which ferritin assay may not add important information. Unfortunately, the available data do not allow exploration of this issue.

It has been apparent for well over a decade that serum ferritin radioimmunoassay results are systematically altered by underlying inflammatory or liver disease. Our overview confirmed this finding. However, contrary to what most investigators have concluded, this does not decrease the value of serum ferritin measurement in these populations. Rather, the interpretation of any given ferritin result in patients with inflammatory or liver disease must differ from the interpretation of the same test result in patients without this disease. The characteristics of the test in these two populations are depicted in Figure 2. For example, a serum ferritin concentration of 30 ng/mL is associated with a likelihood ratio of 2 in a general population and a likelihood ratio of 4 in a population of patients with inflammatory disease. Thus, if one had a patient with a prior probability of iron deficiency of 50% and that patient's serum ferritin concentration was 30 ng/mL, the posttest likelihood would be 0.66 if the patient did not have inflammatory or liver disease or 0.80 if the patient did have inflammatory or liver disease.

Because they allow a precise interpretation of the meaning of any individual test result, likelihood ratio lines, when they can be generated, provide the most powerful guide to application of test results in clinical practice. Data from individual studies, however, are generally too sparse for calculation of likelihood ratio lines. The strength of this overview is that aggregating data across studies provided the power for calculating likelihood ratio lines with relatively narrow confidence intervals.

These results can be applied directly in clinical practice. The clinician begins by making an estimate of the probability of iron deficiency based on information he or she has prior to receiving the results of the serum ferritin determination. This estimate can be referred to as the "pretest probability estimate." For the diagnosis of iron deficiency, this estimate can be based on a number of factors, including: history of previous anemia; dietary history; history of any bleeding (including melena); ingestion of gastric irritant drugs; historical clues to other possible causes of anemia (including: weakness, fatigue, or easy bruising; bone pain suggestive of myeloma or other underlying malignancy; and history of chronic inflammatory conditions such as rheumatoid arthritis); and findings on physical examination such as abdominal mass, spontaneous bruising, lymphadenopathy, splenomegaly, or melena or findings suggesting chronic inflammatory disease.[73] If the pretest probability is neither extremely low (<10% for instance) nor extremely high (>90%), radioimmunoassay for serum ferritin should be ordered, and the result obtained. The clinician should then decide whether the patient fits into the inflammatory disease population or the mixed population, and should find the likelihood ratio associated with the test result by referring to Figure 2. The posttest probability can then be calculated by hand, or using a simple nomogram or "likelihood ratio card."[1] This process is illustrated in Table 4, which provides representative pretest and posttest probabilities given different serum ferritin testing results.

A much simpler approach would treat serum ferritin concentration as having three categories. A value <15 μg/L confirms the diagnosis of iron deficiency, while a value of >100 μg/L rules out iron deficiency. Intermediate values mandate further investigation. While less precise than the approach described in the previous paragraph, this simpler strategy may in many instances be adequate.

In conclusion, the results of this study should alter

TABLE 4

Posttest Probabilities of Iron Deficiency Given Varying Pretest Probabilities and Results of Serum Ferritin Determinations

Serum Ferritin Result	Pretest Probability			
	0.2	0.4	0.6	0.8
120 μg/L	0.05 *0.08**	0.12 *0.19*	0.23 *0.34*	0.45 *0.58*
70 μg/L	0.11 *0.19*	0.25 *0.38*	0.43 *0.58*	0.67 *0.79*
50 μg/L	0.18 *0.30*	0.38 *0.53*	0.57 *0.75*	0.78 *0.87*
30 μg/L	0.35 *0.51*	0.59 *0.73*	0.76 *0.86*	0.90 *0.94*
10 μg/L	0.78 *0.8*	0.90 *0.95*	0.95 *0.98*	0.98 *0.99*

*Values in Roman type are results for the mixed population; values in italics are results for the inflammatory disease population.

clinical recommendations and practice in the diagnosis of iron-deficiency anemia. First, radioimmunoassay for the determination of serum ferritin concentration should be the only blood test ordered. Second, the traditional cutoff point dividing normal and abnormal, which in most laboratories is between 12 and 20 μg/L, is not optimal. The likelihood of iron deficiency does not start to drop until values are higher than approximately 40 μg/L (for general populations) or 70 μg/L (for those with inflammatory or liver disease). Third, the test needn't be abandoned in the management of patients who have inflammatory and liver disease, although the results should be interpreted somewhat differently for such patients than for those without these conditions. Fourth, knowledge of the precise properties of serum ferritin can enhance the power of the laboratory diagnosis of iron deficiency.

The authors thank Dr. Joel Singer for conducting the initial preliminary data analysis and Debbie Maddock for her diligent work in keeping track, and supervising the retrieval, of the citations.

REFERENCES

1. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology, a basic science for clinicians. Boston/Toronto: Little, Brown, 1985.
2. Arkin CF, Wachtel MS. How many patients are necessary to assess test performance? JAMA. 1990;263:275-8.
3. Heckerling PS. Confidence in diagnostic testing (editorial). J Gen Intern Med. 1988;3:604-6.
4. Stuckey WJ. Common anemias: a practical guide to diagnosis and management. Geriatrics. 1983;38:42-8.
5. Fischer SL, Fischer SP. Mean corpuscular volume. Arch Intern Med. 1983;143:282-3.
6. Witte DA. Laboratory tests to confirm or exclude iron deficiency. Lab Med. 1985;16:671-5.
7. Worwood M. Serum ferritin. Clin Sci. 1986;70:215-20.
8. Besa EC. Approach to mild anemia in the elderly. Clin Geriatr Med. 1988;4:43-55.
9. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J. 1988;138:697-703.
10. Chung-hua I, Hseuh T. Iron deficiency in late pregnancy: observation on free erythrocyte protoporphyrin, transferrin saturation, and marrow iron in 16 cases. Chin Med J. 1984;64:108-11.
11. Milman N. Plasma transferrin and the relation to iron status in patients with chronic uraemia. Clin Nephrol. 1981;16:314-20.
12. Milman N, Christensen TE, Pedersen NS, Visfeldgt J. Serum ferritin and bone marrow iron in non-dialysis, peritoneal dialysis and haemodialysis patients with chronic renal failure. Acta Med Scand. 1980;207:201-5.
13. Milman N, Christensen TE, Visfeldgt J. Diagnostic efficiency of various laboratory tests in the assessment of bone marrow iron stores in patients with chronic uraemia. Scand J Haematol. 1981;26:257-64.
14. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213-20.
15. Cicchetti DV, Fleiss JL. A comparison of the null distributions of weighted kappa and the c ordinal statistic. Appl Psychol Measurement. 1977;1:195-201.
16. 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.
17. Department of Clinical Epidemiology and Biostatistics. Interpretation of diagnostic data: 5. How to do it with simple maths. Can Med Assoc J. 1983;129:22-9.
18. Breslow NE, Day NE. Statistical methods in cancer research. Volume 1—the analysis of case-control studies. Lyon, France: IARC Scientific Publications, 1980.
19. Albert A. On the use and computation of likelihood ratios in clinical chemistry. Clin Chem. 1982;28:1113-9.
20. Jakobsen E, Engeset A, Sandstad B, Aas M. Serum ferritin and bone marrow haemosiderin in patients with malignancies and in health controls. Scand J Haematol. 1982;28:264-71.
21. Raper CGL, Rosen C, Choudhury M. Automated red cell indices and marrow iron reserves in geriatric patients. J Clin Pathol. 1977;30:353-5.
22. Blend MJ, Tischler IW, Chernicoff DP, Caleel GT. Correlation of serum ferritin concentration with bone marrow iron stores. J Am Osteopath Assoc. 1981;80:805-9.
23. Barnett MD, Gordon YB, Amess JAL, Mollin DL. Measurement of ferritin in serum by radioimmunoassay. J Clin Pathol. 1978;31:742-8.
24. Brink S, Van Schalkwyk DJ. Serum ferritin and mean corpuscular volume as predictors of bone marrow iron stores. S Afr Med J. 1982;61(12):432-4.
25. Forman DT, Vye MV. Immunoradiometric serum ferritin concentration compared with stainable bone-marrow iron as indices to iron stores. Clin Chem. 1980;26:145-7.
26. Oertel BM, Bombik BM, Stephan M, Gerhartz H. Ferritin in bone marrow and serum in iron deficiency and iron overload. Blut. 1978;37:113-7.
27. Walsh JR, Fredrickson M. Serum ferritin, free erythrocyte protoporphyrin, and urinary iron excretion in patients with iron disorders. Am J Med Sci. 1977;273:293-300.
28. Baynes RD, Bothwell TH, Bezwoda WR, et al. Hematologic and iron-related measurements in rheumatoid arthritis. Am J Clin Pathol. 1987;87:196-200.
29. Witte DL, Angstadt DS, Davis SH, Schrantz RD. Predicting bone marrow iron stores in anemic patients in a community hospital using ferritin and erythrocyte sedimentation rate. Am J Clin Pathol. 1988;90:85-7.
30. Isa L, Jean G, Silvani A, et al. Evaluation of iron stores in patients with alcoholic liver disease: role of red cell ferritin. Acta Haematol. 1988;80:85-8.
31. Brown RD, Benfatto J, Gibson J, Kronenberg H. Red cell ferritin and iron stores in patients with chronic disease. Eur J Haematol. 1988;40:136-41.
32. Lindstedt G, Lundberg P, Bjorn-Rasmussen E, Magnussen B. Serum-ferritin and iron-deficiency anaemia in hospital patients. Lancet. 1980;1:205-6.
33. Ballas SK. Normal serum iron and elevated total iron-binding capacity in iron-deficiency states. Am J Clin Pathol. 1979;71:401-3.
34. Beck JR, Gibbons AB, Cornwell GG, Rawnsley HM. Multivariate approach to predictive diagnosis of bone-marrow iron stores. Am J Clin Pathol. 1979;70(suppl):S665-S670.
35. Krause JR, Stolc V. Serum ferritin and bone marrow biopsy iron stores. II: Correlation with low serum iron and Fe/TIBC ratio less than 15%. Am J Clin Pathol. 1980;74(suppl):S461-S464.
36. Krause JR, Stolc V. Serum ferritin and bone marrow biopsy iron stores. I: Correlation with absence of iron in biopsy specimens. Am J Clin Pathol. 1979;72(suppl):S817-S820.
37. 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 Pathol. 1978;70:79-84.
38. Puolakka J. Serum ferritin in the evaluation of iron status in young healthy women. Acta Obstet Gynecol Scand. 1980;95(suppl):S35-S41.
39. Witte DL, Kraemer DF, Johnson GF, et al. Prediction of bone marrow iron findings from tests performed on peripheral blood. Am J Clin Pathol. 1986;85:202-6.
40. Patterson C, Turpie ID, Benger AM. Assessment of iron stores in anemic geriatric patients. J Am Geriatr Soc. 1985;33:764-7.
41. Blumberg AB, Marti HRM, Graber CG. Serum ferritin and bone marrow iron in patients undergoing continuous ambulatory peritoneal dialysis. JAMA. 1983;250:3317-9.
42. Sharma JC, Roy SN. Value of serum ferritin as an index of iron deficiency in elderly anaemic patients. Age Ageing. 1984;13:248-50.
43. Harju E, Pakarinenen A, Larmi T. A comparison between serum ferritin concentration and the amount of bone marrow stainable iron. Scand J Clin Lab Invest. 1984;44:555-6.

44. Ali MAM, Luxton AW, Walker WHC. Serum ferritin concentration and bone marrow iron stores: a prospective study. Can Med Assoc J. 1978;118:945-6.
45. Mazza J, Barr RM, McDonald JWD, Valberg LS. Usefulness of the serum ferritin concentration in the detection of iron deficiency in a general hospital. Can Med Assoc J. 1978;119:884-6.
46. Kalmin ND, Bobson EB, Bettigole RE. Serum ferritin and marrow iron stores. N Y State J Med. 1978;78:1052-5.
47. Smith RJ, Davis P, Thomson BR, et al. Serum ferritin levels in the anemia of rheumatoid arthritis. J Rheumatol. 1977; 4(suppl):S389-S392.
48. Ali M, Rigolosi R, Fayemi AO, et al. Failure of serum ferritin levels to predict bone-marrow iron content after intravenous iron-dextran therapy. Lancet. 1982;1:652-5.
49. McLaren GD, Carpenter JT, Nino HV. Erythrocyte protoporphyrin in the detection of iron deficiency. Clin Chem. 1975;21:1121-7.
50. Sorbie J, Valberg LS, Corbett WEN, Ludwig J. Serum ferritin, cobalt excretion and body iron status. Can Med Assoc J. 1975;112:1173-8.
51. Hussein S, Prieto J, O'Shea M, et al. Serum ferritin assay and iron status in chronic renal failure and haemodialysis. Br Med J. 1975;1:546-8.
52. Wachtel TJ, Reuben DB, Fulton JP, Wartman SA. The use of "range of uncertainty." Med Decis Making. 1985;5:325-34.
53. Charache S, Gittelsohn AM, Allen H, et al. Noninvasive assessment of tissue iron stores. Am J Clin Pathol. 1987;88:333-7.
54. Paine CJ, Polk A, Eichner ER. Analysis of anemia in medical inpatients. Am J Med Sci. 1974;268:37-44.
55. Bentley DP, Williams P. Serum ferritin concentration as an index of storage iron in rheumatoid arthritis. J Clin Pathol. 1974; 27:786-8.
56. Child Ja, Brozovic B, Dyer NH, et al. The diagnosis of iron deficiency in patients with Crohn's disease. Gut. 1973;14:642-8.
57. Beallo R, Dallman PR, Schoenfeld PY, Humphreys MH. Serum ferritin and iron deficiency in patients on chronic haemodialysis. Trans Am Soc Artif Intern Organs. 1976;22:73-9.
58. Beutler E, Robson MJ, Buttenwieser E. A comparison of the plasma iron, iron-binding capacity, sternal marrow iron and other methods in the clinical evaluation of iron stores. Ann Intern Med. 1958;48:60-80.
59. Oluboyede OA. Iron studies in pregnant and non-pregnant women with haemoglobin SS or SC disease. Br J Obstet Gynaecol. 1980;87:989-95.
60. Lipshchitz DA, Cook JD, Finch CA. A clinical evaluation of serum ferritin as an index of iron stores. N Engl J Med. 1974; 290:1213-6.
61. Gokal R, Millard PR, Weatherall DJ, et al. Iron metabolism in haemodialysis patients. Q J Med. 1979;48:369-91.
62. Addison GM, Beamish MR, Hales CN, et al. An immunoradiometric assay for ferritin in the serum of normal subjects and patients with iron deficiency and iron overload. J Clin Pathol. 1972;25:326-9.
63. England JM, Ward SM, Down MC. Microcytosis, anisocytosis and the red cell indices in iron deficiency. Br J Haematol. 1976;34:589-97.
64. Bartels U, Pederson NS, Jarnum S. Iron absorption and serum ferritin in chronic inflammatory bowel disease. Scand J Gastroenterol. 1978;13:649-56.
65. Thompson WG, Meola T, Lipkin M, Freedman ML. Red cell distribution width, mean corpuscular volume, and transferrin saturation in the diagnosis of iron deficiency. Arch Intern Med. 1988;148:2128-30.
66. Witte DL, Dick F, Goeken J, et al. C-reactive protein aids interpretation of serum ferritin. Clin Chem. 1985;31:1011.
67. Nelson R, Chawla M, Connolly P, LaPorte J. Ferritin as an index of bone marrow iron stores. South Med J. 1978;71:1482-4.
68. Puolakka J, Janne O, Pakarinen A, Bihko R. Serum ferritin in the diagnosis of anemia during pregnancy. Acta Obstet Gynecol Scand. 1980;95:57-63.
69. Wohlenberg H, Panitz N, Wacheck W. Hypoferramie und serumferritin. Dtsch Med Wochenschr. 1980;105:1527-30.
70. Leyland MJ, Ganguli PC, Blower D, Delamore IW. Immunoradiometric assay for ferritin in human serum. Scand J Haematol. 1975;14:385-92.
71. Guyatt GH, Patterson C, Ali M, et al. Diagnosis of iron deficiency anemia in the elderly. Am J Med. 1990;88:205-9.
72. Mulrow CD. The medical review article: state of the science. Ann Intern Med. 1987;106:485-8.
73. Patterson C, Guyatt GH, Singer J, et al. Iron deficiency anemia in the elderly: the diagnostic process. Can Med Assoc J. 1991;144:435-40.