Estimation of Hepatic Hemosiderin by Gross Inspection of Iron-Stained Histologic Sections

HAROLD O. CONN, ROSLYN H. STRAUSS, and MARILYN B. CONN

Medical Service, Veterans Administration Hospital, West Haven, Connecticut and Yale University School of Medicine, New Haven, Connecticut

Received for publication 30 September 1971

In the course of a prospective, controlled evaluation of the effect of portacaval anastomosis on hepatic hemosiderin deposition(1), a large number of iron-stained slides were evaluated blindly, semiquantitatively and serially. Those slides in which excessive deposition of hemosiderin was present microscopically appeared blue in color to gross inspection while those in whom little or no iron was present histologically appeared pink grossly. The present study was undertaken to compare the semiquantitative grading of hemosiderin in liver biopsies by microscopic and gross examination and to determine whether the gross interpretation of iron-stained liver tissue sections is a reliable way of grading hemosiderin deposition.

METHODS AND MATERIALS

Two hundred and sixteen histologic slides of liver tissue stained for iron with the Mallory modification of the technique of Gomori(2) were graded semiquantitatively by microscopic examination on a 1+ to 4+ basis for hemosiderin deposition. The sections of liver were either biopsy or autopsy tissue 6-μ thick obtained from patients entered into two investigations of prophylactic portacaval anastomosis (PCA)(3). Each patient was represented by at least two specimens, one at the time of inclusion into the study and a second one 6 months to 12 years later. Approximately half the patients had PCA and half had not.

Two hundred and one of these slides were randomized, given code numbers

1 Supported by grants from the V.A. Research Service, The Irwin Strasburger Memorial Medical Foundation and the Stratfield Fund.

Copyright © 1972, by Academic Press, Inc.
and read blindly by three observers, one a physician, one a histology technician and the third a research assistant with no laboratory experience. The criteria for microscopic grading were described previously (1). The microscopic interpretations represent the consensus of three separate readings. Criteria for gross grading was as follows: grade 0: pink; grade 1+: questionable bluish tinge; grade 2+: slight bluish tinge; grade 3+: purple color; grade 4+: deep blue color.² The three observers were simply instructed to match by color each unknown slide with 0, 1+, 2+, 3+ and 4+ standards.

One of the observers (Observer A) graded the slides on two separate occasions. After the initial reading the slides were rerandomized, recoded and reexamined. If the two interpretations did not agree, a third evaluation was performed and the consensus of the three readings obtained by determining the mean grade rounded to the nearest whole grade.

Selected standard slides were available to the investigators as points of reference for each of the grades of iron deposition. Histologic interpretations were done in batches of ten slides and each interpretation of the whole series was performed during a 2-week period. The gross interpretations were each performed at one or two sittings.

RESULTS

The gross interpretations of the presence and amount of hepatic hemosiderin are shown in Table 1. Observer A reported iron in 38 (19%), B in 97 (48%) and C in 78 (39%). Although there was wide variation in the interpretation of iron by the three observers, the degree of variation was much smaller in the specimens with excessive iron (3+ and 4+) than in those with small amounts of iron (1+.

| Type of interpretation | Total no. | Grade of hemosiderin |
|------------------------|-----------|----------------------|
|                        |           | 0 | 1+ | 2+ | 3+ | 4+ |
| Histologic*            | 201       | 14 | 26 | 14 | 9  | 4  |
| Gross                  |           |   |    |    |    |    |
| Observer A*            | 201       | 163| 18 | 10 | 6  | 4  |
| Observer B             | 201       | 104| 58 | 24 | 10 | 5  |
| Observer C             | 201       | 123| 39 | 24 | 11 | 4  |
| Consensus              | 201       | 139| 31 | 17 | 9  | 5  |

*Consensus of three interpretations.

² Preparation of sections of varying thickness, from 3 to 15 μ, showed that the criteria remained essentially the same and indicated that small variations in the thickness of individual sections does not affect gross grading. Although iron-stained slides may fade over a long period of time, there was no discernible change in color during the several months of this study.
and 2+). The distribution of hemosiderin determined by consensus of the three observers was quite similar to the pattern reported by microscopic grading (Table 1).

Comparison of Gross and Microscopic Interpretations. Concurrence of the gross grading with the microscopic grading by the three observers is shown in Table 2. Agreement ranged from 83% (Observer A) to 63% (Observer B). The consensus of the three gross interpretations agreed with the microscopic grades in 81%. Furthermore, almost all of the disagreements were by a single grade, and in no instance was the discrepancy greater than two grades. The great majority of the discrepancies were misinterpretations of 0 or 1+ microscopic grades of hemosiderin which were read grossly as 1+ or 0, respectively.

Inter-Individual Variation of Gross Interpretations. The three observers were unanimous in 92 of the 201 specimens (46%), in 72 of which no hemosiderin was detected. In 17 additional patients all three observers agreed that an abnormal amount of iron was present, although they were not unanimous in grading it. Thus, in 109 samples the three observers agreed unanimously about the normality or abnormality of the specimens. Calculated on this basis these observers agreed with each other’s interpretations in 419 of the 608 inter-observer comparisons (69%). This figure is similar to the frequency of inter-observer variability in the interpretation of chest X-rays, barium esophagrams for esophageal varices, liver scans for space-occupying lesions and a variety of other types of subjective medical judgments.

Intra-Individual Variation of Gross-Interpretations. In the initial gross grading of the 201 slides, 162 were read as iron-free and 39 showed hemosiderin (Table 3). In the second interpretation 159 were hemosiderin-free and 42 were read as showing iron. The two readings concurred in 182 of the 201 specimens (91%). In 18 of the 19 discordant readings the difference was one grade or less and only once did the two interpretations differ by more than one grade. This degree of reproducibility compares well with the agreement between microscopic interpretations, which agreed in 92%. In no instance did the two microscopic readings differ by more than one grade.

### Table 2

**Comparison of Gross and Microscopic Interpretations of Hemosiderin Grade**

| Observer | Agreement | Disagreement | Disagreement > 1 grade | Disagreements Associated with 0 or 1+ Hemosiderin |
|----------|-----------|--------------|------------------------|-----------------------------------------------|
|          | No.       | %  | No. | %  | No. | %  | No. | %  |
| A        | 166       | 83 | 35  | 17 | 1   | 0.5 | 26  | 74 |
| B        | 126       | 63 | 75  | 37 | 5   | 2.5 | 56  | 75 |
| C        | 146       | 73 | 55  | 27 | 5   | 2.5 | 36  | 65 |
| Consensus| 162       | 81 | 39  | 19 | 2   | 1.0 | 27  | 69 |

* Percent total.
  
  + Percent disagreement.
  
  * Percent of total number of gross-microscopic disagreements.
DISCUSSION

A number of investigators have previously demonstrated the reliability of the semiquantitative microscopic estimation of hepatic hemosiderin and its close correlation with chemical measurements of hepatic iron content (4–8). None of these investigators has evaluated the gross interpretation of histologic sections stained for iron. The present observations indicate that gross examination of iron-stained liver biopsy slides provides a simple, reliable, but rough estimate of the amount of hepatic hemosiderin.

The major advantage of gross interpretation over microscopic examination is the saving in time of highly trained personnel. In the studies described here, batches of slides could be read grossly in less than one-fifth the time it took for microscopic examination. In addition, the microscopic examination requires expensive equipment and, usually, a more sophisticated, better-trained observer. Reliable technicians, who have not been trained in histologic interpretation, can be easily taught to interpret gross specimens quickly and well. Reliability can be further assured by providing appropriate reference standards and by requiring duplicate estimations by the same or different observers.

Observer variability is ubiquitous and surprisingly constant in all types of subjective interpretations, including physical, radiologic, endoscopic and histologic examinations. The histologic quantitation of hepatic hemosiderin is not free of this hazard, nor is the gross evaluation as this paper shows. The degree of inter-individual variation in a wide variety of interpretations ranges from 15 to 35% and intra-individual error from 15 to 25% (9). In the present study the overall inter-individual error was 31% and the intra-individual variation, in the one observer so studied, 9%, values not different from those cited. Furthermore, the diagnostic agreement between the gross and microscopic interpretations was not different from radiologists' agreement with the endoscopic diagnosis of esophageal varices (10) or other similar comparisons (11).

Although the three observers' interpretations differed from each other and from the microscopic grading, each observer showed a consistent trend. Observer A, for example, tended to undergrade the gross interpretations, especially those slides with small amounts of iron deposition (grades 1+ and 2+). The other ob-
HEPATIC HEMOSIDERIN ESTIMATIONS

servers, especially B, tended to overgrade the sections with little iron. It is clear, and not surprising, that grossly one cannot differentiate iron-free specimens from those with small, clinically insignificant amounts of iron. Gross interpretation can thus serve as a simple screening procedure to separate those slides with significant hemosiderin deposition from those without. More critical analysis, of course, requires experienced microscopic interpretation.

The lesser sensitivity of the gross interpretation than the microscopic examination may well be a blessing in disguise. In our studies(1) and in those of other workers(12–15), for example, small, inconsequential amounts of hepatic hemosiderin are found in 25–50% of patients with alcoholic cirrhosis. Clinically significant amounts of hemosiderin are always easily evident—grade 3+ or 4+, which was invariably recognized grossly. Furthermore, histologic grading systems for hepatic hemosiderin do not correlate linearly with the concentration of iron. The lower histologic grades (1+ to 2+) are associated with relatively small amounts of iron (10–100 mg/100 gm liver) while the higher histologic grades (3+ to 4+) occur in livers with disproportionately large and variable amounts of iron. A liver biopsy with 4+ iron deposition microscopically may contain from 200 to 2,000 mg/100 gm liver. In effect, gross interpretation excludes the low grade ore, which is of doubtful clinical value, and emphasizes the rich deposits of iron, which indicate significant degrees of hemosiderosis.

One disadvantage of gross interpretation is the inability to differentiate between parenchymal and reticuloendothelial deposition of hemosiderin. In patients with excessive hemosiderin accumulation, the pigment is deposited in all types of hepatic cells—parenchymal, Kupffer, macrophagic and biliary. In patients with intravascular hemolysis, however, the pigment may be limited to the reticuloendothelial cells. Since the exclusive deposition of hemosiderin in the reticuloendothelial cells is rarely large in quantity, it is not often apparent grossly, until the process is sufficiently advanced to cause more generalized deposition.

REFERENCES

1. Conn, H. O., Hepatic hemosiderin deposition after portacaval anastomosis. Gastroenterology 62, 61–72 (1972).
2. Mallory, F. B., Pathologic Technique, p. 137. Philadelphia, W. B. Saunders Co. 1942.
3. Conn, H. O., Lindenmuth, W. W., May, C. J. and Ramsby, R. G., Prophylactic portacaval anastomosis. A tale of two studies. Medicine 51, 27–40 (1972).
4. Mayet, F. C. and Bothwell, T. H., Hepatic iron stores in three different racial groups. S. Afr. J. Med. Sci. 29, 55–59 (1964).
5. Lundvall, O., Weinfield, A. and Lundin, P., Iron stores in alcohol abusers. I. Liver iron. Acta Med. Scand. 185, 239–269 (1969).
6. MacDonald, R. A. and Pechet, G. S., Tissue iron and hemochromatosis. AMA Arch. Intern. Med. 116, 381–391 (1965).
7. Frey, W. G., III, Gardner, M. H. and Pillsbury, J. A., Quantitative measurement of liver iron by needle biopsy. J. Lab. Clin. Med. 72, 52–57 (1968).
8. Barry, M. and Sherlock, S., Measurement of liver-iron concentration in needle-biopsy specimens. Lancet 1, 100–103 (1971).
9. Garland, L. H., Studies on the accuracy of diagnostic procedures. *Amer. J. Roentgenol.* 82, 25–38 (1959).

10. Conn, H. O., Mitchell, J. R. and Brodoff, M. G., A comparison of the radiologic and esophagoscopy diagnosis of esophageal varices. *N. Engl. J. Med.* 265, 160–167 (1961).

11. Conn, H. O. and Spencer, R. P., Observer error in liver scans. *Gastroenterology* (in Press).

12. Zimmerman, H. J., Chomet, B., Kulesh, M. H. and McWhorter, C. A., Hepatic hemosiderin deposits. Incidence in 558 biopsies from patients with and without intrinsic hepatic disease. *AMA Arch. Intern. Med.* 107, 494–503 (1961).

13. Bell, E. T., The relation of portal cirrhosis in hemochromatosis and to diabetes mellitus. *Diabetes* 4, 435–446 (1955).

14. MacDonald, R. A. and Mallory, G. K., Hemochromatosis and hemosiderosis. *AMA Arch. Intern. Med.* 105, 686–700 (1960).

15. Cachin, M., Pergola, F. and Veron, P., Hemochromatosis in alcoholism. *Sem. Hop. Paris* 38, 1773–1777 (1962).