Liver Function Testing in a Working Population: Three Strategies to Reduce False-Positive Results

Curtis Wright, MD, MPH; Julio C. Rivera, MD; and Joan H. Baetz, RN, MPH

A very high rate of mildly abnormal results on a liver panel of five serum chemistries was observed when these tests were performed on a group of asymptomatic, normal workers. These results often led to lengthy delays in hiring, which benefited neither worker nor employer. Three strategies which markedly reduce the number of false-positive examinations with little or no reduction in test sensitivity are available.

The monitoring of five common serum chemistries (bilirubin, alkaline phosphatase, lactic dehydrogenase, and alanine and aspartate aminotransferases) has become the usual standard of care for the routine screening of asymptomatic workers for occult liver damage or disease. Although effective in some settings, such screening can pose a distinct and difficult clinical problem of false-positive results when low-risk populations are screened. In such cases the occupational physician may not be able to use the same techniques as would be used in interpreting the results of testing in symptomatic, high risk, or chemical industry workers. Tamburro and Liss,[1] in a prior article in this journal, describe how a test for high risk or symptomatic individuals should have a high sensitivity for disease, whereas in the asymptomatic worker a test having a high specificity may be more appropriate.

In either setting, the impact on both worker and employer of medical screening is significant. If the results of screening tests are normal, then the worker begins or resumes his duties. If the results are abnormal, then it is the duty of the physician to evaluate the results of such screening in order to follow up and intervene in cases where test abnormalities may reflect hepatic injury or disease.

In the experience of the authors, if there is no history or physical finding that places that individual in a high-risk group, the first step of such follow-up is to compare the worker's test result with a clinical range of normal values supplied by the vendor of laboratory services. Abnormal values usually lead the physician to request that the worker return to the clinic for repeat evaluation and testing, which often results in lost time, delay in promotion or reassignment, and additional expense to the worker and the employer. This follow-up is worthwhile if it results in additional protection to the health of the worker, but is a legal liability and a simple waste of resources if it does not.

This study is the result of the observation by the authors that 25% to 30% of asymptomatic workers seen at pre-employment and surveillance examinations in our clinic had serum chemistries on a five-test liver panel in excess of the normal range established by our clinical laboratory, despite the fact that they had absolutely no risk factors for liver disease. The study reported here was our response to this situation, and was designed to provide our Occupational Medicine residents with practical strategies to determine how to interpret the screening results of the low-risk worker.

Methods

By an anonymous technique approved by the Human Subjects Committee of the Wyman Park Health System, Baltimore, MD, approximately 250 individuals were sequentially selected from a group of asymptomatic normal workers who were having a routine pre-employ-

From the Division of Occupational Medicine, The Johns Hopkins University, School of Hygiene and Public Health, Baltimore MD (Dr Wright, Clinical Fellow; Dr Rivera, Asst Professor; Ms Baetz, Instructor).

Address correspondence to Dr Wright: Behavioral Pharmacology Research Unit, D-5-West, Francis Scott Key Medical Center, Baltimore, MD 21224.

0096-1736/88/3009-0693$02.00/0

ment or health maintenance examination which routinely included liver function testing. This group was then screened to eliminate workers who were known to be exposed to any hepatotoxic drug or chemical, had any history of liver disease, or had any signs or symptoms of alcoholism, drug dependence, and/or hepatitis at clinical examination. Of the 250 screened, 238 met these criteria and were eligible for the study. The remaining 238 had their blood drawn and processed, and the results were reported by the hospital's clinical laboratory in accordance with the standards of the Joint Commission on Hospital Accreditation, College of American Pathologists,[2] and the State of Maryland laboratory proficiency program.

Evaluation of abnormal values and repeat testing was done by the ordering physician based on his review of each case and was completed prior to any data analysis. No attempt was made to interfere with the usual practice of the clinic physicians, and the laboratory was kept blind as to which patients were included in the study, although both clinic and laboratory were aware that a study was being performed. At the close of the study the charts of all patients who were noted to have abnormal values by the then-current laboratory standards were reviewed, and the actions, diagnoses, and follow-up laboratory values were abstracted and coded for confidentiality. Three records could not be located, and the final number of workers in the group studied was reduced to 235. As will be described below, 14 people were found who had clinical or sample handling abnormalities, leaving us with a group of 221 normal patients, 11 cases of occult conditions, 3 sample handling errors, and 3 missing cases from the original 238 eligible for the study.

Results

Of the 235 workers tested, 64 people had one or more tests which exceeded the upper limits of normal established by our clinical laboratory. Table 1 shows the number of analytes which exceeded this limit and gives the number of workers with abnormal results for each test of the panel. The charts of each person who had a value above the 90th percentile for the group as a whole or who exceeded the clinical limit of normal (whichever was lower) were reviewed to determine what action was taken by the clinician and what diagnosis was reached, and the means by which the diagnosis was made. Of the cases summarized in Table 2, three patients were diagnosed as having alcoholic hepatitis on the basis of a clinical history of excessive alcohol use, other physical findings of alcoholism, and an improvement in serum chemistries following a period of abstention from alcohol. Two patients were found to have failed to report part-time employment which had resulted in significant chemical exposure and probable chemical hepatitis. Four patients were re-evaluated by gastroenterology; one being diagnosed as having viral hepatitis on serologic evidence, and three as having Gilbert's syndrome. One female worker was found to have a high alkaline phosphatase as a result of pregnancy, and one worker had failed to report he was taking a medication expected to cause a mild hepatitis. No patient had a clinical condition which warranted liver biopsy in the opinion of the treating clinician, and the extent to which the workers under-reported self-prescribed medication was unknown. Of this group of 235 asymptomatic low-risk workers, eight had abnormal results on the basis of benign conditions or sample handling problems, and six had significant disease. The 14 workers who were thus found to have occult disease, normal variations, or sample handling abnormalities were excluded from the group used to define the normal ranges which was reduced to a total of 221 normal persons.

TABLE 1
Analytes from the Entire Group Which Exceeded the Laboratory Range of Normal

Test Performed	Laboratory Range	Abnormal Test Results
Bilirubin	0–1.1 mg/dL	19/235 (8.0%)
Alkaline phosphatase	28–98 U/L	14/235 (5.9%)
Lactic dehydrogenase	102–228 U/L	32/235 (13.6%)
Aspartate aminotransferase	3–33 U/L	10/235 (4.2%)
Alanine aminotransferase	3–30 U/L	17/235 (7.2%)
Total number of analytes abnormal (five tests per individual)		92/1,175 (7.8%)
Persons with one or more abnormal tests		64/235 (27.23%)

TABLE 2
Diagnoses Found on Evaluation of the Study Group

Diagnosis	No. (N = 235)
Alcoholic hepatitis	3
Hepatitis attributed to occult chemical exposure	2
Acute viral hepatitis	1
Gilbert's syndrome	3
Side effect of medication	1
Known pregnancy	1
Sample hemolysis	3

The results shown in Table 1 suggest that the laboratory had achieved an upper limit of normal very close to the 95th percentile (95% of individuals in the group test below this level). To examine the distribution of the abnormal test results, a set of hypothetical expected frequences was calculated from the assumptions that there were 221 normal patients and a 5% false-positive rate, and that those elevated values were due to chance alone. This was done by assuming a binomial distribution, statistical independence, a mean false-positive rate of 5%, and the binomial distribution.[3 (pp 116–117)] The results are shown in Table 3 along with the actual observed frequencies.

Fig. 1 shows the distributions of the values for the five tests in our final series of 221 normal patients. The upper limits of normal were determined by the nonparametric method of Herrera,[4] which consists of ranking the observed values from smallest to largest and picking the desired percentiles directly from a rank-ordered list. Table 4 gives the relevant percentiles of our sample.

TABLE 3
Predictive Value of Abnormal Test Results*

	Frequency (N = 235)		Percentage with Clinical Illness
	Expected	Observed	
Patients with one abnormal test	48	50	0/50 (0%)
Patients with two abnormal tests	5	9	4/9 (45%)
Patients with three abnormal tests	0	4	3/4 (75%)
Patients with four abnormal tests	0	1	1/1 (100%)
Patients with five abnormal tests	0	0	Not defined

* At the prevalence of abnormality found in this study.

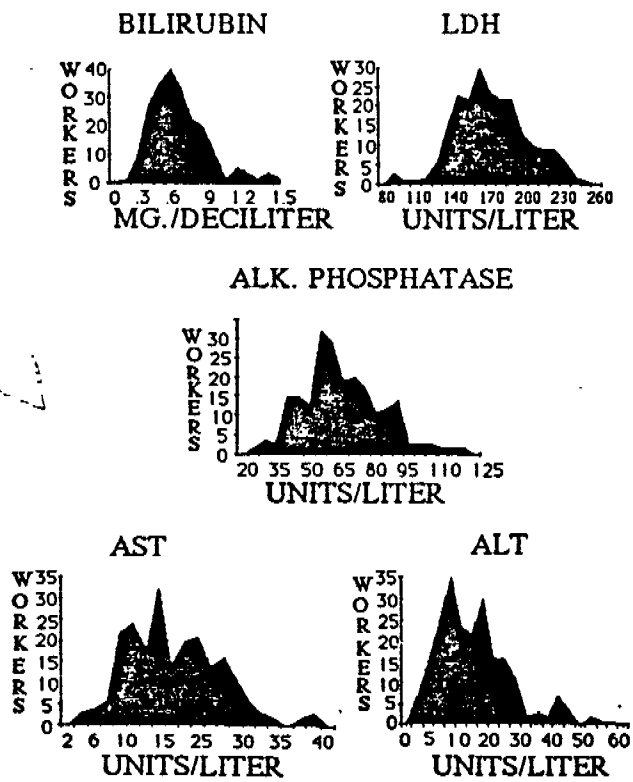

Fig. 1. Frequency polygons for the results of five common serum chemistries performed on a sample of 221 healthy workers. *LDH*, lactic dehydrogenase; *AST*, aspartate aminotransferase; *ALT*, alanine aminotransferase.

Discussion

The most significant result of this study is the fact that we were repeating an excessive number of falsely positive examinations. The results shown in Table 1 indicate that the upper limit of the clinical range of normal given by the laboratory was exceeded in one or more tests in 64 of 235 people tested. Of these 64 people, only six had conditions of medical significance. Our objective was to to devise ways to avoid having to re-evaluate 25% of the work force to find the 2.5% who had disease which required intervention.

Articles in the clinical literature by Dixon and Lazlo[5] and Parkerson and Eisenson[6] suggested that there was wide variation among clinicians as to what strategy they used to determine which tests to follow up. We believed that it was essential that some of these strategies be made explicit in our training program. We thus searched the literature further to select several good strategies.

TABLE 4
The 90th through 99th Percentiles in 221 Normal Workers

Analysis	Laboratory Upper Limit	Percentile			
		90th	95th	97.5th	99th
Bilirubin (mg/dL)	1.1	1.0	1.2	1.4	1.6
Alkaline phosphatase (IU/L)	98	90	100	110	118
Lactic dehydrogenase (IU/L)	228	232	243	253	258
Aspartate aminotransferase (IU/L)	33	27	30	33	39
Alanine aminotransferase (IU/L)	30	31	42	46	53

Strategy 1—Establish Normal Limits for Your Own Clinic

Our first strategy was most strongly advocated by Hoffmann in 1963[7] and consisted of making our own range of normal for patients drawn from our clinical population. When we inquired of our pathologist, we found that the clinical laboratory reviews its range of normal whenever it purchases a new analytical system or at periodic intervals as a quality assurance practice. It does so by selecting samples from the patient samples submitted, making a graph of the distribution, and selecting an upper limit somewhere between the 85th and 99th percentiles (the suggestions of the makers of the equipment, published ranges from the literature, and the clinical judgment of the responsible pathologist).

The laboratory does not usually have access to either the patient or the chart, and can only guess that it has selected a representative sample of the population served. In our case the laboratory had done an excellent job, and had set the upper limit of normal between the 90th and the 95th percentiles for all but ne f the elements of the panel. In many cases the laboratory does not have access to a good sample of the population, and may set the standard either too high or too low for use in a cohort of healthy workers.

Any clinic can collect several hundred samples from healthy, asymptomatic workers who have no known exposure to hepatotoxins and create their own range of normal by the methods described above. In an excellent pair of articles, Reed et al[8] discuss this percentiles method and suggest that a sample size of 150 to 250 will give adequate accuracy, whereas Mainland[9] strongly advises against attempts to establish clinical ranges by use of means, standard deviations, normal approximations, or log-normal transformations.

In our case we found the laboratory's upper limit of normal was set slightly too low to reflect the 95th percentiles of our series of workers, and the slightly broader 95th percentile limits of Tabl 4 did reduce the burden of false-positive results from 58 to 41 in our

TABLE 5
True Cases, True Negative Results, False-Positive Results, and False-Negative Results Found by Each Strategy

Strategy	Results			
	True Negative	True Positive	False-Negative	False-Positive
Use laboratory normals	171	6	0	58
Strategy 1 Use clinic normals at 95th percentiles	188	6	0	41
Strategy 2 Use clinic normals at 99th percentiles	217	6	0	12
Strategy 3 Use laboratory normals, and accept one elevated value (if clinically reasonable)	221	6	0	8

series. As can be seen from Table 5, use of this first strategy would not result in our failing to detect any of our cases of illness.

Strategy 2—Use a Wider Range if Doing Multiple Tests

Our second strategy was based on the well-known problem of performing multiple tests on the same person. In doing a group of five tests in combination on a given worker we have increased the chances that that person's test results will be false positive. The more tests performed on one person, the more likely that one or more will have abnormal test results. If the upper limit of normal is set at the 95th percentile, the chance of a false-positive result on the second test will be 5%, and so on for as many tests as you wish to do.

To be "normal" (have no abnormal test results), the person has to avoid a false-positive result on each test, so that the percentage of the population remaining as each test is added will be 0.95 to the Nth power, where N is the number of tests. In the case of a five test panel, the theoretical likelihood of one or more falsely abnormal tests was calculated at $1.00 - (0.95 \times 0.95 \times 0.95 \times 0.95 \times 0.95)$ or 22.6%[3 (pp 116–117)] and we verified that our observed result was similar at 56/221 or 24%.

The simplest way to correct for this factor is to set the normal range for each test so that the five test panel *as a whole* has whatever false-positive rate you desire. This can be done using the Bonferroni inequalities[3 (pp 116–117)] and allows you set an upper limit for each test which will give a false-positive rate of 5% of all of the tests combined. For five tests this means that if the upper limit of normal is set at the 99th percentile for each test, the overall false-positive rate will be 5% for the five tests as a group.

This was done in our sample and we found the false-positive rate using the 99th percentiles to be 12/221 or 5% as predicted. Our second strategy, as shown in Table 5, was to set the upper limits of normal for each single test at the 99th percentile, so that the upper limit for the whole set of tests was set at the 95th percentile.

Strategy 3—*Knowingly* Interpret the Test Results

Our third and simplest strategy was to utilize the biologic patterns of the known types of hepatic injury. Zimmerman et al[10] classifies the major types of hepatic injury as either hepatocellular or cholestatic, and offers a schema for interpretation of enzyme patterns which was nicely demonstrated by Ferraris et al.[11] This scheme, which is outlined in Fig. 2, shows lactic dehydrogenase and γ-glutamyltranspeptidase as nonspecific indicators of general hepatic insult (lactic dehydrogenase being elevated from many nonhepatic sources), alanine and aspartate aminotransferases as being elevated preferentially in hepatocellular diseas , and alkaline phosphatase and bilirubin being elevated preferentially in cholestatic conditions. Alanine and aspartate aminotransferases are thus "paired" tests, as are bilirubin and alkaline phosphatase, and would be expected to rise and fall in concert.

In contrast to this systematic variation of enzyme patterns in disease, false-positive results due to analytic variation would be expected to occur in a random fashion. Table 3 shows the predicted number of abnormal tests to be expected if laboratory variation is randomly distributed and compares that prediction and the actual values for our sample. In addition, the last column of the table gives the prior predictive value of disease in this sample if one or more tests are found to be abnormal. It confirms what was expected: that a high value on one test due to an analytic variation does not generally result in a significant increase in the chance of elevation of other tests run on the same sample.

Most hepatic disease processes would not be expected to elevate just one value in the panel, and it would be very unusual for any of the most common forms of liver injury (alcohol, hepatitis, chemical injury, or biliary disease) to produce a pattern of four low values and one high one. It is thus possible to check a modest elevation of alanine aminotransferase against the aspartate aminotransferase value, a high bilirubin against the alkaline phosphatase, and a high lactic dehydrogenase against the rest of the panel.

As shown in Table 3, none of the people who were proven to actually have liver disease had an isolated elevation of only one value in our sample. Our simplest strategy for reducing the number of false-positive results would be to discount a moderate elevation in one *isolated* value when the other four tests were well within normal limits. This is the most effective strategy in our population, giving only eight false-positive results in 221 workers screened, while missing no cases. We cannot claim credit for this strategy, since Parkerson and Eisenson[6] showed that th single greatest predictor of

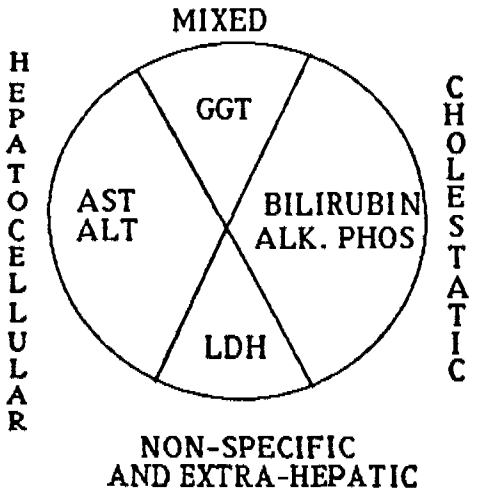

Fig. 2. Patterns of serum chemistry elevations by types of hepatic injury. *GGT*, γ-glutamyltranspeptidase. Other abbreviations are as in Fig. 1.

whether or not a physician would follow up a screening panel was the number of abnormal tests.

No powerful and effective tool is without risk, and this last technique requires that the clinician thoughtfully review the differential diagnosis of each isolated elevated value in terms of the entire clinical picture. Once this is done, a modest elevation in a paired enzyme with *no corresponding elevation* in the other member of the pair need not be automatically repeated if, in the *opinion of the responsible physician*, there is no reason to suspect a nonhepatic source of the elevation.

Conclusion

We began this investigation with profound skepticism of the accuracy of the clinical laboratory, yet we had faith in the range of normal provided by that same laboratory. In fact, neither attitude is warranted. Effective use of the laboratory demands that the responsible clinician understand the process of establishing a normal range, just as the pathologist and the laboratory technologist must understand the uses to which their results are put.

In an era in which laboratory testing is used as a profit-making service in outpatient clinical settings, it is as much an abuse of trust to over-test as to under-test in screening the asymptomatic population. It is the physician's ethical obligation to order the *right test* at the *right time* for the *right person*, and to *rightly* interpret the results.

It was our experience that the clinical range of normal appropriate for a hospital was not specific enough for use in screening an asymptomatic, low-risk working population, and we have presented three strategies for reducing the number of false-positive results without serious risk of missing true cases. It is not the intent of this paper to establish the normal range for any population, to claim any of these strategies as our invention, or to prove the sensitivity or specificity of our strategies for any specific disease. Our intent is to exhort the physician to examine his or her use of the "normal values" and to encourage the explicit use of reasoned strategies which make best use of these powerful tests.

The best of our three strategies requires no data collection and little equipment, nd asks only that the clinician carefully inspect the liver function panel and accept a single isolated value as being due to analytic variation if it does not correlate with any reasonable pattern of hepatic injury or disease. We wish to explicitly warn the reader that these results are intended for use in the routine screening of asymptomatic *clinically healthy* low-risk worker populations. If the population being screened has clinical symptoms of hepatic disease, works with a known hepatotoxin, or is at high risk for hepatitis, interpreting their laboratory tests on the basis of a "normal range" is not appropriate. In such cases the upper limit of "normal" should be a cutoff value selected to give the best discrimination between the sick and the well, determined by the methods described in the very readable text by Sacket et al[12] on the interpretation of such data.

It is the intent of this report to provide the clinician with the reassurance of our experience that careful clinical interpretation is still the most powerful tool for evaluating the significance of any test result.

Acknowledgments

The authors would like to express their grateful appreciation to the staff of the Occupational Medicine Clinic, the clinical chemistry laboratory, and all of the staff of the Wyman Park Health System, without whose faithful and careful service this study would have been impossible.

References

1. Tamburro C, Lisa G: Tests for hepatotoxicity: Usefulness in screening workers. *J Occup Med* 1986;28:1034–1044.

2. Elevitch FR, Noce PS: *Data ReCAP-1980*, Skokie, IL, College of American Pathologists, 1981, pp 1–21.

3. Snedecor GW, Cochran WG: *Statistical Methods*, ed 7, Ames, IA, Iowa University Press 1980, pp 116–117.

4. Herrera L: The precision of percentiles in establishing normal limits in medicine. *J Lab Clin Med* 1958;52:34–43.

5. Dixon RH, Lazio J: Utilization of clinical chemistry services by medical house staff. *Arch Int Med* 1974;134:1064–1067.

6. Parkerson GR, Eisenson HJ: Association of patient and physician characteristics with follow-up of abnormal laboratory results. *J Fam Pract* 1980;11:943–948.

7. Hoffmann RG: Statistics in the practice of medicine. *JAMA* 1963;185:864–873.

8. Reed AJ, Henry RJ, Mason WB: Influence of statistical method used on the resulting estimate of the normal range. *Clin Chem* 1971;17:275–284.

9. Mainland D: Remarks on clinical norms. *Clin Chem* 1971;17:267–274.

10. Zimmerman HJ, Seef LB: Enzymes in hepatic disease, in Coodley EL (ed): *Diagnostic Enzymology*. Philadelphia, Lea & Febiger, 1970, pp 1–18.

11. Ferraris R, Colombatti G, Fiorentini MT, et al: Diagnostic value of serum bile acids and routine liver function tests in hepatobiliary diseases. *Dig Dis Sci* 1983;28:129–136.

12. Sacket DL, Haynes RB, Tugwell P: *Clinical Epidemiology, A Basic Science for Clinical Medicine*. Boston, Little Brown & Co. 1985, pp 59–136.