Determinants of Physician Recognition and Follow-Up of Abnormal Laboratory Values

George R. Parkerson, Jr, MD, MPH
Durham, North Carolina

Potential determinants of physician recognition and follow-up of abnormal laboratory values were studied in the ambulatory primary care setting. Data support the hypothesis that a significant positive association exists between the clinical importance of a laboratory result and physician response. Clinical importance was indicated by degree of abnormality of the laboratory value, the type of test, and the indication for obtaining the test. Response was not significantly associated with type of laboratory report or resident's year of training, but a relationship was shown with resident's National Board scores. A model of 12 laboratory tests was found to be more appropriate for studying recognition and follow-up than one of 30 because of fewer repetitious tests and fewer results of doubtful clinical usefulness. With such a select model, recognition and follow-up of abnormals can be used as process measures of quality of medical care.

Clinical medicine is a mixture of science and management expertise, the quality of which is difficult to measure. Yet it becomes increasingly important to identify methods of measurement as society struggles to provide adequate medical care for all people. Physician recognition and follow-up of abnormal laboratory values have been used as such measures. In past studies, the recognition and follow-up rates have been uniformly low,[1-5] implying poor quality of care.

The present study investigates possible determinants of physician recognition and follow-up, seeking to explain the overall low rates and determine the extent to which quality of care is measured. The hypothesis is that physician recognition and follow-up of abnormal laboratory values is primarily determined by, and directly related to, the clinical importance of the abnormality. The term "clinical importance" reflects the quasi-scientific nature of clinical medicine because its definition varies from patient to patient, disease to disease, and physician to physician. In spite of the difficulties with definition, there are some indicators of clinical importance that can be measured, namely: (1) the indication for obtaining the test, (2) the type of test, and (3) the degree of abnormality of the laboratory value. If the indication for the test is sound, if the test is appropriate to the clinical problem and reliable as a diagnostic aid, and if the degree of abnormality of the resulting value is great enough to be free of technical or individual variation, then the physician is very likely to recognize the abnormality and follow it up. Under such circumstances, if he/she fails to do so, the likelihood of poor quality of care becomes a possibility and further inquiry is indicated.

From the Department of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina. Requests for reprints should be addressed to Dr. George R. Parkerson, Department of Community and Family Medicine, Duke University Medical Center, Durham, NC 27710.

0094-3509/78/0801-0341$02.25

Methods

Study Design

The study was conducted in the Duke-Watts Family Medicine Center during the first four months of 1976, at which time there were 6,196 registered patients and 7,066 patient encounters. Of these, only the adult patients (age 18 years or older) of resident physicians were included.

The study patients were 64.8 percent female and 35.2 percent male, 73.7 percent white and 26.3 percent nonwhite, 31.6 percent aged 18 to 39 years, 35.9 percent aged 40 to 59 years, and 32.4 percent aged 60 years and older. This compares with the entire practice population, in which patients were 64.4 percent female and 35.6 percent male, 81.9 percent white and 18.1 percent nonwhite. Considering only the adults (excluding the 21 percent below age 20), the age distribution for the entire practice population was 56.4 percent aged 20 to 39 years, 23.5 percent aged 40 to 59 years, and 20.1 percent aged 60 years and older.

Laboratory results were entered into the computer through a terminal in the center and were converted into special computer printout laboratory reports, approximately half of which showed normal ranges following the conventional laboratory values, and half, standardized units following the conventional values.

The standardized units were those of Labstand,[6] a system which translates laboratory values into standard units (su) on a scale of 0 to 100 su, uniform for all tests, with the goal that brief inspection of any laboratory report will reveal immediately whether the result is normal or abnormal, high or low, and to what degree. This process requires of the reader no knowledge of the normal or abnormal ranges for specific laboratory tests.

Results on 30 frequently requested blood tests were studied. These included 23 chemistries: alanine amino transferase (SGPT), asparate amino transferase (SGOT), direct bilirubin, indirect bilirubin, total bilirubin, calcium, carbon dioxide content, chloride, cholesterol, creatinine, glucose, lactate dehydrogenase (LDH), phosphatase alkaline, phosphorus, potassium, sodium, total protein, albumin, globulin, albumin to globulin (A/G) ratio, triglycerides, urea nitrogen (BUN), and uric acid. Also there were seven hematology tests: erythrocyte count (RBC), hematocrit, hemoglobin, leukocyte count (WBC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC). These were chosen because they were the major constituents of laboratory panels ordered by the residents.

A test result was considered to be abnormal if its value lay outside the normal range published by the commercial laboratory which performed all tests in the study (National Health Laboratories, Winston-Salem, NC).

Physicians in the study were 27 family medicine residents, 12 of whom were in their first year of training, 11 in their second, and four in their third. Of these, 25 were recent graduates of 20 different US medical schools and two were foreign medical graduates. They were stratified according to year of training and randomly divided into two groups. While one group received laboratory reports showing normal ranges, the other received Labstand reports. Each group used each type of report for half of the study period. The residents were aware that a study was being conducted but they had only minimum details.

The computer identified all patients with abnormal laboratory results, and the remaining data for the study were collected from their medical records. This was facilitated by the problem-oriented system in use by all residents. All medical record reviews were conducted by the principal investigator. Most were done one to three months after the date of the laboratory test, with a minimum of one month, a maximum of six months, and a mean of 63 days.

Recognition and follow-up of abnormal laboratory values were used as outcome variables. Criteria were as follows:

A. Both Recognition and Follow-Up
   1. Repeat of the laboratory test, or
   2. Change in diagnostic and/or therapeutic plans as a result of the laboratory result, or
   3. Explanation in the progress notes as to why such steps (as 1 and 2 above) were not taken.

B. Recognition, But No Follow-Up
   1. Notation in the progress notes or on the problem list that the result was abnormal, but then no follow-up, or
   2. Circling or otherwise specially marking the abnormal result on the laboratory sheet, but then no follow-up.

C. Neither Recognized Nor Followed Up
   No evidence in the medical record of acknowl-

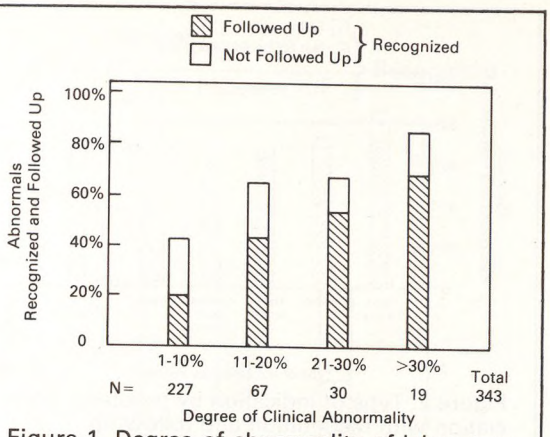

Figure 1. Degree of abnormality of laboratory values: Association with recognition and follow-up.

edgement of the abnormality, according to the criteria for A and B above.

For the analysis, recognition was defined as acknowledgement by the physician on the medical record that the test result was abnormal. Therefore, it included both those values recognized and followed up, and those recognized and not followed up (categories A and B combined). Follow-up included only those abnormals both recognized and followed up (category A). The two dependent variables were not intended to be mutually exclusive and were analyzed separately.

Determinants

Six factors were studied as potential determinants of physician response to abnormal laboratory results. The first three listed below were considered the principal study variables because they measure clinical importance of abnormal values. The others were primarily used as controlling variables in the analyses.

1. Degree of Abnormality of the Laboratory Value

Degree of abnormality of each abnormal result was determined from the Labstand standard unit value for that result. This is logical because the standard unit scale is based on degree of abnormality, where 0 to 39.99 su represents the clinical low abnormal range, 40 to 59.99 su, the normal range, and 60 to 100 su, the high abnormal range. Maximum abnormality is determined from medical literature review appropriate for each type of laboratory test and is indicated by 0 su and 100 su. For example, a laboratory result of 64 su would be considered ten percent of maximum clinical abnormality for a high abnormal value, since it is located at a point one tenth of the range from 60 su to 100 su. Likewise, 36 su would be ten percent of maximum for a low abnormal value.

2. Type of Laboratory Test

The full data consisted of all 30 types of laboratory tests previously listed. A smaller model of 12 of these was used for most analyses and will be described in the Results section.

3. Type of Indication

From review of medical records of all patients with abnormal laboratory values, the clinical problems listed in the progress note assessment when the laboratory test was obtained were used as indications. The classification rubrics are those of the International Classification of Health Problems in Primary Care (ICHPPC).[7] Some analyses used the individual problems while others used a classification into two groups, ie, those indicated primarily for health maintenance, and those primarily for medical problems.

4. Type of Laboratory Report

The conventional normal range laboratory report and the experimental Labstand standard unit report were compared to observe the effect on recognition and follow-up rates. The type of report for each laboratory value depended upon which

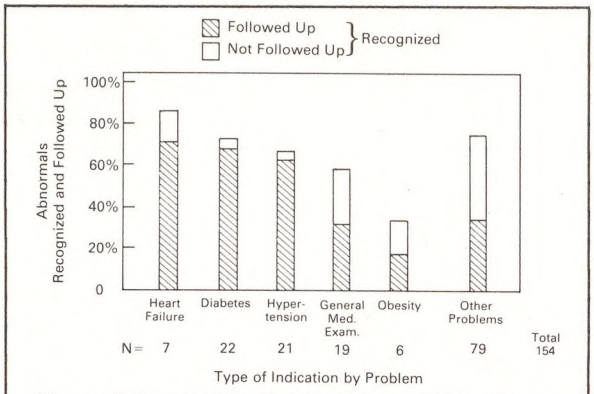

Figure 2. Type of indication by problem: Association with recognition and follow-up (Includes only abnormal values of high degree of abnormality, ie, > 10% of clinical maximum).

type the resident's group was programed to receive at the time the laboratory test was obtained.

5. Resident's Year of Training

One class was used for each of the three years of residency training.

6. Resident's Board Score

Twenty-three of the 27 residents took Part II of the Internal Medicine National Board Examination. When their performance is compared with that of all 4,300 Board candidates in September 1975, their mean score is found to be higher than the national average (500 vs 489), the variance is smaller (standard deviation of 84 vs 94), and the range is not as wide (360 to 670 vs 240 to 710). In spite of these differences, there is a remarkable similarity, indicating that the residents' scores are to a high degree representative of the entire candidate population.

For analysis in this study, the residents were divided into a high score group (12 with scores above the mean for the entire group) and a low score group (11 with scores below the mean).

Statistical Methods

All statistical analyses in this study were done using the chi-square statistic.

Results

Descriptive Data

There were a total of 6,635 laboratory values on 392 different patients, an average of 16.9 values per patient. Of these, 5,627 (84.8 percent) were normal and 1,008 (15.2 percent) were abnormal. Patients having all normal results numbered 96, while there were 296 having at least one abnormal. This study observes only the 1,008 abnormal values on 296 patients, an average of 3.4 abnormals per patient.

The original data consisted of 1,412 values labeled abnormal, from which 404 were excluded. Of these, 197 resulted from a change in normal limits by the laboratory, which went unrecognized initially by the investigator. These alterations accompanied change to a different type of auto-analyzer. Another group of 56 exclusions resulted from faulty auto-analyzer operation. These inaccurate determinations involved serum albumin, which in turn caused misleading results in globulins and A/G ratios. The trouble was discovered as a result of the study and the equipment defect was corrected promptly by the laboratory. Repeat tests on the same patient accounted for 83 exclusions, missing data for 31, and various other causes for the remaining 37.

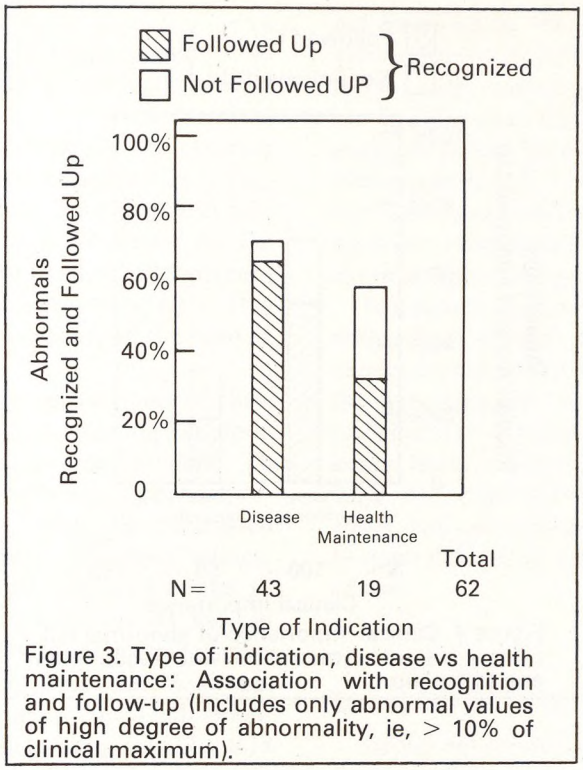

Figure 3. Type of indication, disease vs health maintenance: Association with recognition and follow-up (Includes only abnormal values of high degree of abnormality, ie, > 10% of clinical maximum).

Selection of a Model for Analysis

In addition to the large number of values that unquestionably warranted exclusion, there were others whose recognition and follow-up would hardly be useful as a measure of quality of care. In many instances purposeful lack of follow-up would be more indicative of good care than follow-up. An example of this was the impossible situation presented by frequencies of abnormal bilirubin values, which showed 151 abnormal direct bilirubins and 29 abnormal indirect bilirubins, but only eight abnormal total bilirubins, on the same group of data. Further investigation revealed that all but a few of these were below normal, ie, in the abnormal low range, rather than high. This explained how the relationships were possible mathematically, but from the clinical standpoint none of these values were valid indicators of abnormality. Some laboratory tests were of extremely low clinical worth, such as the A/G ratio, which Davidsohn[8] recommends abandoning. Other tests were for most purposes repetitious, such as hemoglobin and hematocrit, and BUN and creatinine.

For these reasons it was decided to formulate a model for analysis consisting of tests selected by the following criteria:

1. Each test is not clinically repetitious of another one in the model.
2. In general, the test results are likely to be important clinically.

The select model included the following tests: SGOT, total bilirubin, calcium, cholesterol, creatinine, glucose, potassium, sodium, triglycerides, uric acid, hematocrit, and WBC. These 12 tests produced 343 abnormals from 212 different patients, whereas the full data had contained 1,008 abnormals on 30 tests from 296 patients. Abnormal hematocrits were the most frequent (21.6 percent), followed by glucose (18.1 percent), and WBC (12.2 percent). The remaining tests each constituted less than ten percent of the total.

Study of Determinants

Significant differences in both recognition and follow-up rates were found for different degrees of abnormality, types of laboratory tests, and types of indication. In contrast, no significant rate

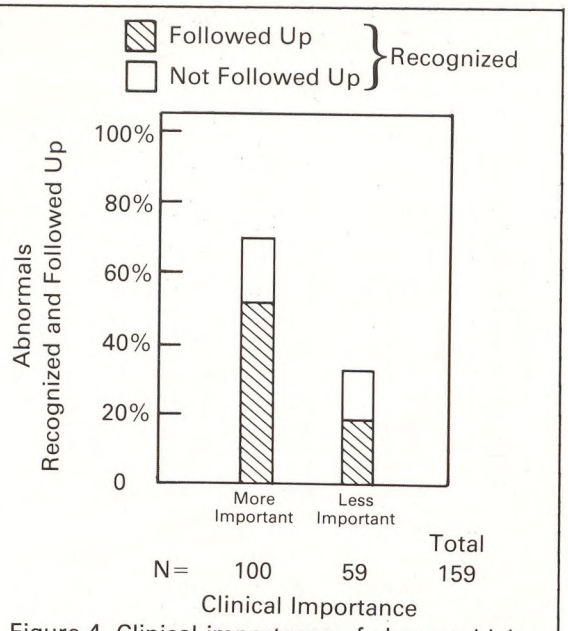

Figure 4. Clinical importance of abnormal laboratory values: Association with recognition and follow-up.

differences were identified for different laboratory report formats or resident's year of training. Residents with higher than average Board scores had higher follow-up rates than those with lower scores, but when degree of abnormality was controlled for, this higher rate was found to concern only those values of low degree abnormality.

The association of degree of abnormality with recognition and follow-up is shown in Figure 1. As the degree of abnormality increased, both rates increased proportionately. Thus, recognition increased from 41.9 percent for the lowest degree of abnormality to 84.2 percent for the highest (χ^2=23.58; df=3; P<.0005), while follow-up increased from 20.3 to 68.4 percent (χ^2=36.77; df=3; P<.0005). When controlling for all the other five determinants, this relationship persisted.

Among different types of laboratory tests, recognition rates varied from 100 percent for creatinine and 80 percent for total bilirubin to 27 percent for sodium and triglycerides. Follow-up rates varied from 80 percent for creatinine to 6 percent for triglycerides. When only the more abnormal values (greater than ten percent of maximum clinical abnormality) were considered, marked changes occurred in the individual rates. For example, recognition rate for glucose increased from 53.2 percent for all abnormals to 93.3 percent for those abnormal values greater than ten percent, and the follow-up rate, from 33.9 to 73.3 percent. Similarly, hematocrit recognition rate increased from 58.1 to 73.1 percent, and follow-up rate from 40.5 to 69.2 percent.

Recognition and follow-up also differed according to the indication for obtaining the test. Figure 2 illustrates these differences among the various indication problems, when only the more abnormal values were analyzed. The potentially urgent and frequently encountered disease categories, ie, heart failure, diabetes, and hypertension, had both high recognition and follow-up rates, while the disorder that is less urgent and more difficult to manage, ie, obesity, had low rates. Recognition for heart failure was 85.7 percent and follow-up was 71.4 percent. For diabetes, the rates were 72.7 and 68.2 percent, and for hypertension, 66.7 and 61.9 percent. In contrast, for obesity, recognition was 33.3 percent and follow-up, 16.7 percent. General medical examination and the miscellaneous problem group occupied an intermediate position, having relatively high recognition rates (57.9 and 74.7 percent, respectively), but low follow-up

rates (31.7 and 32.9 percent).

If diabetes and hypertension are combined in the analysis to represent serious medical problems and then compared with general medical examination to represent health maintenance, the contrast in physician response can be illustrated as in Figure 3. Both recognition rate (69.8 percent) and follow-up rate (65.1 percent) were higher for the diseases than for health maintenance (57.9 percent recognition and 31.6 percent follow-up). This difference is statistically significant in the case of follow-up ($.0005<P<.005$).

When the effects of all three measures of clinical importance are combined, the impact upon recognition and follow-up is more evident. To illustrate this, a model representing "more important" abnormal laboratory values is contrasted with one of "less important" abnormal values. The "more important" group consists of the 12 selected types of laboratory tests already used in the other analyses, but only those having abnormal values greater than ten percent of maximum abnormality and obtained for a medical indication. The "less important" group is composed of the 18 types of tests from the original 30 which were not included in the select analysis group, and those having abnormal values of only one to ten percent of maximum abnormality, and those obtained purely for health maintenance indications. The results of this analysis are shown in Figure 4. For the "more important" group, recognition rate (70.0 percent) and follow-up rate (53.0 percent) are more than double those for the "less important" group (33.9 and 18.6 percent, with $P<.0005$ for both comparisons).

Analysis of Abnormal Values Not Followed Up

Further understanding of physician response to abnormal results can be gained by close examination of those specific abnormal values that were not followed up.

Of the 343 abnormal values, a total of 239 (69.7 percent) were not followed up. Only 58 of these (16.9 percent of all abnormals) were of higher degrees of abnormality, ie, greater than ten percent of maximum clinical abnormality. Table 1 shows some of the characteristics of these values. Cholesterol had the highest individual percentage not followed up (41.7 percent). All ten of these values were in the low abnormal range. This seems to be a large number of low cholesterols, indicating in retrospect that the operation of the autoanalyzer should have been checked or the normal limits questioned. Likewise, the large number of low abnormal WBCs appears unusual and raises questions concerning transport of blood specimens as well as function of laboratory equipment.

In a patient population like the one under study, with a high number of hypertensives, one might expect a relationship between thiazide therapy and the eight high uric acids and five low potassiums, for which the clinician usually accepts greater variation than in untreated patients. On the other hand, the eight low hematocrits, three low glucoses, and one high glucose of 206 mg/100 ml may be more difficult to explain when defending quality of care.

Detection of New Problems

Of the 343 abnormal laboratory values, 38 (11.1 percent) led to new, previously unsuspected, patient problems. Anemia was the most frequent new problem (15 cases). There were three cases of gout, one of diabetes mellitus, and four of lipid disorders. The remaining were less important or incompletely established conditions. Only eight (21.1 percent) of these newly detected problems resulted from tests ordered solely for health maintenance indications. However, these included the one case of diabetes, one of the lipid disorders, and three of the anemias.

Discussion

This study suggests that if recognition and follow-up of abnormal laboratory results are to be considered as measures of quality of medical care, it is essential to be highly selective in the choice of laboratory tests used as indicators and to use definitions of abnormality that reflect true clinical abnormality, insofar as possible. To include all tests and all values branded as abnormal according to present day unrealistic normal ranges may appear to be a more comprehensive approach, but it may reflect inappropriate application of laboratory technology in the clinical sphere more than it measures physician performance. An alternative

Table 1. Characteristics of Abnormal Laboratory Values Not Followed Up (Includes Only Abnormal Values of High Degree of Abnormality, ie,>10% of Clinical Maximum)

Name of Laboratory Test	Low Abnormal Range		High Abnormal Range		Total Values Not Followed Up	Total All Abnormal Values for Laboratory Test	Percentage Not Followed Up
	Laboratory Values	Count	Laboratory Values	Count			
1. Cholesterol (mg/100 ml)	23,121,122,125,126, 128,129,132,134,134	10	0	0	10	24	41.7
2. Uric Acid (mg/100 ml)	1.4,1.6,2.2	3	9.4,9.4,9.6,9.9,10.0, 10.1,10.2,11.3	8	11	31	35.5
3. WBC ($\times 10^3/\mu l$)	3.1,3.3,3.4,3.4,3.5, 3.8,3.8,3.9,4.0,4.0, 4.2,4.2,4.3	13	0	0	13	42	31.0
4. Sodium (mEq/liter)	131,132,132	3	150,150	2	5	22	22.7
5. Potassium (mEq/liter)	3.1,3.1,3.2,3.3,3.3	5	0	0	5	26	19.2
6. Calcium (mg/100 ml)	7.8	1	0	0	1	8	12.5
7. Hematocrit (%)	30.1,30.6,32.1,32.7, 33.5,34.4	8	0	0	8	74	10.8
8. Glucose (mg/100 ml)	46,56,59	3	207	1	4	62	6.5
9. Triglycerides (mg/100 ml)	27	1	0	0	1	33	3.0
10. SGOT (U/ml)	*	0	0	0	0	11	0.0
11. Total Bilirubin (mg/100 ml)	*	0	0	0	0	5	0.0
12. Creatinine (mg/100 ml)	*	0	0	0	0	5	0.0
Totals		47 (81.0%)		11 (19.0%)	58	343	16.9

* No abnormal range

strategy would be to include all test types and all abnormal values but distinguish between clinically appropriate and inappropriate physician responses. For example, follow-up of an abnormal low serum glucose might be considered correct, while follow-up of abnormal low BUN, incorrect. It appears from the present data that the resident's decision not to follow-up certain abnormalities often represented high rather than low quality of care.

The findings support the hypothesis that recognition and follow-up of abnormals is directly related to the clinical importance of the laboratory result, as defined by degree of abnormality, type of test, and type of indication. This means that a laboratory value of an appreciable degree of abnormality, such as greater than ten percent of maximum, from a test of high clinical relevance such as glucose or hematocrit, obtained because of an established clinical problem such as diabetes mellitus or hypertension, can be expected to have both a high recognition and a high follow-up rate. On the other hand, an abnormal value of low degree or questionable abnormality, on a test of uncommon or debatable relevance, such as abnormally low triglycerides, obtained from a patient either entirely without medical problems, or because of an enigmatic problem such as obesity, can be predicted to have low recognition and follow-up rates.

Some explanations are suggested for physicians' behavior in response to three of the dilem-

mas they confront in ambulatory primary care. One is how to decide which laboratory values are really abnormal and which result from inappropriate normal ranges or laboratory and clerical inaccuracies. Residents appear to cope with this dilemma with the strategy: pay little attention to most abnormal values of low degree of abnormality. In effect, they establish their own unwritten sets of clinical normal ranges, upon which they rely more than upon the laboratory normal ranges.

The data support pleas of other investigators for revision of the present system of normal ranges.[9-13] Since abnormality can be more accurately defined than normality, data on people with definite disease may eventually constitute the basis for precise abnormal range distribution curves. Then normality can be defined in terms of the statistical likelihood that a laboratory value is *not abnormal*, rather than whether or not it is normal. With such an approach, there will no longer be a need for normal ranges.

Another dilemma is how to selectively utilize the large number of different laboratory tests available. Here residents seem to respond by purposely ignoring results of those types of tests which are unlikely to benefit their patients. The paradox is that they continue to order those same tests, apparently because of the widespread custom of ordering laboratory tests by multiple-test panels, even when only one or two tests are clinically indicated. The usual rationalization for this strategy is that, because of modern technology, it is just as cheap for the patient to have the panel done as it is to run a single test.

Only tests ordered by panels were included in this study. Interestingly, none of the panels resemble the model of 12 tests chosen for most analyses. While it is not clear what criteria are used to formulate panels and who establishes those criteria, it is apparent that present panels are not appropriate for the ambulatory primary care clinical setting.

A third dilemma is how to choose laboratory tests for screening purposes only and what to do when unexpected abnormal values are returned for persons who are not sick. The physician behavior here seems to be: get multiple tests with the idea of providing comprehensive screening, hope that all results will be normal, and then if abnormal values do appear, pay them less attention than if those same values were from sick people. This is not logical and indicates that more realistic guidelines for selecting health maintenance laboratory tests are greatly needed.

The following recommendations are made for change in laboratory utilization by primary care facilities:

1. Stop routine use of panels when only specific tests are needed.
2. Formulate new panels based on criteria appropriate for ambulatory care.
3. Require a clinically defensible indication for each laboratory test or panel requested.
4. Routinely monitor follow-up of abnormal laboratory values, particularly those of higher degrees of abnormality.

Acknowledgements

The author is indebted to Dr. Barbara Hulka, Dr. Michel Ibrahim, and Dr. Seymour Grufferman for counsel and critical review of the manuscript, and gratefully acknowledges the collaboration of Dr. W.E. Hammond, Dr. W.W. Stead, and Ms. Ellen Garvey with computer applications, Dr. Dennis Tolley with the study design, and Ms. Suzanne Rosenblum with implementation.

This work was partially funded through grants from the North Carolina United Community Services and the Robert Wood Johnson Foundation.

References

1. Huntley RR, Steinhauser R, White KL, et al: The quality of medical care: Techniques and investigation in the outpatient clinic. J Chron Dis 14:630, 1961
2. Lashof J, Turner I: Periodic evaluation of outpatients. Arch Environ Health 8:531, 1964
3. Williamson JW, Alexander M, Miller GE: Continuing education and patient care research: Physician response to screening test results. JAMA 201:938, 1967
4. Schneiderman LJ, De Salvo L, Baylor S, et al: The "abnormal" screening laboratory result. Arch Intern Med 129:88, 1972
5. Korvin CC, Pearce RH, Stanley J: Admissions screening: Clinical benefits. Ann Intern Med 83:197, 1975
6. Parkerson GR: Labstand: A computerized system for reporting clinical laboratory data in standard units. J Fam Pract 6:611, 1978
7. International classification of health problems in primary care. Report of the Classification Committee of the World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians. Chicago, American Hospital Association, 1975
8. Davidsohn I, Henry JB: Clinical Diagnosis by Laboratory Methods—Todd-Sanford, ed 15. Philadelphia, WB Saunders, 1974, p 815
9. Elveback LR, Guillier CL, Keating FR: Health, normality, and the ghost of Gauss. JAMA 211:69, 1970
10. Harris EK, Kanofsky P, Shakarji G, et al: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects: Part 2: Estimating biological components of variation. Clin Chem 16:1022, 1970
11. Holland WW, Whitehead TP: Value of new laboratory tests in diagnosis and treatment. Lancet 2:391, 1974
12. Sackett DL: The usefulness of laboratory tests in health-screening programs. Clin Chem 19:366, 1973
13. Sunderman FW Jr: Current concepts of "normal values," "reference values," and "discrimination values" in clinical chemistry. Clin Chem 21:1873, 1975