The Determination and Interpretation of Reference Intervals for Multichannel Serum Chemistry Tests

James W. Mold, MD; Cheryl B. Aspy, PhD; Kenneth E. Blick, PhD; and Frank H. Lawler, MD, MPSH
Oklahoma City, Oklahoma

BACKGROUND. When interpreting the results of clinical chemistry tests, physicians rely heavily on the reference intervals provided by the laboratory. It is assumed that these reference intervals are calculated from the results of tests done on healthy individuals, and, except when noted, apply to people of both genders and any age, race, or body build. While analyzing data from a large screening project, we had reason to question these assumptions.

METHODS. The results of 20 serum chemistry tests performed on 8818 members of a state health insurance plan were analyzed. Subgroups were defined according to age, race, sex, and body mass index. A very healthy subgroup (n=270) was also defined using a written questionnaire and the Duke Health Profile. Reference intervals for the results of each test calculated from the entire group and each subgroup were compared with those recommended by the laboratory that performed the tests and with each other. Telephone calls were made to four different clinical laboratories to determine how reference intervals are set, and standard recommendations and the relevant literature were reviewed.

RESULTS. The results from our study population differed significantly from laboratory recommendations on 29 of the 39 reference limits examined, at least seven of which appeared to be clinically important. In the subpopulation comparisons, "healthy" compared with everyone else, old (≥75 years) compared with young, high (≥27.1) compared with low body mass index (BMI), and white compared with nonwhite, 2, 11, 10, and 0 limits differed, respectively. None of the contacted laboratories were following published recommendations for setting reference intervals for clinical chemistries. The methods used by the laboratories included acceptance of the intervals recommended by manufacturers of test equipment, analyses of all test results from the laboratory over time, and testing of employee volunteers.

CONCLUSIONS. Physicians should recognize when interpreting serum chemistry test results that the reference intervals provided may not have been determined properly. Clinical laboratories should more closely follow standard guidelines when setting reference intervals and provide more information to physicians regarding the population used to set them. Efforts should be made to provide appropriate intervals for patients of different body mass index and age.

KEY WORDS. Reference values; clinical chemistry; laboratory; data analysis. *(J Fam Pract 1998; 46:233-241)*

Serum chemistry panels, made possible by multichannel clinical chemistry analyzers, have revolutionized the manner in which physicians evaluate patients. After complete blood counts and urinalyses, they are the most frequently ordered laboratory tests in both primary care and hospital settings. It is generally assumed that the laboratory's reference intervals for the tests included in such panels contain the central 95% of the values obtained from a healthy population. Based on this assumption, it has been calculated that one can anticipate, for a given healthy patient, a probability of at least one abnormal result, given n separate tests in the panel of $1-(0.95)^n$ (assuming that the results of each test are independent of the others, which, of course, they are not).[1] It is also assumed that the reference intervals, with a few exceptions (eg, uric acid and gamma-glutamyl transpeptidase) apply equally well to men and women, whites and African Americans, young and old adults, and that they are unaffected by body mass.

From the Department of Family and Preventive Medicine (J.W.M., C.B.A., F.H.L.) and the Department of Pathology (K.E.B.), University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma. Requests for reprints should be addressed to James W. Mold, MD, University of Oklahoma Health Sciences Center, 900 NE 10th St, Oklahoma City, OK 73104. E-mail: jim-mold@uokhsc.edu

© 1998 Appleton & Lange/ISSN 0094-3509

During analysis of the data from a study designed to measure the effects of a screening panel (Chem 25) offered as a promotion by a local insurance company, we had the opportunity to test these assumptions regarding the reference intervals for the tests included in the panel. We reviewed distributions of test results in our population and compared them with the laboratory's reference intervals for 20 of the serum chemistry tests. In addition, we contacted four other laboratories to determine the procedures used by them to set reference intervals.

METHODS

Between February 1 and March 15, 1994, the Oklahoma State and Educational Employees Insurance Program offered its approximately 100,000 adult members older than 25 years the opportunity to have a Chem 25 test and lipid profile for $15. Phlebotomy sites were established in 300 locations throughout the state, and 8818 enrollees chose to have the blood tests done. Participants were asked to fast for 8 to 12 hours before testing. The tests were performed by a single laboratory in Oklahoma City using a DAX multichannel analyzer (Miles Inc, Tarrytown, NY). Procedures for the storage and transfer of specimens were standardized and in accordance with the procedures of the laboratory that performed the analyses.

Prior to having blood drawn, each individual was invited to participate in a research project involving the completion of questionnaires before and again several months after the blood chemistry testing; 4150 subjects agreed to participate and completed the first questionnaire at the time that their blood was drawn. Follow-up questionnaires were mailed on July 22, 1994. The research participants who had not returned their follow-up questionnaire by October 4, 1994, were sent a second questionnaire. Response to the first mailing was 1746; an additional 576 responded to the second mailing for a total of 2322 persons with both initial and follow-up questionnaire data (response rate = 56%). Late responders were somewhat older than early responders, but they were similar in respect to sex, prior history of heart attack, family history of heart attack, systolic blood pressure, diastolic blood pressure, height, and body mass index. It seems reasonable to assume that responders were younger and healthier than nonresponders.

The first questionnaire included several questions designed to identify individuals with a history of abnormal blood chemistry results, known medical problems, and any use of medications likely to affect the clinical chemistry test results. It also included the Duke Health Profile (DUKE), a reliable and well-validated, 17-item self-administered questionnaire that measures health-related quality of life.[2,3] Using these instruments, we retrospectively identified a "healthy" subgroup who had no history of abnormal liver tests, a high or low blood protein level, a high or low blood calcium level, high blood urea nitrogen (BUN), a low potassium level, a low sodium level, liver disease, thyroid disease, parathyroid disease, iron deficiency, diabetes, difficulty passing urine or pain with passage of urine, kidney disease, high blood cholesterol, gout, muscle disease (eg, muscular dystrophy), and who denied use of any of the following medicines during the previous week: blood pressure medicine, diuretic (fluid) medicine, potassium, medicine to lower cholesterol, iron (including iron in vitamins), estrogen (female hormones), thyroid medicine, or arthritis medicine, and who scored 90% or higher on the General Health subscale of the DUKE. Two hundred seventy individuals met these criteria for "health."

The distributions of results were examined for each test in the Chem 25 panel for all subjects, for those in the healthy and less healthy subgroups, and for subgroups defined by age (<75 or ≥75), race (white or nonwhite), and body mass index (<27.1 or ≥27.1; 27.1 was the mean for the study population). The distributions were tested for normality using the Kolmogorov-Smirnov one-sample test for sample sizes greater than 2000 and the Shapiro-Wilk test for smaller samples. Analyses were carried out using Statistical Analysis System software (SAS Institute, Inc). Percentiles were determined at 2.5 and 97.5 for all tests with both upper and lower reference limits, and 97.5 percentiles for those with a zero lower limit. These values were then compared with the upper and lower reference limits recommended by the laboratory by calculating 95% confidence intervals around the percentiles for the study groups, using a method described by Conover[4] for collecting confidence intervals for quantiles. Nonparametric methods were used since none of the test results were normally distributed and since outliers did not appear to be a significant problem.

Comparisons were made for 20 of the serum

chemistry tests. These included: BUN, creatinine, uric acid, sodium, potassium, chloride, carbon dioxide, calcium, phosphorus, total protein, albumin, globulins, total bilirubin, alkaline phosphatase, iron, alanine aminotransferase (ALT, SGPT), aspartate aminotransferase (AST, SGOT), lactate dehydrogenase (LD, LDH), gamma-glutamyl transpeptidase (GGT, GGTP), and creatine phosphokinase (CK, CPK). Excluded from our analyses were glucose, triglycerides, high-density lipoprotein (HDL) cholesterol, and low-density lipoprotein (LDL) cholesterol, all of which are affected by dietary intake. Total cholesterol was also excluded since the suggested reference intervals are not based on population distribution. We chose not to include prostate specific antigen (PSA) results in this set of analyses for similar reasons. For uric acid and GGT, the laboratory reported separate sets of reference intervals for men and women, and so they were analyzed separately. For four of the tests (including GGTP), the laboratory's lower reference limit was zero, leaving only one limit to analyze. In total, we evaluated 17 lower and 22 upper reference limits for 39 separate comparisons.

We contacted the clinical laboratory that performed the tests on our subjects to determine how they set their reference intervals. Unfortunately, the laboratory's ownership had changed; we were able to determine only the methods currently being used in that laboratory. Supervisors from three other well-known clinical reference laboratories, as well as the University of Oklahoma's laboratory, were also contacted. The methods used by these laboratories were compared with the methods suggested by the National Committee for Clinical Laboratory Standards (NCCLS, 1994)[5] and two standard textbooks.[6,7]

As a way to compare the general health of our population sample with that of a sample of primary care outpatients, we computed age- and sex-adjusted means for the general health subscale of the DUKE for our population and for a population reported by Parkerson.[8] This was accomplished by the direct method using the combined populations of the two samples as the reference population. Since the outpatient comparison population contained no one older than 65 years, we could only calculate an age- and sex-adjusted mean for the subset of our sample younger than age 65.

RESULTS

Comparisons of Calculated Reference Limits

Table 1 profiles the study population with regard to age, sex, race, "health" (as we defined it), and body mass index (BMI). Age and BMI were available for all subjects, while the data on which "health" was determined was available only for those who completed the first questionnaire (n=4150), and race was available only for those who completed the second questionnaire (n=2322).

None of the test results were normally distributed. Therefore, reference limits were set nonparametrically. Twenty-nine of the 39 quantile limits from the study population were significantly different from those suggested by the laboratory. Those that are particularly pertinent, either because of the size of the discrepancy or clinical relevance (determined

TABLE 1

Description of Total Study Population (N=8818)

	No. (%)
Age	
<75	8321 (94)
≥75	497 (6)
Gender	
Male	3636 (41)
Female	5157 (58)
Unknown	25 (1)
Race	
White	2161 (25)
Nonwhite	139 (2)
Unknown	6518 (74)
Health	
Healthy*	270 (3)
Not healthy	2053 (23)
Unknown	6495 (74)
Body mass index	
< Mean	4916 (56)
≥ Mean	3846 (44)
Unknown	56 (1)

*"Healthy" determined by a score of ≥ 90 on the Duke Health Profile General Health Scale and answers of no to questions regarding previous abnormal tests for liver, blood protein, blood calcium, BUN, potassium, or sodium, no to any current medical problems, no to having taken any medication for these conditions during the past week, and consuming <5 alcoholic drinks in a month.

retrospectively) are shown in Table 2.

Of the 4150 persons who completed the first questionnaire, 270 fit our definition of health (no history of medical problems known to affect serum chemistry test results and a general health score of 90 or better on the DUKE). Because of the relatively low number of "healthy" subjects, confidence intervals around the upper and lower quantiles for that group were somewhat wider, but 16 of the 39 comparisons still showed significant differences (no overlap of confidence intervals) from the laboratory reference limits. When the healthy group was compared with everyone else in the study population, the only significant differences were for the upper limits for GGTP and uric acid for men (Table 3).

Of the 39 comparisons between our old and younger subjects, 11 were significantly different. The most important of these are displayed in Table 4. Differences between subjects with low BMI and those with high BMI totaled 10, the most important of which are shown in Table 5. There were no significant differences between whites and nonwhites.

The results of the 4395 tests that fell within the reference intervals recommended by the laboratory (3% of a total of 168,167 tests) fell outside of the limits set from our healthy subgroup. Conversely, 2646 test results that initially fell outside of the laboratory's reference intervals (1% of a total of 6993) were within the intervals set from our healthy subgroup. One thousand three-hundred eighty-

TABLE 2

2.5th and 97.5th Percentiles for Selected Tests Comparing Reference Intervals Derived from Our Study Population (N=8818) with the Reference Intervals Recommended by the Laboratory that Performed the Tests

	Laboratory Reference Intervals	Study Population Intervals
	2.5th and 97.5th Percentiles	2.5th and 97.5th Percentiles (95% CI)
Potassium	3.3 - 5.0 mEq/L	3.6 - 5.4 mEq/L (3.5, 3.6 - 5.3, 5.4)
Calcium	8.5 - 10.6 mg/dL	8.8 - 10.4 mg/dL (8.8, 8.8 - 10.3, 10.4)
Albumin	3.0 - 5.0 g/dL	3.7 - 4.9 g/dL (3.7, 3.7 - 4.8, 4.9)
Creatine phosphokinase (CK)	0 - 270 U/L	35 - 290 U/L (34, 36 - 277, 306)
Lactate dehydrogenase (LD)	60 - 225 U/L	111 - 231 U/L (111, 113 - 228, 237)
Globulin	0.9 - 5.8 g/dL	2.4 - 4 g/dL (2.4, 2.4 - 4, 4)
Alanine aminotransferase (ALT)	0 - 45 U/L	9 - 59 U/L (9,9 - 57, 62)

CI denotes confidence interval.

TABLE 3

2.5th and 97.5th Percentiles for Healthy Group (N=270) Compared with All Others (N=8549) in Study Population

	Healthy*	Remaining Sample
	2.5th and 97.5th Percentiles (95% CI)	2.5th and 97.5th Percentiles (95% CI)
Uric acid (men)	3.5 - 7.9 (2.9, 4 - 7.6, 8.5)	3.5 - 8.8 (3.4, 3.6 - 8.7, 9.1)
Gamma-glutamyl transpeptidase (men)	11 - 64 (10, 12 - 51, 74)	11 - 102 (10, 12 - 51, 74)

*"Healthy" determined by a score of ≥ 90 on the Duke Health Profile General Health Scale and answers of no to questions regarding previous abnormal tests for liver, blood protein, blood calcium, BUN, potassium, or sodium, no to any current medical problems, no to having taken any medication for these conditions during the past week, and consuming <5 alcoholic drinks in a month.
CI denotes confidence interval.

eight individuals who had no results outside of the laboratory's reference intervals (33% of a total of 4260 individuals) had at least one result outside of the intervals set from the healthy subgroup, while 2448 individuals with at least one "abnormal" result (54% of a total of 4498) had no results that fell outside of the reference intervals set from the healthy subgroup.

TABLE 4

2.5th and 97.5th Percentiles for Selected Tests Comparing the Young Group (N=8321) with the Old Group (n=497) in the Study Population

	Young, <75 years	Old, ≥ 75 years
	2.5th and 97.5th Percentiles (95% CI)	2.5th and 97.5th Percentiles (95% CI)
Blood urea nitrogen (BUN)	7 - 22 (7, 7 - 21, 22)	8 - 30 (7, 9 - 29, 38)
Creatinine	0.7 - 1.4 (0.7, 0.7 - 1.4, 1.4)	0.7 - 1.7 (0.7, 0.7 - 1.7, 2.0)
Albumin	3.7 - 4.9 (3.7, 3.7 - 4.9, 4.9)	3.5 - 4.6 (3.4, 3.6 - 4.6, 4.7)
Alkaline phosphatase	39 - 116 (38, 40 - 114, 188)	41 - 133 (37, 45 - 124, 150)
Creatine phosphokinase (CK)	36 - 292 (35, 37 - 282, 309)	29 - 213 (24, 32 - 178, 251)
Lactate dehydrogenase (LD)	111 - 230 (110, 113 - 226, 235)	117 - 250 (89, 125 - 238, 282)
Alanine aminotransferase (ALT)	9 - 60 (9, 10 - 57, 63)	9 - 38 (7, 10 - 36, 50)

CI denotes confidence interval.

Methods Used by Laboratories to Set Reference Limits

The methods used to set reference intervals vary considerably among the five laboratories we contacted. One laboratory relies almost entirely on the reference intervals provided by the manufacturer of the multichannel chemistry analyzer when it can be shown through a series of analytical performance tests that the laboratory's technicians can accurately reproduce the results predicted by the manufacturer on known samples. We were referred to two manufacturers. Both included a disclaimer paragraph in their instruction manual stating that reference intervals may vary from location to location and should be determined by each laboratory. One disclosed that its suggested reference intervals were obtained from analysis of tests performed on a sample of 200 blood donors in New York City.

Two other laboratories (including the one that purchased the laboratory that conducted the tests on our subjects) use aggregate data from all tests performed by the laboratory over a 3- to 6-month period. A Gaussian distribution is assumed and reference limits are set 2 standard deviations from the mean. The intervals are then checked against "published standards" and may be adjusted to some degree if discrepant.

The fourth reference laboratory and the University of Oklahoma Hospital laboratory base their reference intervals on the results of tests run on samples from 100 (the university) to as many as 2000 (the reference laboratory; 200 from each of 5 to 10 sites) "healthy adults." These are, for the most part, employees of the hospital or laboratory who claim to be in good health and agree to refrain from taking medications and food for 8 to 12 hours prior to testing. The university also uses a very brief health questionnaire; the reference laboratory does not. The reference laboratory assumes a normal distribution while the university uses more sophisticated statistical methods to set reference intervals.

Recommendations (excerpted and paraphrased) from the National Committee for Clinical Laboratory Standards[5] (NCCLS) are listed in Appendix A. These are essentially identical to the recommendations of Blick and Lyles[6] and *Tietz Fundamentals of Clinical Chemistry.*[7] None of the laboratories contacted were following all of these recommendations.

DISCUSSION

Our findings regarding reference intervals and the methods used to set them were unexpected, the data having been collected for other reasons. We believe,

TABLE 5

2.5th and 97.5th Percentiles for Selected Tests by Body Mass Index (BMI) in the Study Population

	BMI < 27.1*	BMI > 27.1
	2.5th and 97.5th Percentiles (95% CI)	2.5th and 97.5th Percentiles (95% CI)
Uric acid (men)	3.4 - 8.4 (3.2, 3.5 - 8.2, 8.6)	3.7 - 9.1 (3.5, 3.9 - 8.9, 9.2)
Uric acid (women)	2.1 - 6.6 (2.0, 2.1 - 6.5, 6.8)	2.7 - 8 (2.5, 2.8 - 7.7, 8.3)
Alkaline phosphatase	38 - 113 (36, 38 - 111, 117)	43 - 121 (42, 44 - 117, 125)
Aspartate aminotransferase (AST)	10 - 33 (10, 11 - 32, 35)	10 - 42 (10, 11 - 39, 44)
Lactate dehydrogenase (LD)	110 - 225 (108, 111 - 220, 231)	114 - 239 (112, 116 - 234, 247)
Alanine aminotransferase (ALT)	9 - 49 (8, 9 - 47, 52)	11 - 70 (10, 11 - 67, 76)
Gamma-glutamyl transpeptidase (GGT)	7 - 57 (7, 8 - 53, 68)	9 - 87 (8, 9 - 77, 108)

*BMI of 27.1 was the mean of the total study population.

however, that these results have important implications because physicians depend heavily on the suggested reference intervals as a guide to the presence of abnormalities.[10] The four outside reference laboratories that we contacted all have locations throughout the country and perform thousands of clinical chemistry panels daily. If physicians assume that these reference intervals (commonly referred to as "normal ranges") are based on results from healthy adults who are representative of persons in the physician's own patient populations and were determined using appropriate statistical methods, they may be misinterpreting the results.

Reference intervals suggested by different laboratories obviously differ, as perhaps they should. The values suggested by the laboratory that performed the tests on our subjects may be more aberrant than those suggested by other laboratories, and thus the differences we found may be more striking than those derived from the comparison of our results with other standards. For instance, the lower reference limit for serum albumin (3.0 g/dL) suggested by our laboratory is lower than that generally accepted for ambulatory patients (3.5 g/dL).

Laboratories that calculate reference intervals based on data from all tests run over time will reflect the average health of the population tested. Therefore, hospital laboratories might be expected to have proportionately less healthy reference populations than laboratories doing mostly outpatient testing. In either case, the resultant intervals will probably reflect a less than healthy population.

The group who had blood drawn during the insurance company-sponsored screening program appear to have been healthier, based on their laboratory tests results, than those used to set the reference intervals for the laboratory that did the testing. For example, the lower reference limits for potassium and albumin were higher and the upper reference limit for lactate dehydrogenase was lower than those suggested by the laboratory, even though 6% of the study population was older than 75 and only 3% met our strict criteria for excellent health. Our sample was also probably healthier than the typical population of patients seen by family physicians in their practices, based on mean DUKE scores. The mean age and gender adjusted subscale scores for physical health for 1916 family practice outpatients seen at Duke University in 1991 and 1992 was 66.7 (95% CI, 69.0 to 70.2) compared with 69.6 (95% CI, 69.0 to 70.2) for the 3048 comparably aged subset of our subjects who completed the DUKE.[8]

When we compared the reference intervals that we would have set using our entire population with those from the "healthy" subset, the only statistically significant differences were for the upper reference limits for uric acid and GGTP (they were lower for males in the healthy group). These differences may

be the result of greater alcohol use among the "less healthy" group or they could be due to chance alone. The actual intervals that we would have chosen would not have been substantially different for any of the other tests, and the confidence intervals were extremely narrow. Since we used more rigorous criteria to determine "health" than any of the laboratories we contacted, we have included the reference intervals that we calculated from the healthy subgroup, since our intervals may be of some value to practitioners and clinical pathologists (Appendix B). In fact, a pathologist in one of the laboratories urged us to publish the intervals obtained for our healthy subgroup because he has a great deal of difficulty defining and recruiting a sample of healthy volunteers.

None of the laboratories that we contacted had attempted to determine whether test results vary with race, body mass index, or age. Although we found a higher 97.5th percentile for CK (379 vs 271) and lower 97.5th percentile for GGTP (61 vs 98) for nonwhite men, the confidence intervals were wide and the differences did not quite reach statistical significance. Blacks tend to have higher concentrations of total protein due to greater concentrations of gamma globulins, as well as higher levels of LD and CK, related to a tendency to have greater muscle mass than whites.[9] Our data are consistent with these generalizations for total protein and CK, but not LD. Greater body mass index, on the other hand, was associated with a number of significant differences that might be important to clinicians. Increasing body weight is typically associated with higher concentrations of LD, glucose, and phosphate.[9] Men with greater body mass are said to be more likely to have higher levels of AST, creatinine, and total protein, and heavier women are likely to have higher serum calcium concentrations. Our data tend to support these contentions.

The issue of age differences in laboratory test results is complicated by the greater prevalence of illnesses and abnormalities in older people. Many of our findings appear to demonstrate this. However, we wonder if the somewhat lower 97.5th percentile values for CK and ALT are not of some importance, much as creatinine levels are often lower for a given creatinine clearance in older people because of decreased muscle mass. Serum concentrations of albumin, total protein, calcium, phosphorus, and uric acid decrease and concentrations of alkaline phosphatase, AST, and glucose increase with age.[9] Also mentioned are changes occurring in the serum of women following menopause, including increases in ALT, AST, alkaline phosphatase, phosphorus, and uric acid. Hammerman-Rozenberg et al[10] looked at this issue carefully and recommended a different set of reference intervals for elderly patients.

It was not particularly surprising that the test results were not normally distributed. A number of biological measures with physiologic "set-points" do not follow a Gaussian distribution pattern.[11] The results were roughly bell-shaped and could certainly have been assumed to have been normally distributed. For purposes of setting reference intervals, the middle 95% of results must be determined. Several methods could be used to do this, including parametric methods assuming a normal distribution, parametric methods following log transformation of the test results to achieve more nearly normal distributions, and nonparametric methods. We used the nonparametric method of determining reference intervals because it is extremely accurate, since it simply involves discounting 2.5% of the values from either end of the distribution, unless there are a significant number of outliers.[7,12,13] We did not have a significant number of outliers for our data from the healthy subgroup, and it was only potentially problematic for CK readings for the whole group.

CONCLUSIONS

It may be erroneous to assume that reference intervals (normal ranges) for laboratory test results are based on results from healthy persons that resemble those in a typical clinical practice population. In many cases, the reference population is neither known to be healthy nor is it representative of the patient being tested. Because of the manner in which reference intervals are commonly set, the probability that an abnormal test result is truly abnormal is greater than predicted mathematically. Furthermore, when evaluating test results, physicians should take into account factors such as age and body mass index that appear to significantly affect the distribution of results in the population.

Clinical chemistry laboratories should more closely adhere to the recommended standards published in textbooks or suggested by NCCLS. Reference population characteristics should probably be provided to the physician to facilitate inter-

pretation of reference intervals and test results. Reference intervals based on patient age, sex, and BMI could be provided on laboratory reports. Separate reference intervals are already provided for men and women for some tests (eg, uric acid). Alternatively, all results could be adjusted for these factors and reported as percentiles rather than concentrations. This would have the additional advantage of standardizing results across laboratories. Other methods such as "multivariate reference regions" and "subject-based reference values" have also been suggested.

Finally, it is important for clinicians to remember that laboratory tests, whether within or outside of the published reference intervals, must be interpreted in the context of the clinical situation. It is not sufficient to scan laboratory reports for those test results marked abnormal.

ACKNOWLEDGMENTS

We'd like to thank George Parkerson, MD, for his invaluable assistance in how to use the Duke Health Profile; to the Oklahoma State and Educational Employees Insurance Program for their assistance in data procurement; and to Jan Wilmoth for manuscript preparation.

REFERENCES

1. Ober KP. Uncle Remus and the cascade effect in clinical medicine. Am J Med 1987; 82:1009-13.
2. Parkerson GR, Broadhead WE, Tse CKJ. The Duke Health Profile: a 17-item measure of health and dysfunction. Med Care 1990; 28:1056-72.
3. Parkerson GR, Broadhead WE, Tse CKJ. Comparison of the Duke Health Profile and the MOS Short-Form in healthy young adults. Med Care 1991; 29:679-83.
4. Conover, WJ. Practical nonparametric statistics. 2nd edition. New York, NY: John Wiley & Sons; 1980:111-6.
5. Sasse EA, Aziz KJ, Harris EK, et al. How to define and determine reference intervals in the clinical laboratory; approved guideline. NCCLS Document C28-A. Wayne, Pa: The National Committee for Clinical Laboratory Standards, 1994.
6. Blick KE, Liles SM. Principles of clinical chemistry. New York, NY: John Wiley & Sons, 1985:73-9.
7. Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER, eds, Aldrich JE, consulting ed. Tietz fundamentals of clinical chemistry. Philadelphia, Pa: WB Saunders Company, 1995:182-91.
8. Parkerson, GR. User's guide for the Duke Health Profile (DUKE). Durham, NC: Duke University; 1996.
9. Burtis CA, Ashwood ER, eds. In: Tietz textbook of clinical chemistry, 2nd edition. Philadelphia, Pa: WB Saunders Company; 1994:91-101, 461.
10. Hammerman-Rozenberg R, Cohen A, Ginsberg, et al. Laboratory reference values for the 70 year olds. Isr J Med Sci 1996; 32:611-20.
11. Lawler FH, Brace RA. Fetal and maternal arterial pressures and heart rates: histograms, correlations, and rhythms. Am J Physiol 1982; 243:R433-R444.
12. Young DS. Determination and validation of reference intervals. Arch Pathol Lab Med 1992; 116:704-9.
13. Reed AH, Henry, RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. Clin Chem 1971; 17:275-84.
14. Solberg HE. Multivariate reference regions. Scand J Clin Lab Invest 1995; 55:(suppl 222)3-5.
15. Solberg HE. Subject-based reference values. Scand J Clin Lab Invest 1995; 55:(suppl 222)7-10.

Appendix A

Selected Recommendations of The National Committee for Clinical Laboratory Standards for Setting Reference Intervals

The reference population should be determined with considerable forethought based on a review of the literature pertinent to the specific test or tests in question. Specific exclusion criteria and partitioning criteria should be determined and a questionnaire designed to identify these criteria. "Reference individuals should not be hospital or clinic patients unless absolutely necessary . . ."

"It may be necessary, for certain analytes, to define conditions for establishing reference intervals in different subclasses. Many of these . . . constitute partitioning factors . . . and they may require separate reference intervals."

"The reference interval is . . . the interval between and including two numbers, an upper and lower reference limit, which are estimated to enclose a specified percentage (usually 95%) of the values for a population from which the reference subjects have been drawn. In some cases, only one reference limit is of medical importance, usually an upper limit, say the 97.5th percentile."

"Because the reference values of many analytes do not follow the Gaussian form, non-parametric methods for determining reference limits are both simpler and more appropriate. Unless outliers are known to be aberrant observations, the emphasis should be on retaining rather than deleting them."

Adapted with permission from Sasse EA, Aziz KJ, Harris EK, et al. How to define and determine reference intervals in the clinical laboratory; approved guideline. NCCLS Document C28-A. Wayne, Pa: The National Committee for Clinical Laboratory Standards; 1994.

Appendix B

Reference Intervals for Selected Tests for Healthy Subgroup

Test	Healthy Subgroup Reference Intervals
Blood urea nitrogen	6-21 mg/dL
Creatinine	0.7-1.4 mg/dL
Uric acid (men)	3.5-7.9 mg/dL
Uric acid (women)	1.8-6.7 mg/dL
Sodium	134-144 mEq/L
Potassium	3.6-5.2 mEq/L
Chloride	101-110 mEq/L
Carbon dioxide	21-31 mEq/L
Calcium	8.9-10.3 mg/dL
Phosphorous	2.7-4.7 mg/dL
Protein	6.7-8.3 g/dL
Albumin	3.8-4.9 g/dL
Bilirubin	0.3-1.2 mg/dL
Alkaline phosphatase	38-114 U/L
Aspartate aminotransferase	10-34 U/L
Creatine phosphokinase	38-318 U/L
Lactate dehydrogenase	106-238 U/L
Iron	45-183 µg/dL
Globulin	2.3-4 g/dL
Alanine aminotransferase	10-56 U/L
Gamma-glutamyl transpeptidase (men)	11-64 U/L
Gamma-glutamyl transpeptidase (women)	8-53 U/L

Characteristics of Healthy Subgroup (N=270)	
Age	
mean (SD)	47.7 (12.8)
range	25 to 96
Body mass index	
mean (SD)	25.7 (4.1)
range	13.4 to 38.8
Sex, %	
Men	51
Women	48
Unknown	1
Race, %	
White	52
Nonwhite	3
Unknown	44

SD denotes standard deviation.