Reference Interval Computation Using Robust vs Parametric and Nonparametric Analyses, *Paul S. Horn,[2*] Amadeo J. Pesce,[1] and Bradley E. Copeland[1]* ([1] Department of Pathology and Laboratory Medicine, College of Medicine, and [2] Department of Mathematical Sciences, College of Arts & Sciences, University of Cincinnati, P.O. Box 210025, Cincinnati, OH 45221-0025; * author for correspondence: fax 513-556-3417, e-mail Paul.Horn@UC.edu)

The clinical chemist is faced with the problem of defining reference intervals for many analytes. Problems that hinder such a determination are the presence of outliers in the data set and the inability to accumulate the recommended sample size *(1)*. We previously have demonstrated the theoretical basis for the application of robust methods to resolve these problems *(2)*. In particular, we consider the problem of establishing that the reference population is "healthy" to be a nearly impossible task because many disease processes may be missed in the examination process. Our own experience shows that diabetics were initially classified as healthy in our test population, and only after a thorough review of all of the data was it possible to elicit the presence of this disease. In this report, we propose the use of a robust estimator we have described previously *(2)*. The advantage of this approach is that it is more tolerant of outliers in the reference population data and does not require as large a sample size as the nonparametric calculation method, nor does it require the reference data to be transformed to a gaussian distribution, which is not always possible. We then apply and compare this robust estimator with both the traditional nonparametric and parametric analysis in determining reference intervals for a well-studied population.

The Fernald Medical Monitoring Program provided us with a documented healthy sample (T) to test the three methods: parametric, nonparametric, and robust. Our computer-generated sample (W) offered the possibility to test our estimates of reference intervals in a population with a greater potential incidence of diseases *(3)*. The robust approach offered the opportunity to look at a more conservative estimate of a "healthy reference interval" *(2)*. A second advantage of the robust approach is that it provides a greater degree of confidence in the calculation of reference intervals for those "bins" that have small numbers of samples *(2)*.

The methods used for deriving reference intervals were the same as those described in our previous work *(2)*. We used a nonparametric approach that uses weighted order statistics, where the weights are functions of the incomplete beta function *(4)*. The second method uses the Box-Cox methodology to transform the original data to achieve normality. A further transform is applied, if necessary, to remove residual excess kurtosis *(5)*. The endpoints of this interval are then transformed back to the original scale. The robust method we use is based on the robust quantile estimator for skewed population described by Horn *(6)*.

The final method examined was the full-sample version of the robust prediction interval described by Horn *(7)*, computed on transformed data. The robust interval requires only symmetry, and not normality; therefore, it can be computed after the initial Box-Cox transformation, if such a suitable transformation is possible. Because this interval does not require removal of excess kurtosis or assessment of data normality by the Anderson-Darling test, there are fewer cases, compared with the traditional approach, where its computation is not possible.

To test these statistical methods for calculating the reference interval, we have evaluated samples from two populations. Reference intervals were calculated in these populations for 27 analytes, two genders, and six age groups for males and seven age groups for females.

The T population represents a unique data set *(3)*. Each of the 2948 patients entered as T were clinically examined by disease history, family history, psychological history, physical examination, chest x-ray, and laboratory values. Each participant's health was scored on a rating system of 1–6, with 6 being in excellent health and 5 being healthy. Only those participants who were scored a 5 or 6 were used for the initial calculation of the reference intervals. The population is representative of the greater Cincinnati metropolitan area.

The W group was selected from those patients treated by either a general internist or family practitioner who were the only physicians allowed to order this battery of "screening tests" *(8)*. Thus, the patients were drawn from the general population pool. Our other criteria were that no other tests were ordered on that visit and that we had no record of any other tests being ordered for those patients as either inpatients or outpatients over a 1-year time span. Because the individuals were seeing a physician, this implied that in some cases there would be some underlying disease. With the T group, ~2% of the patients were classified by physician assessment as not healthy, and these were deleted from the T group calculations. This was not done for the W group. The Wide Age Range Population W and the T population had similar geographic and ethnic demographics. The T population was 99% Caucasian, whereas the W population was 97% Caucasian. We ensured comparability of the laboratory data between the two patient groups by using the same NIST standardized laboratory assays *(9)*. Approval for use of the patient data was obtained from the University of Cincinnati Institutional Review Board and the Trustees of the Fernald Medical Monitoring Programs. The methods for quality control are those described by Copeland *(10)*. The chemistry procedures are those reported by Faulkner and Meites *(11)*. The hematology values were obtained on a Coulter STKR (Coulter Electronics).

Examples of the data generated are shown in Table 1. Table 1 describes 95% reference intervals of four analytes for males 50–59 years of age for each of the two populations, T (Fernald Medical Monitoring Program) and W (Wide Age Range Population), calculated by the four methods. Each analyte-method cell consists of two numbers defining the endpoints of the appropriate reference interval. For example, the 210 cholesterol values for males 50–59 years of age from the T group gave a 95% reference

Table 1. 95% reference intervals for males 50–59 years of age.

	T group (n = 210)				W group (n = 178)			
Analyte	**Nonparametric**	**Robust**	**Transformed**	**Transformed robust**	**Nonparametric**	**Robust**	**Transformed**	**Transformed robust**
ALT,[a] U/L	9.65	9.65		8.18	10.78	10.78	10.66	10.62
	55.21	53.36		52.00	67.27	56.20	58.98	59.52
Cholesterol, mmol/L	3.97	3.97	4.01	3.99	4.02	4.02	4.10	4.12
	7.79	7.72	7.70	7.72	7.90	7.79	7.80	7.76
Glucose, mmol/L	4.64	4.64		4.50	3.62	3.62		3.76
	13.73	11.53		8.01	10.46	9.35		8.35
Hemoglobin, mmol/L	2.00	2.00	2.02	2.02	1.90	1.90	1.92	1.93
	2.61	2.57	2.57	2.57	2.65	2.62	2.62	2.61

[a] ALT, alanine aminotransferase.

interval of 3.97–7.79 mmol/L, using the nonparametric procedure. The other three procedures provided similar, although slightly narrower, intervals. (All of the tables are available by request. Each table consists of 27 analytes, 5 of which are hematological.)

Where a transformation of the data for a particular test to a gaussian distribution was not possible or if the Anderson-Darling test rejected normality, the reference interval was not computed and the appropriate cell was left blank. It is noteworthy that in approximately one-third of the cases, no suitable transformation was found to allow for use of traditional normal theory methods. However, it is reassuring that when a successful transition was achieved, the resulting reference interval was comparable to that of the transformed robust (which requires only symmetry) as well as the untransformed robust method. It is also noteworthy that the robust method, although designed primarily for right-skewed populations, gives results that are comparable to the other, traditional methods for left-skewed data such as hemoglobin.

The reference interval width for the W group was compared to that of the T group by using their ratio because this dimensionless quantity allows for comparisons across analytes. For the 27 analytes tested, the interval width for the W data was on average 25% wider than the interval width based on the T data using either the nonparametric or robust analysis. To evaluate the effect of using different methods of estimating the reference interval, the ratio of the interval width calculated by the nonparametric procedure to that of the robust procedure was examined. For the 27 analytes tested, the nonparametric interval was 2.7% wider ($P < 0.01$) for the T group and 3.7% wider ($P < 0.01$) for the W group. The smaller percentage of difference in the carefully screened and documented healthy T group shows that under these conditions, the robust and nonparametric methods are in better agreement. These observations are consistent with our simulation data.

Our recommendation for the development of reference intervals includes the use of both nonparametric and robust estimators where the relationship between the intervals estimated by the two methods can be used as an "ad hoc" estimate of the reference population quality. Evaluation of the reference intervals by both of these techniques should yield similar results. If they do not agree, then consideration must be given to the possibility that a significant part of the sample could have diseased individuals or that multiple populations are being measured.

References

1. Sasse EA, Harris EK, Krishnamurthy S, Lee HT Jr, Rutland A, Seamonds B. How to define, determine, and utilize reference intervals in the clinical laboratory. NCCLS Document C28-P. Wayne, PA: NCCLS, 1992.
2. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. Clin Chem 1998;44:622–31.
3. Copeland BE, Pesce AJ. The medical heritage concept: a model for assuring comparable laboratory results in long-term longitudinal studies. Ann Clin Lab Sci 1992;22:110–24.
4. Harrel FE, Davis CE. A new distribution free quantile estimator. Biometrika 1982;69:635–70.
5. Harris EK, Boyd JC. Statistical bases of reference values in laboratory medicine. New York: Marcel Dekker, 1995:361pp.
6. Horn PS. Robust quantile estimators for skewed populations. Biometrika 1990;77:631–6.
7. Horn PS. A biweight prediction interval for random samples. J Am Stat Assoc 1988;83:249–56.
8. Pesce AJ, Copeland BE, Jenkins S, Henry J, Henderson J. Healthy human reference intervals for an adult population using computer analysis of an ambulatory outpatient data base. In: Sunderman FW, ed. Applied seminar on clinical science of aging. Philadelphia: Institute for Clinical Science, 1994:101–10.
9. De BK, Karr BA, Gosn S, Copeland BE. Problems and practical considerations in assaying accuracy with NIST SRM 909a: report of defective vials. Clin Chem 1996;42:1832–7.
10. Copeland BE. Quality control, In: Kaplan LA, Pesce AJ, eds. Clinical chemistry: theory analysis and correlation. St. Louis, MO: CV Mosby, 1984:310–37.
11. Faulkner WR, Meites S, eds. Geriatric clinical chemistry. Washington, DC: AACC Press, 1993:190–584.

Measurement of Thyroid-stimulating Hormone Receptor Autoantibodies by ELISA, *Jane Bolton,[1] Jane Sanders,[1] Yasuo Oda,[1] Ceri Chapman,[1] Reiko Konno,[2] Jadwiga Furmaniak,[1] and Bernard Rees Smith[1*]* ([1] RSR Ltd., Avenue Park, Pentwyn, Cardiff CF23 8HE, UK; [2] Cosmic Corporation, Tomisaka Bldg., 2-7-3 Koishikawa, Bunkyo-ku, Tokyo 112, Japan; * author for correspondence: fax 44-1222-732704)

Hyperthyroidism in Graves disease is attributable to autoantibodies to the thyroid-stimulating hormone receptor (TSHR), and measurement of these TSHR autoantibodies (TRAbs) can be useful in disease diagnosis and