STATISTICS IN MEDICINE, VOL. 2, 191–198 (1983)

DETERMINING REFERENCE ('NORMAL') LIMITS IN MEDICINE: AN APPLICATION

P. D. BEZEMER,* J. C. NETELENBOS,† C. MULDER,‡ J. A. THEUNE,* I. H. STAMHUIS,* AND J. P. STRAUB‡

Free University, P.O. Box 7161, 1007 MC Amsterdam, The Netherlands

SUMMARY

We provide an account of a study to assess reference limits for eight routine laboratory determinations at the Academic Hospital of the Free University, Amsterdam and emphasize methodological issues rather than results. We argue that reference limits have use mainly in the first phase of the diagnostic process. Reference and target populations should be grossly comparable, and therefore patients (after slight selection) could well serve as references. However, we found major differences between in- and out-patients, so we suggest that this factor, together with age and sex, be taken into account. To arrive at reliable limits, the size of the reference sample should be at least 100. Laboratory reports should provide percentiles, which enable a more flexible decision than do fixed limits.

KEY WORDS Reference values Diagnostics Sample size Percentiles

INTRODUCTION

Reference or normal ranges have, as Benson[1] remarked in 1972, a vague but comforting role in laboratory medicine. Despite the advent of medical decision making,[2,3] this statement is just as true ten years later. Isn't it comforting to have limits at hand with which one can immediately compare a laboratory result? On further consideration, however, most existing reference limits have little practical use, because of various inadequacies, as explained in this paper. Elion-Gerritzen's[4] conclusion that some clinicians do not use existing limits in judging a laboratory result for a patient but use others (action levels) instead, is pertinent in this respect.

The study reported in this paper sought to find reference limits less vague and of more practical use than those existing[5] at the Academic Hospital of the Free University (AZVU) at Amsterdam. We stress in this paper methodological issues rather than results. Strictly speaking, the latter are interesting only for the hospital concerned. We will confine ourselves to univariate ranges.

We stress one important point concerning the proper context for use of reference ranges. One compares a patient value with only one population, without taking into account other populations (Figure 1). When the value does not fit the pattern of values found in the reference population (e.g. healthy persons), it is generally not clear what pathology, if any, exists. One seldom reaches a final diagnosis using reference values. To this end, discrimination values,[6] demarcating two or more

* Department of Medical Statistics
† Department of Internal Medicine
‡ Department of Clinical Chemistry

0277–6715/83/020191–08$01.00
© 1983 by John Wiley & Sons, Ltd.
Received September 1982

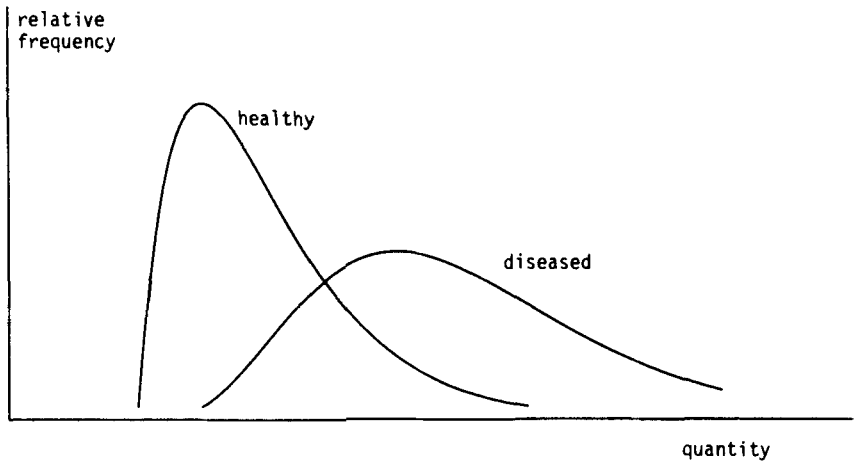

Figure 1. Overlapping distributions for healthy persons and a certain population of diseased persons

populations, are suited. The important consequence is that reference values have use only in the limited context of the first phase of a diagnostic process, i.e. to screen or to raise ideas about possible pathology. The finding of a value outside the range, and sometimes even within (because of overlap), usually indicates closer examination of the patient.

DESIGN OF THE STUDY

Reference population

There is a major controversy in the literature with regard to the choice of reference populations. Some plead for objectively healthy (i.e. after careful examination) people,[7] others propose hospital patients,[8,9] whereas, in practice, often donors, medical personnel, etc are taken as the reference. We believe that the choice depends on the target population, i.e. the population of persons for whom the range is intended. One can make a proper comparison between patient values and the range only when target and reference population are comparable with respect to characteristics influencing the quantity concerned.

In this study the target population consists of in- and out-patients referred to the department of Internal Medicine of our hospital. We decided to take as reference persons all newly admitted patients to this department in the period from October 1978 until March 1979, with the exception of patients with final diagnoses known to influence strongly the quantity or quantities concerned. We specified these exception criteria beforehand; they differed from one quantity to another. The arguments for this choice are:

(1) Reference persons and (later) patients are comparable in as many aspects as possible (age, sex, social class, stress, etc).
(2) It is well-known that the vast majority of routine laboratory results are 'normal'. We note that we avoid the circular reasoning inherent in the papers by Hoffmann[10] and many others largely because we exclude patients, not laboratory data.
(3) It is relatively simple to form a group of reference persons as proposed here. Therefore it is also feasible to determine reference limits periodically afresh (or adaptively).

We gathered a total of over 1000 reference persons, about 55 per cent of whom were in-patients, and excluded about 5 per cent in-patients and 1 per cent out-patients. For each person we recorded in-/out-patient, sex, age, height, weight, race (caucasian or not), fasting blood sample or not and final diagnosis (in nine broad classes).

Determinations

Since it was the first study of this kind at the AZVU, we confined our attention to eight routine determinations. In addition to practical advantages, the routine nature of the determinations ensured that they did not receive special care. Should this be the case, we would lose comparability between results in reference and target population to some extent. More precise measurement for reference persons than for (later) patients, would result in ranges that are too narrow!

The eight determinations were: albumin, total protein, calcium, phosphorus, creatinine, urea, uric acid and iron. We employed standard methods and took the blood samples after a light breakfast in about 67 per cent of patients and after fasting in the remaining patients (only relevant for phosphorus).

Sample size

The size of the reference sample should depend on the desired precision. Most important is the population fraction of values outside the reference limits. Figure 2 provides 90 per cent intervals for this fraction, assuming random sampling, and for a one-sided reference limit that provides an unbiased estimate of the 97·5th percentile. For the normal distribution[11] this limit has the form 'mean $\pm k$ times standard deviation', where k depends on the sample size, and in the distribution-free case it is an order statistic. The computation of the intervals was straightforward,[12] using the non-central Student distribution (*), Monte Carlo methods ($\diamond$), and the binomial distribution (○). Figure 2 shows that:

(1) The 'real' fraction of values above a sample reference limit may vary considerably from one

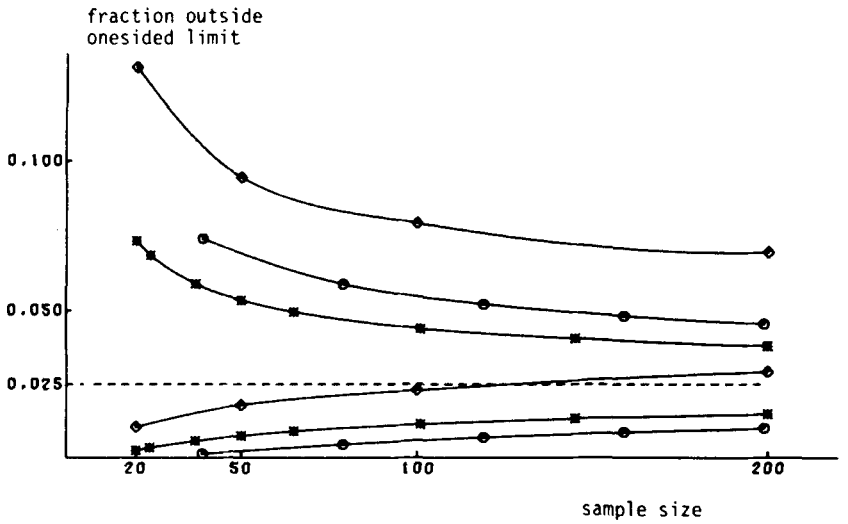

Figure 2. Ninety per cent-intervals for the population fraction above the reference limit which estimates the 97·5th percentile unbiased, plotted against the sample size; reference limit (*) rightly based on the normal distribution, ($\diamond$) faulty based on the normal distribution (distribution is lognormal) and (○) distribution free

sample to the other. Even with a sample size of 100, a desired percentage of 2·5 results in a 90 per cent interval of about 1–5 per cent (* and ○). Clearly, one should discourage reference samples of sizes, say, between 20 and 50, which, in practice, have frequent use.

(2) Using limits as if the distribution were normal, when it is really lognormal, leads to a serious bias (without loss of generality we assumed a two parameter lognormal distribution).

In our study we foresaw a partitioning of the reference sample. Because of Figure 2, we wanted to have at least 100 persons per group, and therefore decided to gather about 1000 patients.

Limits

Reference limits may serve several purposes, and of course each purpose calls for its own type of limit. We will not discuss these here. In the majority of cases, however, and in this study too, one needs 'routine limits'. Then percentiles in the reference sample (without interpolation), i.e. order statistics, which estimate percentiles in the reference population, are adequate. Several authors[8,13] have advocated the use of percentiles. We mention the arguments for this choice:

(1) As is well known, many quantities do not distribute normally. Although a positive skewness is the most usual deviation, others occur. Therefore percentiles in the reference sample are a good general choice, the more so when the tail of the distribution is most important.

(2) Order statistics are simple, readily understood limits. They are measured values themselves, not (artificially) computed ones.

Finally, which percentile(s) should one choose? We decided to cut off 2·5 and 10 per cent from either side, but for urea, because of clinical relevancy, only from the upper side. Such a choice is always somewhat arbitrary. In our view, the important point is that one chooses more than one

Table I. Reference limits from various sources

	This study	Free University Hospital Amsterdam[5]	University of Alabama Hospital[8]	Massachusetts General Hospital[14]
Reference sample	± 1000 patients, slight selection	?	± 13000 patients, no selection	?
Method	Percentiles 5 and 95 per cent	?	Percentiles 5 and 95 per cent	?
Albumin (g/l)	37–54	38–50	28–48	35–50
Total Protein (g/l)	63–79	60–75	58–83	60–84†
Calcium (mmol/l)	2·17–2·52	2·15–2·50	2·0–2·5	2·1–2·6
Phosphorus (mmol/l)	0·68–1·28*	0·8–1·3	0·78–1·68	1·0–1·5†
Creatinine (μmol/l) ♂	60–118	65–110	53–168	60–130
Creatinine (μmol/l) ♀			44–142	
Urea (mmol/l) ♂	2·9–8·9	3·0–7·5	2·2–11·5	2·9–8·9
Urea (mmol/l) ♀			1·4–9·4	
Uric Acid (μmol/l) ♂	240–500	280–480	215–565	180–420
Uric Acid (μmol/l) ♀	185–490	180–380	145–530	
Iron (μmol/l)	5–30	14–28	—	9·0–26·9

* After light breakfast
† Fasting

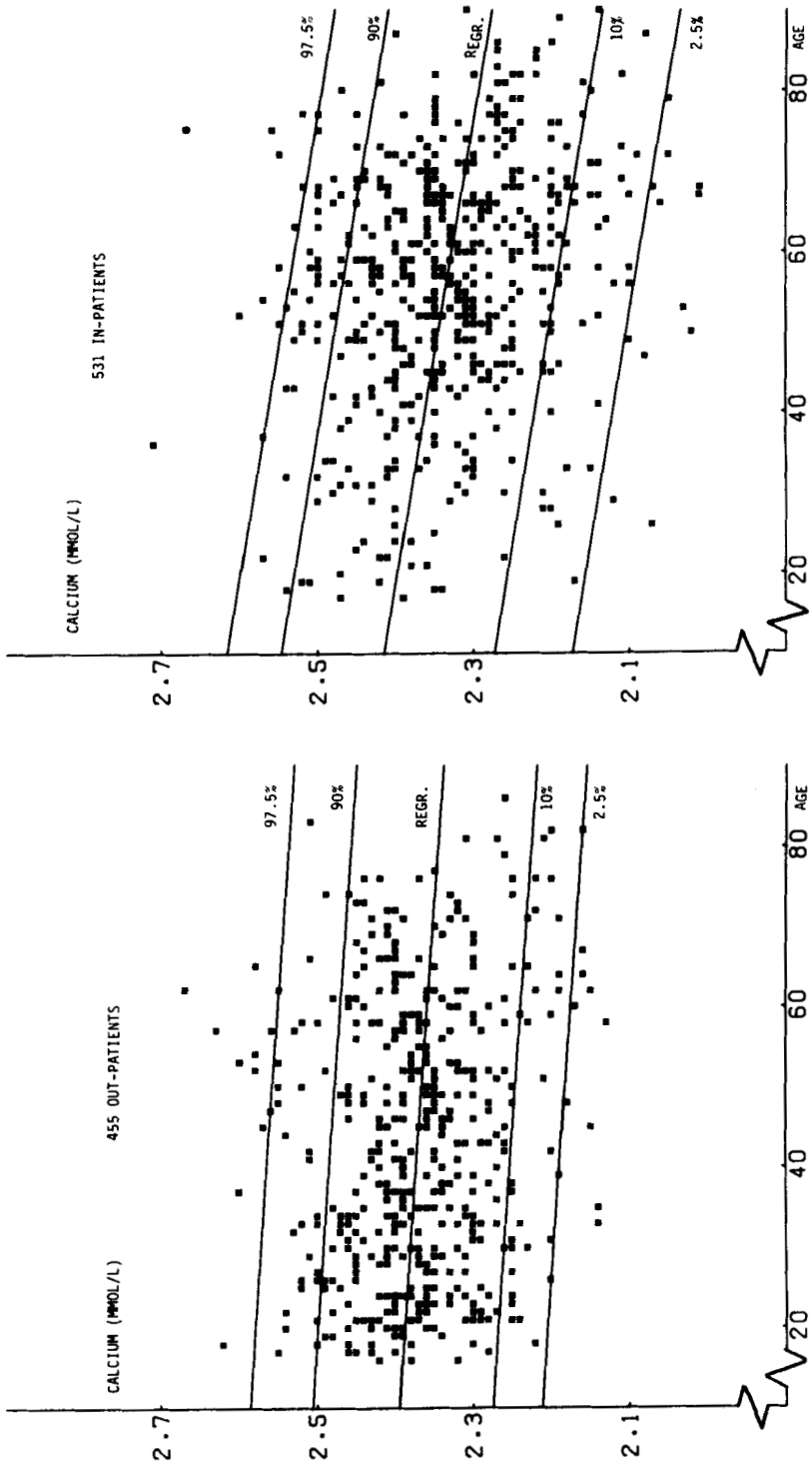

Figure 3. Linear regression and parallel percentile lines for calcium against age

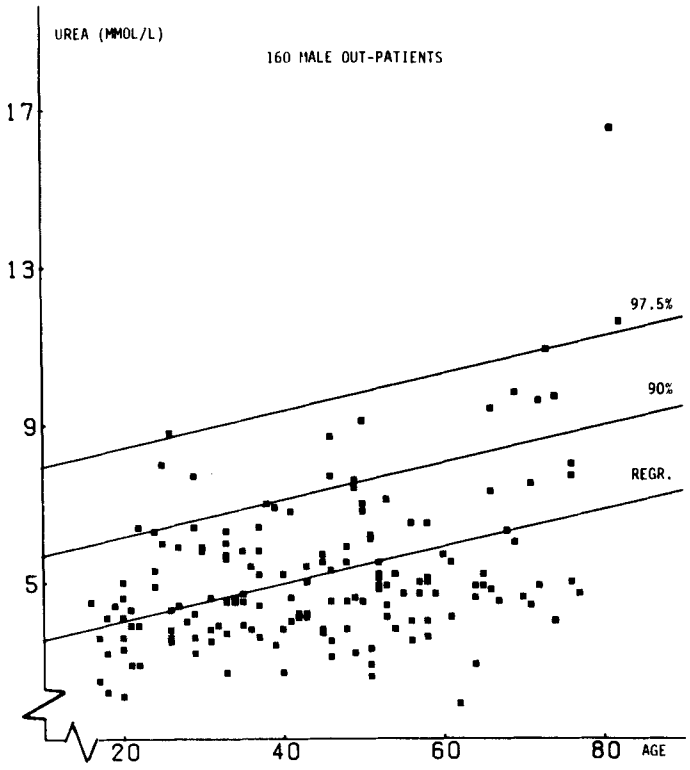

Figure 4(a)

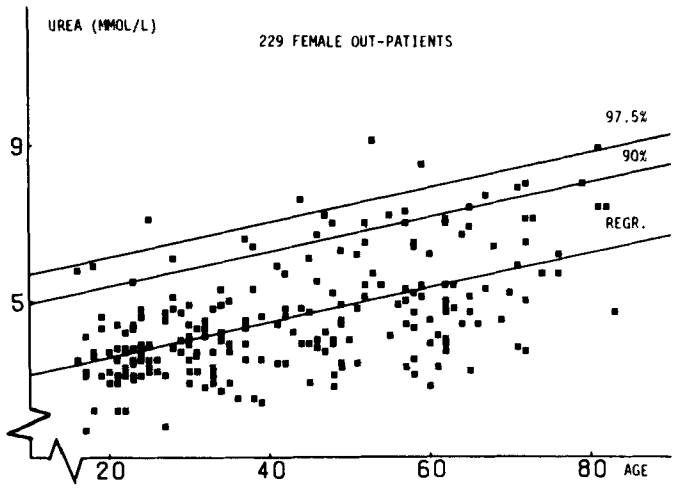

Figure 4(b)

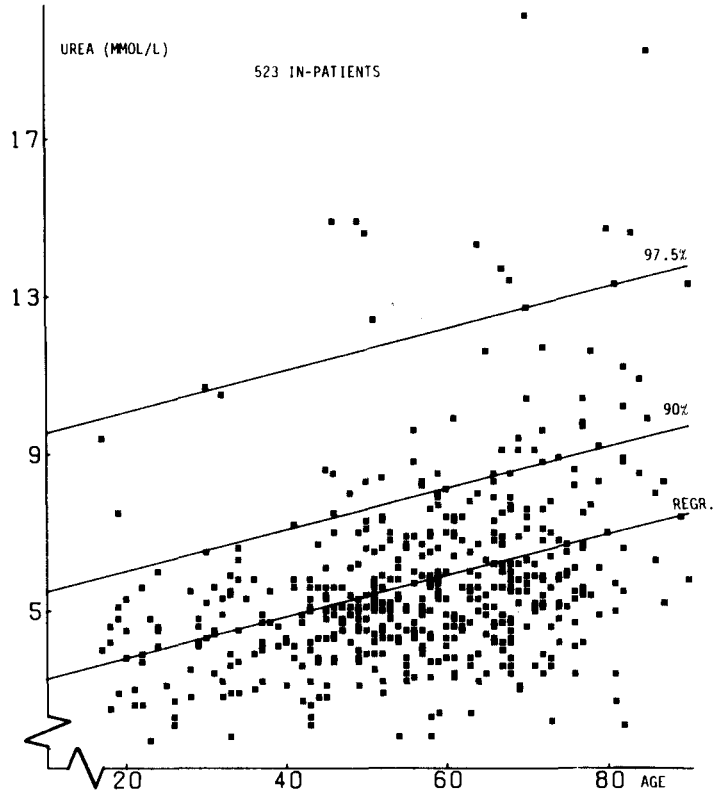

Figure 4(c)

Figure 4. Linear regression and parallel percentile lines (one-sided) for urea against age

limit (on either side). This stresses the fact that it is not one fixed limit that 'diagnoses the patient as normal or abnormal'. It is the physician who has responsibility for diagnosing his patient. Ideally, each value supplied by the laboratory should have an accompanying percentile value (in the reference distribution), so that the physician, together with other information concerning the patient, can make a decision. Provision of two percentiles is a step in that direction.

RESULTS

As indicated in the introduction, we shall discuss only a portion of the results of the study.

First, we determined the 5th and 95th percentiles for all patients (not the 2·5th and 97·5th percentiles because the reference sample consists of patients). Table I gives the results, together with those from the University of Alabama Hospital[8] (based on unselected patients), the limits reported for the Massachusetts General Hospital[14] and those in use at the AZVU.[5] The apparent differences are explainable by various factors such as region, reference population, statistical method, etc.

We list the following results:

(1) As expected, there were scarcely differences between the nine diagnosis classes. Also, we found no differences between caucasians and non-caucasians (about 10 per cent of the total).

(2) Covariance analyses with age as a covariable revealed significant differences in almost all cases between in- and out-patients and between men and women. However, the clinical relevance of the differences, especially between men and women, is sometimes questionable (the groups are relatively large). Differences between men and women have often been reported in the literature[15, 16]; those between in- and out-patients are interesting in relation to the choice of the reference population.

(3) Multiple linear regression on age, height and weight led to the conclusion that, for uric acid, weight is the one important variable and, for the other constituents (except phosphorus), age. These results, again, are in accordance with the literature.[9, 15, 16]

(4) We also examined the residuals in the regression analyses. Total protein, calcium and phosphorus appeared approximately normal; the other constituents, except albumin, exhibited positive skewness.

As an example, Figures 3 and 4 indicate the regression lines, together with parallel percentile lines, for calcium and urea. Comparison with the global limits in Table I reveals a much improved precision of the reference ranges by taking into account age, in- or out-patient, and sex (in the examples only for urea in the out-patients). We suggest consideration of these factors when determining reference ranges or, better, percentiles.

REFERENCES

1. Benson, E. S. 'The concept of the normal range', *Human Pathology*, **3**, 2 (1972).
2. McNeil, B. J., Keeler, E. and Adelstein, S. J. 'Primer on certain elements of medical decision making', *New England Journal of Medicine*, **293**, 211 (1975).
3. Wulff, H. R. *Rational Diagnosis and Treatment*, Blackwell, Oxford, 1976.
4. Elion-Gerritzen, W. E. 'Requirements for analytical performance in clinical chemistry', *American Journal of Clinical Pathology*, **73**, 183 (1980).
5. Academisch Ziekenhuis der Vrije Universiteit, *Referentiewaarden van klinisch-chemische bepalingen voor volwassenen*, Amsterdam, 1978.
6. Sunderman, F. W. 'Current concepts of "normal values", "reference values", and "discrimination values" in clinical chemistry', *Clinical Chemistry*, **21**, 1873 (1975).
7. Alström, T. *et al.* 'Recommendations concerning the collection of reference values in clinical chemistry and activity report Ch 1', *Scandinavian Journal of Clinical and Laboratory Investigation*, **35**, Suppl 144 (1975).
8. Harwood, S. J. and Cole, J. W. 'Reference values based on hospital admission laboratory data', *Journal of the American Medical Association*, **240**, 270 (1978).
9. O'Kell, R. T. and Elliott, J. R. 'Development of normal values for use in multitest biochemical screening of sera', *Clinical Chemistry*, **16**, 161 (1970).
10. Hoffmann, R. G. 'Statistics in the practice of medicine', *Journal of the American Medical Association*, **185**, 864 (1963).
11. Documenta Geigy. *Wissenschaftliche Tabellen*, Basel, 1968.
12. Bezemer, P. D. *Referentiewaarden*, Thesis, Free University, Amsterdam, 1981.
13. Elveback, L. R. 'How high is high? A proposed alternative to the normal range', *Mayo Clinic Proceedings*, **47**, 93 (1972).
14. Scully, R. E., McNeely, B. U. and Galdabini, J. J. 'Case records of the Massachusetts General Hospital', *New England Journal of Medicine*, **302**, 37 (1980).
15. Barnett, R. N. *Clinical Laboratory Statistics*, Little Brown, Boston, 1979.
16. McPherson, K. *et al.* 'The effect of age, sex and other factors on blood chemistry in health', *Clinica Chimica Acta*, **84**, 373 (1978).