Determination of reference limits: statistical concepts and tools for sample size calculation

Abstract: Reference limits are estimators for ‘extreme’ percentiles of the distribution of a quantitative diagnostic marker in the healthy population. In most cases, interest will be in the 90% or 95% reference intervals. The standard parametric method of determining reference limits consists of computing quantities of the form $\bar{X} \pm c \cdot S$. The proportion of covered values in the underlying population coincides with the specificity obtained when a measurement value falling outside the corresponding reference region is classified as diagnostically suspect. Nonparametrically, reference limits are estimated by means of so-called order statistics. In both approaches, the precision of the estimate depends on the sample size. We present computational procedures for calculating minimally required numbers of subjects to be enrolled in a reference study. The much more sophisticated concept of reference bands replacing statistical reference intervals in case of age-dependent diagnostic markers is also discussed.

Keywords: coverage proportion; non-linear regression; order statistics; quantile; reference band.

Introduction

Reference limits are an indispensable tool of medical diagnostics by means of quantitative laboratory tests. Without reference limits which have been established by means of well-grounded statistical rules and procedures, quantitative diagnostic methods would be of little if any use for clinical practice. Except for cases where clinical decision limits are available, comparison with reference limits is currently clearly the major approach to the clinical interpretation of laboratory data. Speaking in general terms, a result of a quantitative laboratory test is classified as being diagnostically suspect if it falls in a range of values which have to be considered as extreme, in the sense, that there is only a small proportion of subjects in the healthy population providing values which are still larger or smaller. The proportion of values in the reference population exceeding these reference limits are usually fixed at 5.0% or 2.5%, depending on whether deviations from the mean to only one direction are supposed to have diagnostic relevance, or both too small and too large values are potentially pathologic. The proportion of values measured in healthy subjects which do not fall outside the reference interval is called the coverage of the latter. The coverage proportion coincides with the specificity [1] of the corresponding diagnostic decision rule being equal to 95% when the reference limits are set in the way described before. Thus, if for a given diagnostic test higher specificity is needed the reference interval may be enlarged. It is important to note that observing a value beyond the reference limits must not be taken as definitely indicating that a subject suffers from the disease under consideration. Such values also occur in non-diseased individuals but at a low frequency.

In the following, we will focus on statistical considerations about the planning of studies to determine reference limits and in particular the estimation of the appropriate sample size. The present review should aid clinical researchers in the design of such studies. Detailed
considerations about the correct definition of reference individuals and their health status are outside the scope of this review. Both selection strategies and statistical criteria are potentially useful for avoiding inclusion of values from individuals in a reference sample who actually belong to a diseased subpopulation. Selection strategies are based on filters tailored for recognizing individuals of whom it has to be suspected that they suffer from a disease influencing the marker under investigation. Examples of such a filter are repeated hospitalization prior to blood sampling and extreme values of other diagnostic tests. Statistically, a mandatory step is the elimination of outliers. However, one has to be aware of the fact that non-inclusion of outlying values is only a necessary but not a sufficient condition for representativeness for the non-diseased population.

Basic options for determining reference limits

Parametric approach

Stated in statistical language, reference limits are quantiles or percentiles of the probability distribution which the diagnostic marker under evaluation follows in the population of non-diseased individuals. If this distribution is Gaussian with population mean μ and variance σ², the true reference limits are given by the numbers μ ± cσ, where in the one- and two-sided case c has to be set equal to 1.645 and 1.960, respectively. It should be noted that the parametric approach requires a normal distribution either of the primary measurement values or some suitable transform of them. The statistical literature contains a wealth of different proposals for checking normality of distributions. Among the most recommendable criteria is a combination of the skewness and kurtosis statistics – for details see [2]. As holds generally true in parametric statistics, μ and σ² are unknown constants which have to be estimated from empirical data. Assuming that one has to deal with a single diagnostic marker X which does not depend on additional variables (in particular on age), the data obtained from a reference value study consists of a list X₁,...,Xₙ of n values measured in subjects that should be ideally free of the disease to be diagnosed by means of X. Estimating the reference limits simply requires substituting the sample mean X̄ and the empirical standard deviation S for their population analogs μ and σ using X̄−c⋅S and X̄+c⋅S as empirical lower and upper reference limit, respectively. A crucial point to be observed in carrying out these very simple computations is that the factor c is the same for the reference limits estimated from empirical data as for the theoretical limits which would have to be used if the underlying population were completely known. This means, in particular, that c need not be replaced with a quantity which depends on the sample size as has to be used in computing confidence limits for μ [3] which depend on the standard error of the mean rather than the standard deviation of a single observation. Actually, a reference interval is in general substantially wider than the corresponding confidence interval (CI).

Nonparametric estimation

Whenever the distribution of X exhibits features which are not consistent with the Gaussian model and cannot be removed through applying some known transformation (e.g., taking logarithms), estimation of reference limits should be based on nonparametric techniques. With a distribution of any shape, the upper reference limit for the population is obtained through inverting the cumulative distribution function F(x). Precisely speaking, this means that, given the value of the coverage proportion q (usually chosen as q = 0.95 or q = 0.975), one has to find the smallest value of X which satisfies the condition F(x) ≥ q. Denoting the solution by x_q, the process of determining x_q is easily visualized graphically (see Figure 1A). Assuming F(x) to be continuous, the lower reference limit for the population, say x̄_q, is obtained through performing the same step with the function 1−F(x) instead of F(x) itself (Figure 1B). Adopting this rule for the computation of reference limits with the data obtained from a sample of non-diseased subjects only requires substituting the empirical for the theoretical population distribution function, taking into account that the former will never be continuous but a step function increasing from left to right. As will be illustrated in the subsequent example, computing the nonparametric estimates of x_q and x̄_q eventually reduces to sorting the measured values X₁,...,Xₙ in ascending order and picking from the resulting list of ordered values some specific element to be identified by means of a simple rule as conforming with the pre-specified coverage proportion q.

An illustrating example for a standard application of both estimation procedures

The following illustration uses parts of the data set published in Appendix 4.2 of the book by Harris and Boyd [4]. After elimination of a few data lines containing outliers it consists of the results of performing routine clinical chemistry in n = 593 male medical students. Both estimation
Procedures will be demonstrated for blood urea nitrogen (BUN), serum creatinine (SCR) and uric acid (UA) (given in mg/dL). Computing the standard descriptive statistics with these data gave the values shown in Table 1.

Although one-sided reference intervals for these markers will rarely be considered in practice, their computation will be demonstrated here as well.

**Parametric estimates for $q=0.95$**

With the data of Table 1, the lower and upper one-sided reference limit for BUN is computed to be $\bar{X} - 1.645 \cdot S = 15.3 - 1.645 \cdot 3.347 = 9.794$ and $\bar{X} + 1.645 \cdot S = 15.3 + 1.645 \cdot 3.347 = 20.806$, respectively. Calculating the limits of the two-sided reference interval of coverage 95% is likewise fairly simple: The left-hand limit has value $\bar{X} - 1.96 \cdot S = 15.3 - 1.96 \cdot 3.347 = 9.071$ and the right-hand limit has value $\bar{X} + 1.96 \cdot S = 15.3 + 1.96 \cdot 3.347 = 20.938$.

**Table 1** Basic summary statistics for the distribution of three markers of renal dysfunction in a sample of $n=593$ male medical students (data from 4, Appendix 4.2 [4]; all values given in mg/dL).

<table>
<thead>
<tr>
<th></th>
<th>BUN</th>
<th>SCR</th>
<th>UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>15.3</td>
<td>1.07</td>
<td>6.24</td>
</tr>
<tr>
<td>Median</td>
<td>15.0</td>
<td>1.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.347</td>
<td>0.145</td>
<td>1.144</td>
</tr>
</tbody>
</table>
15.3–1.96 \cdot 3.347 = 8.740, and its right-hand counterpart is obtained as 15.3+1.96 \cdot 3.347 = 21.860. For SCR and UA, performing the same calculations yields the other entries in the Table 2.

**Nonparametric estimates for \( q = 0.95 \)**

With the same specification of the target coverage proportion as in the parametric case and \( n = 593 \), there holds \( n \cdot q = 563.35 \) and \( n \cdot ((1+q)/2) = 578.175 \). Rounding these numbers up to the nearest integer yields for the ranks of the order statistics determining the upper nonparametric 95% reference limits in the one- and two-sided cases the values \( r_1 = 564 \) and \( r_2 = 579 \), respectively. The corresponding ranks for determining the lower reference limits are \( r_1 = n+1-r_1 = 594–564 = 30 \) and \( r_2 = n+1-r_2 = 594–579 = 15 \). All what remains to do is to identify the order statistics associated with these ranks in the list (being not reproduced here) of the ascending sorted measurement values, yielding the entries in Table 3. An alternative approach to determining the nonparametric estimates of all reference limits of interest as suitable order statistics actually observed in the sample is the so-called interpolation method (averaging the values of adjacent order statistics). However, the latter is not in line with basic statistical usage and has the disadvantage of producing results with more significant decimal figures than contained in the individual measurements.

**Sample size planning of reference value studies**

Until fairly recently, the topic of sample size calculation for reference value studies was not given much attention in the biostatistical literature. In view of this it is not surprising that in the practice of laboratory medicine, sample size planning is still mostly done by means of rather crude rules of thumb. One of these rules specifies that \( n \) should not be smaller than 120 since otherwise the result is obtained as 15.3+1.96 \cdot 3.347 = 21.860. For SCR and UA, performing the same calculations yields the other entries in the Table 2.

**Table 2** Results of parametric reference limit estimation with the data of Table 1 (all values in mg/dL).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Left</th>
<th>Right</th>
<th>Two-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCR</td>
<td>0.831</td>
<td>1.309</td>
<td>[0.786, 1.354]</td>
</tr>
<tr>
<td>UA</td>
<td>4.358</td>
<td>8.122</td>
<td>[3.998, 8.482]</td>
</tr>
</tbody>
</table>

**Table 3** Results of nonparametric reference limit estimation with the data of Table 1 (all values in mg/dL).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Left</th>
<th>Right</th>
<th>Two-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUN</td>
<td>10.0</td>
<td>21.0</td>
<td>[10.0, 22.0]</td>
</tr>
<tr>
<td>SCR</td>
<td>0.8</td>
<td>1.3</td>
<td>[0.8, 1.4]</td>
</tr>
<tr>
<td>UA</td>
<td>4.4</td>
<td>8.3</td>
<td>[4.0, 8.7]</td>
</tr>
</tbody>
</table>

exact nonparametric CI for the lower limit of the 95% reference interval cannot be calculated [5]. Another proposal being, from a statistician's viewpoint, likewise largely ad hoc requires \( n \) to be determined in a way ensuring that the width of the 90% CI for either one of both reference limits does not exceed some specified proportion (e.g., 20%) of the length of the theoretical reference interval for the underlying population [6, 7]. None of these criteria enables one to ensure that the coverage proportion provided by the reference interval estimated from the sample to be drawn comes satisfactorily close to the target value \( q \) with high probability. An approach being tailored to fill this gap has been developed in [8]. The following description of the rationale of the procedure (see also the flow-chart shown below as Figure 2) primarily applies to the planning of studies aiming to establish an upper reference limit.

Step 1: Specification of tolerances \( \delta \) and \( \delta' \), for the deviation of the actual coverage proportion from its target value \( q \) to the left and right, respectively. If, for example, one is willing to accept that the proportion of values in the underlying population covered by the estimated reference region becomes as small as 94% and 96% at most with \( q \) chosen as 95%, this would mean that both \( \delta' \)'s were set equal to 1%. Although in principle there is no need to set both \( \delta' \)'s to a common value, we assume in the sequel that the same difference is to be tolerated in both directions so that the subscript will be omitted. A good guidance for choosing the numerical value of \( \delta \) is to keep fixed the ratio to \( 1-q \) at the same value, implying, that if ±0.01 is thought to be appropriate for \( q = 0.95 \), in another application with \( q = 0.99 \), one should specify \( \delta = 0.002 \), and so on. A much more liberal specification is to set both \( \Delta \)s equal to 0.04 when the target coverage is \( q = 0.95 \). Then, the proposed criterion essentially ensures with high probability that the specificity which should be 95%, neither exceeds that of a reference interval with coverage 99% nor fails to be at least as large as that of an interval with one-sided coverage 90%. As can be seen from the entries in the first row of Table 4 the sample sizes required in order to satisfy the criterion in this weakened form are only about half as
large as the traditionally recommended lower bound of 120 even when the nonparametric estimation procedure shall be used.

Step 2: Specifying a lower bound \( \beta \) (e.g., 90%) to the probability that the coverage proportion attained by the estimated reference region complies with the tolerances specified in Step 1. In [8], this probability whose role in planning a reference value study is largely analogous to that of power in statistical hypothesis testing, is termed ‘confidence probability’ (not to be confused with the confidence level associated with a CI for some distributional parameter [3]).

Step 3: Exploiting the functional relationship between the confidence probability and the sample size for determining by means of a simple iteration process the minimum value of \( n \) required to guarantee that the former becomes at least as large as specified in Step 2.

Step 3 can be carried out in a numerically exact manner. A selection of exactly calculated sample sizes for \( \beta = 0.90 \) and common value \( \delta \) of the tolerances \( \delta_1 \) and \( \delta_2 \) is shown in Table 4. As has to be expected, it turns out that given the value of the target coverage \( q \) and the pre-specified lower limit \( \beta \) to the confidence probability, the required sample size increases with decreasing tolerance \( \delta \). Furthermore, ceteris paribus, the sample size required for a reference value study to be evaluated by means of the nonparametric method, is up to 2.5 times larger than that required in the parametric case.

Alternatively, Step 3 can be carried out by means of the approximation formulae

\[

n = (1 + 0.5 z_j^2) / \phi z_j / \delta^2 \quad [ \text{parametric case}], \\
\quad n = q(1 - q) z_j^2 / \delta^2 \quad [ \text{nonparametric case}].
\]

In these approximate equations, \( z_j \) and \( z_j \) stands for the \( 100 \cdot (1-\text{q}) \) and \( 100 \cdot (1+\beta) / 2 \) – percentage point of the standard normal distribution, respectively, and \( \phi \) denotes the ordinate of the standard normal density function at \( x = z_j \).

Remarkably, the sample sizes calculated by means of the approach described in the preceding paragraph hold likewise for studies aiming to establish a lower instead of an upper reference limit. If the focus lies on reference intervals bounded on both sides, the following procedure can be applied: The same formulae as in the one-sided case are used, but with halved tolerances and \( q \) and \( \beta \) replaced with \( (1+q)/2 \) and \( (1+\beta)/2 \), respectively. According to this latter rule, the values of \( n \) appearing in the third line from bottom of Table 1 (i.e., \( n = 272 \) and \( n = 636 \)) are sufficient for determining a two-sided reference interval of coverage \( q = 95\% \) with tolerances \( 2\% \) and confidence probability of \( \beta = 80\% \) at least.

**Table 4** Results of exact sample size calculation by means of the approach of Jennen-Steinmetz and Wellek [8].

<table>
<thead>
<tr>
<th>( q )</th>
<th>( \delta )</th>
<th>( \delta/(1-q) )</th>
<th>( n_{par} )</th>
<th>( n_{nonpar} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>0.0400</td>
<td>0.8</td>
<td>46</td>
<td>62</td>
</tr>
<tr>
<td>0.950</td>
<td>0.0300</td>
<td>0.6</td>
<td>77</td>
<td>125</td>
</tr>
<tr>
<td>0.950</td>
<td>0.0250</td>
<td>0.5</td>
<td>110</td>
<td>196</td>
</tr>
<tr>
<td>0.950</td>
<td>0.0200</td>
<td>0.4</td>
<td>171</td>
<td>305</td>
</tr>
<tr>
<td>0.950</td>
<td>0.0150</td>
<td>0.3</td>
<td>302</td>
<td>559</td>
</tr>
<tr>
<td>0.950</td>
<td>0.0100</td>
<td>0.2</td>
<td>678</td>
<td>1276</td>
</tr>
<tr>
<td>0.975</td>
<td>0.0150</td>
<td>0.6</td>
<td>123</td>
<td>257</td>
</tr>
<tr>
<td>0.975</td>
<td>0.0125</td>
<td>0.5</td>
<td>176</td>
<td>397</td>
</tr>
<tr>
<td>0.975</td>
<td>0.0100</td>
<td>0.4</td>
<td>272</td>
<td>636</td>
</tr>
<tr>
<td>0.975</td>
<td>0.0075</td>
<td>0.3</td>
<td>482</td>
<td>1153</td>
</tr>
<tr>
<td>0.975</td>
<td>0.0050</td>
<td>0.2</td>
<td>1081</td>
<td>2607</td>
</tr>
</tbody>
</table>
Example illustrating the use of the approximation formulae

Let us assume that the protocol of a reference value study under planning specifies \( q = 0.95, \delta = 0.01, \beta = 0.90 \). The \( z \)-values appearing in the formulae are both equal to 1.645, then, and \( \varphi \) is computed to be \( \varphi = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{1}{2} \cdot (z^2 - 1)} = 0.1031 \).

Plugging in these values yields:

\[
(1+0.5 \cdot 1.645^2) \cdot (0.1031 \cdot 1.645 / 0.01)^2 = 676.82, \tag{1}
\]

\[
0.05 - 0.95 \cdot (1.645 / 0.01)^2 = 1285.36. \tag{2}
\]

After rounding them up to the next integer the approximated sample sizes are the same as the exact values to be read from Table 4. Since the same happens in a multitude of other instances, it can be concluded that the accuracy of the approximation formulae is very good. (For a choice of additional values of \( q \) and \( \beta \), the values of the constants required for applying the approximation formulae are provided in Supplemental Data, Table 1, which accompanies the article at http://www.degruyter.com/view/j/cclm.2014.52.issue-12/issue-files/cclm.2014.52.issue-12.xml.)

Reference regions for age-dependent diagnostic markers

In not a few clinical disciplines, it is more the rule than the exception that the reference limits of a diagnostic marker have to depend on age or some other covariable of the continuous type. As becomes obvious from the example shown in Figure 3A, using the same reference limits for all ages would lead to grossly misleading conclusions. Of the PAPP-A concentrations measured later than at a gestational age of 90 days, only a single one would be classed as substantially less than normal whereas for younger gestational ages, the proportion of such values would be almost twice as large as the usually required false positive rate of 5%.

In principle, it is fairly clear of what shape a ‘reasonable’ reference region should be in the situation of Figure 3A. The boundaries of the reference band must be given by curves which increase from left to right, and the width of the band should likewise be an increasing function of time since the same holds true for the variance of the measurement values. However, making these notions precise in terms of a statistical algorithm is a comparatively demanding task involving computationally intensive techniques. In the original biostatistical literature several approaches to the construction of age-dependent reference centiles have been proposed which differ distinctly both in terms of the underlying principles and the algorithms required for their implementation. Most frequently applied are smoothed age-specific \( \bar{X} \pm 1.96 \cdot S \) ranges for normally distributed observations [10, 11] or data which can be normalized through applying a Box-Cox type transformation [12, 13], procedures based on the nonparametric quantile regression [14], and a strictly distribution-free construction which ensures that the proportion of values covered by the band in the empirical joint distribution of age and the diagnostic marker of interest always coincides with the pre-specified target coverage proportion \( q \) [15]. In this introductory exposition, we confine ourselves to give an outline of the last of these approaches which has successfully been applied in a multitude of settings, in particular within the field of prenatal medicine (cf. [16]).

Major steps of a distribution-free construction of reference bands for age-dependent diagnostic markers

Step 1: Fitting a linear regression model in order to determine a central line around which the band is to be spanned. The form of this regression function must reflect the way in which the age-specific means change over time. In absence of specific knowledge about the mechanism which underlies this process, an ordinary polynomial of sufficiently high degree will usually be appropriate.

Step 2: Determining the ratio between the bandwidth at the beginning, around the middle and at the end of the interval within which the measurements were taken, in a way which reflects possible differences in variability between these parts of the time range.

Step 3: Calculating (by means of an iteration algorithm) the minimum width of a band of the form resulting from Steps 1 and 2 guaranteeing coverage of at least \( q = 90\% \) of the data points in the sample. In cases where the distribution of the measurement values lacks symmetry, the proportions of values lying above and below the band are controlled separately, subject to the condition that both of them are approximately equal to 5%. Thus, the ordinates of the points on the upper and lower boundary will correspond to smoothed 5th and 95th percentiles, respectively. For the PAPP-A data of Figure 3A, the 90% reference band obtained by means of the distribution-free approach is shown in Figure 3B.
Discussion

Even when the diagnostic marker under consideration is technically easy to measure, acquiring a database of sufficiently large size for a reference value study is often a serious problem. Diagnostic laboratories usually have easy access to blood samples from patients. However, the gold-standard design for a reference value study requires persons free from certain diseases or having at least a well-defined health status and showing sufficient variability with respect to age and sex to be recruited. However, in practice, appropriate and in particular complete information regarding the health status is in most cases of out-patients and in-patients not available. The time and effort to be spent on obtaining the data needed and to remove diseased individuals from such a mixed population seems often forbidding. Students as reference individuals have the disadvantage of the limited age range. Thus, there are many obstacles in recruiting subjects for reference value studies making sample size planning a major issue, except one explicitly admits to base reference
limit estimation on samples consisting of mixtures of diseased and non-diseased individuals. Unfortunately, such an indirect approach (cf. [17]) necessarily requires to rely on model assumptions which are not open to direct empirical validation.

Concerning the question of whether or not it is admissible to tolerate in a study to be evaluated by means of direct methods a certain amount of values measured in diseased individuals rather than healthy controls, no general recommendations can be given. Nevertheless, it is clear that the proportion of values from diseased individuals should be as low as possible. In order to protect oneself against the potential biases entailed in estimating reference limits with samples including also diseased individuals, it is advisable to rely on statistical methods which, like the nonparametric estimation procedure described above, are robust against outliers. However, it is worth highlighting that the distributions which a diagnostic marker follows in diseased individuals and healthy controls, respectively, might exhibit considerable overlap in which case outlier robustness of the estimation procedure will not be sufficient for solving the problem. Moreover, our results on sample size calculation imply that the price to be paid for using nonparametric rather than parametric methods of reference limit estimation is substantial. In view of this, it is important to note that given biases due to inclusion of values from diseased individuals is not an issue, application of standard parametric methods is also admissible when the assumption is warranted that the distribution of the diagnostic marker under consideration is not normal itself but can be normalized through applying some mathematical transformation (e.g., the log transformation) to the primary measurements. Whenever this holds true, it is recommendable to calculate reference limits on the transformed scale by means of parametric methods and eventually retransform the results by means of the inverse of the respective transformation (taking, e.g., \( \exp(Y + 1.96 S) = e^{\hat{\theta}_{0.12} + 1.96 \cdot 0.83} = 5.74 \) as upper reference limit of coverage 97.5\% in a setting where the logarithms \( Y_i = \log(X_i), \ldots, Y_n = \log(Y_n) \) of sample values yield an arithmetic mean and standard deviation of 0.12 and 0.83, respectively). The correctness of sample size planning by means of the above formulae and tabulated results remains invariant under this transformation/retransformation procedure.

It is also worth noticing that the methods described in this article only apply to studies dealing with a single quantitative marker. In order to account for the often more realistic scenario that in diagnosing some given disease several different markers are considered taking a positive diagnostic decision if at least one of these markers turns out to exceed its reference limits, multivariate reference regions of ellipsoidal (see, e.g., [18]) or rectangular shape [19] are needed. Such multivariate reference regions enable one to control, in analogy with multiple testing of statistical hypotheses (cf. [20]), the ‘multiple risk of a false positive diagnostic decision’ which should not exceed the usual bound of 5\% (corresponding to a ‘multiple specificity’ of at least 95\%).

Another generalization of the methods discussed in this paper is required when an age-independent marker is repeatedly measured at several different points of time so that the data upon which the diagnostic decision relies, consist of longitudinal profiles. As was shown in [21], the distribution-free procedure used to construct the reference band shown in Figure 3B can be modified in such a way that it yields reference bands for longitudinal data as well.

A currently open question is whether and in what way the criterion of estimation precision underlying Table 1 can be adopted for the estimation of age-dependent reference centiles and the sample size planning of studies dealing with age-dependent diagnostic markers.

As a concluding remark, we point out that there are reasons to expect that the majority of diagnostic markers to be assessed in laboratory medicine actually depend in a systematic way on age (or some other covariable). In view of this, it seems tempting to examine also those laboratory parameters which traditionally are assessed by comparing them for all ages with the same reference limits, for the possibility of improving the diagnostic accuracy attainable by means of them through using a reference band of the kind constructed in the preceding section for a sample of PAPP-A values.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Financial support: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References


**Supplemental Material:** The online version of this article (DOI 10.1515/ccm-2014-0226) offers supplementary material, available to authorized users.

Stefan Wellek served as a Professor of Biostatistics at the University of Heidelberg and Head of the Department of Biostatistics at the Central Institute of Mental Health Mannheim from 1994 to the present. In addition, he has been affiliated with the Department of Medical Biostatistics, Epidemiology and Informatics of the University of Mainz as adjunct senior research professor since 2010. He took his PhD in mathematical and his ‘habilitation’ in medical statistics (from the universities of Düsseldorf and Mainz, Germany). He is the author of a research monograph on statistical equivalence testing methods and published numerous papers on a broad range of topics (including methods of reference limit estimation) in biostatistical and medical journals.

Karl J. Lackner studied medicine at the University of Heidelberg and received his MD degree in 1982. After a 2½-year postdoctoral stay at the National Heart, Lung, and Blood Institute in Bethesda, Maryland he began his training in internal medicine and cardiology at the university hospital in Heidelberg which he completed in 1991. Thereafter, he moved to the university hospital in Regensburg where he specialized in laboratory medicine. In 1995 he was board certified as a specialist in laboratory medicine. In 1996 he got his certification as clinical chemist. Since 2001 he has been Director of the Institute of Clinical Chemistry and Laboratory Medicine at the University Medical Centre in Mainz. His major research interest is in cardiovascular disease and hemostaseology. From 2007 to 2011 he served two terms as president of the German United Society of Clinical Chemistry and Laboratory Medicine (DGKL).
Christine Jennen-Steinmetz studied mathematics at the Faculty of Mathematics of Heidelberg University and received her PhD degree from the same faculty. Since 1983 she has been working at the Department of Biostatistics at the Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg. She provides statistical and methodological expertise to the research projects of the departments represented at this institute. She has acquired experience in statistical modeling and data analysis. Furthermore, she has expertise in statistical computation and simulation. During the last years her research has focused on statistical methods for reference value studies.

Iris Reinhard is a member of the Biostatistics group of the Central Institute of Mental Health, Mannheim, Germany. She received her MSc in Mathematics from the University of Heidelberg, Germany, in 1990. Her areas of interest include mixed models, generalized linear models, survival models and the construction of reference regions among other fields. She was involved in projects in the field of Gynecology, Addiction Medicine, Psychiatry and Psychosomatics.

Isabell Hoffmann is a postgraduate fellow in the Biostatistics Unit of the Department of Medical Biostatistics, Epidemiology and Informatics of the University of Mainz, Germany. She received her Diploma in Statistics from the Technical University of Dortmund, Germany, in 2010. Currently, her main research interests relate to the analysis of interactions in models for high-dimensional data.

Blettner majored in statistics/economics/social sciences at the University of Dortmund, Germany. After her diploma in Statistics in 1978, she worked at the University of Dortmund, at the WHO in Lyon, and at the National Cancer Institute Bethesda/USA, where she completed her PhD thesis in 1987. After being a lecturer in medical statistics at the University of Liverpool, she joined the Germany Cancer Research Centre (DKFZ) in Heidelberg as Assistant Professor. From 1999 until 2003 she was full Professor and Chair of the Department of Epidemiology and Medical Statistics and Dean of the School of Public Health at the University of Bielefeld (2001–2003). In 2003 Prof. Blettner became Director of the Institute of Medical Biometry, Epidemiology and Informatics at the Johannes Gutenberg-University of Mainz. Her main research areas are statistical methods in epidemiology and radiation and cancer epidemiology.