Laboratory Management **Edited by: E. Wieland**

Georg Hoffmann*, Ralf Lichtinghagen and Werner Wosniok

Simple estimation of reference intervals from routine laboratory data

DOI 10.1515/labmed-2015-0104
Received September 3, 2015; accepted October 22, 2015

Abstract: According to the recommendations of the IFCC and other organizations, medical laboratories should establish or at least adapt their own reference intervals, to make sure that they reflect the peculiar characteristics of the respective methods and patient collectives. In practice, however, this postulate is hard to fulfill. Therefore, two task forces of the DGKL ("AG Richtwerte" and "AG Bioinformatik") have developed methods for the estimation of reference intervals from routine laboratory data. Here we describe a visual procedure, which can be performed on an Excel sheet without any programming knowledge. Patient values are plotted against the quantiles of the standard normal distribution (so-called QQ plot) using the NORM. INV function of Excel. If the examined population contains mainly non-diseased persons with approximately normally distributed values, the respective dots form a straight line. Very often the values are rather lognormally distributed; in this case the straight line can be detected after logarithmic transformation of the original values. Values, which do not match with the assumed theoretical distribution, deviate from the linear shape and can easily be identified and eliminated. Using the reduced data set, the mean value and standard deviation are calculated and the reference interval ($\mu \pm 2\sigma$) is estimated. The method yields plausible results with simulated and real data. With the increasing number of results, which do not match with the model, it tends to underestimate the standard deviation. In all cases, where the QQ plot does not yield a substantial linear part, the proposed method is not applicable.

***Correspondence: Prof. Dr. med. Georg Hoffmann,** Trillium GmbH, Medizinischer Fachverlag, Jesenwanger Strasse 42b, 82284 Grafrath, Germany, Tel.: +49-(0)-8144/93905-0, Fax: +49-(0)-8144/9390-29, E-Mail: georg.hoffmann@trillium.de
Ralf Lichtinghagen: Institute of Clinical Chemistry at Hanover Medical School (MHH), Hanover, Germany
Werner Wosniok: Institute of Statistics, University of Bremen, Bremen, Germany

Keywords: indirect estimation; MS Excel; quantile quantile plot; reference interval.

Introduction

Most laboratory medical decisions are based on the comparison of measured values with reference intervals. These determine which results are to be classified as "normal" – or more precisely, which people are "not ill". Given its enormous relevance for diagnostics, prognostics, and follow-up/treatment monitoring, one would assume that the term reference interval should have been defined clearly and that the procedures for determining decision limits should have been established precisely. In theory, this is true according to the current IFCC/CLSI guideline from 2008 [1]:

- The reference interval is defined as the central 95% range of observed values of an apparently healthy population.
- Each medical laboratory must determine its own reference intervals for all analytes in use on the basis of at least 120 reference individuals by way of the 2.5 and 97.5 percentiles, or must at least check existing information on the basis of 20 healthy controls.

According to DIN EN ISO 15189, each laboratory must document and communicate to users the foundations of the corresponding decision values [2].

But these specifications are often deviated from in real life. On the one hand, the term "healthy" is imprecise (for example, concerning the elderly), and on the other hand, medical laboratories are required to examine suitable reference individuals for hundreds, if not thousands, of tests, which appears to be unrealistic and – in the case of healthy newborns – even unethical [3, 4]. The requirement of 120 reference individuals may seem reasonable at first sight, but this number must be multiplied by 2 for gender-based values and by even higher factors with respect to age, so that the total number of individuals may indeed reach a 1000 or more. As a result, most laboratories

Table 1: Reference intervals published on the Internet for common parameters (as of 05.10.2015).

Source	Sodium, mmol/L	Calcium, mmol/L	Fibrinogen, mg/dL	Total protein, g/dL
DocCheck.com	135–148	2.20–2.65	180–350	6.6–8.3
Laborlexikon.de	135–145	2.02–2.60	150–450	6.1–8.1
Thieme.de	135–150	2.3–2.6	200–400	6.0–8.4
Wikipedia.org	135–145	2.2–2.6	180–350	6.0–8.0

accept the information provided by the manufacturers of diagnostic products or other authors without the necessary verification and documentation. This can produce substantial discrepancies between different sources [5]. For example, the reference intervals for electrolytes or proteins stated on much-visited Internet sites differ by as much as 30% (Table 1).

One way out of this dilemma is to resort to indirect methods using routine laboratory results for estimation of reference intervals. The underlying idea is to eliminate values that do not fit the expected distribution based on model hypotheses. Then, theoretic percentiles are calculated from the remaining values. The foundation for this method was laid as early as around 1960 (see overview in Ref. [6]), but it was not until the arrival of the personal computer in the 1980s that enough computing power was available to apply it generally.

Worthy of special mention here is the reference limit estimator (RLE) method published by Arzideh et al. in 2007 [7]. It was developed by the "Reference Values Working Group" of the German Society for Clinical Chemistry and Laboratory Medicine (DGKL) by means of the statistical software package R. This mathematically sophisticated method is based on transformed raw data, calculates the theoretical density function at the center of the distribution, and eliminates potential outliers due to local deviations from this function at "truncation points". The group succeeded in estimating the reference values of enzymes [7], electrolytes [8] and blood count parameters [9], which was also evaluated in multicenter studies [10, 11]. But the RLE method requires very large datasets of several 1000 cases, which are not always available. In addition, the software package R often poses an obstacle to users without experience in computer science and statistics.

This is why we evaluated less sophisticated methods, which should be manageable with little effort and without programming experience. In a joint activity of the DGKL working groups Reference Values and Bioinformatics [12], we focused in particular on the original methods of J. Pryce 1960 [13] and R. Hoffmann 1963 [14] that were based on real data with theoretical normal distributions, and performed then without computers, using "probability paper" instead [15]. In this paper, we present a method where the visual approach is simulated via Microsoft Excel [16], and compare it to another "modified Hoffmann method" [4] from the more recent literature.

Materials and methods

Analysis results from 246 healthy test subjects from a study done by the Hanover Medical School (MHH) were available for the evaluation of the method. The foundations of the study are described in Ref. [17]. The reference for the comparison of methods was the RLE method of the Reference Values Working Group [7]; Dr. F. Arzideh, University of Bremen, supplied validated patient datasets.

The calculations shown in Figures 1 and 9 were done using the English version of R; the others were perfomed with the German version of Microsoft Excel. There were no fundamental differences between the different Excel versions. Only when calling the function NORM.INV (see below), one must keep in mind that there was no period in older versions (prior to 2010) (NORMINV instead of NORM.INV). To simplify the generation of random numbers, histograms, etc., one should install the "Analysis ToolPak" add-in included in Excel (for instructions, see http://support.office.com, for example).

Probability plots ("probability paper") for the visual check of distributions without the use of computers are available on the Internet (for example, http://gpr.physik.hu-berlin.de/Downloads/Papiere.html). This is a kind of graph paper whose scale on the Y-axis has been adapted to a density function. In the case of Gaussian bell-shaped curve, the lines are thus the densest in the center at a probability of 50%, and drawn far apart at the margins beyond 1% and 99%.

For the reference method [7], the freely accessible statistics package R (www.r-project.org) with the add-on packages *geoR* and *msm* was used. To facilitate the use, the Reference Values Working Group developed an Excel frontend called "Reference Limit Estimator", which is also available online (www.dgkl.de, look for "Arbeitsgruppe Richtwerte" under the menu item "AGs & Sektionen" – only available in the German version of the website).

Practical implementation

The following section is to help readers reproduce the results presented and to enable them to run their own

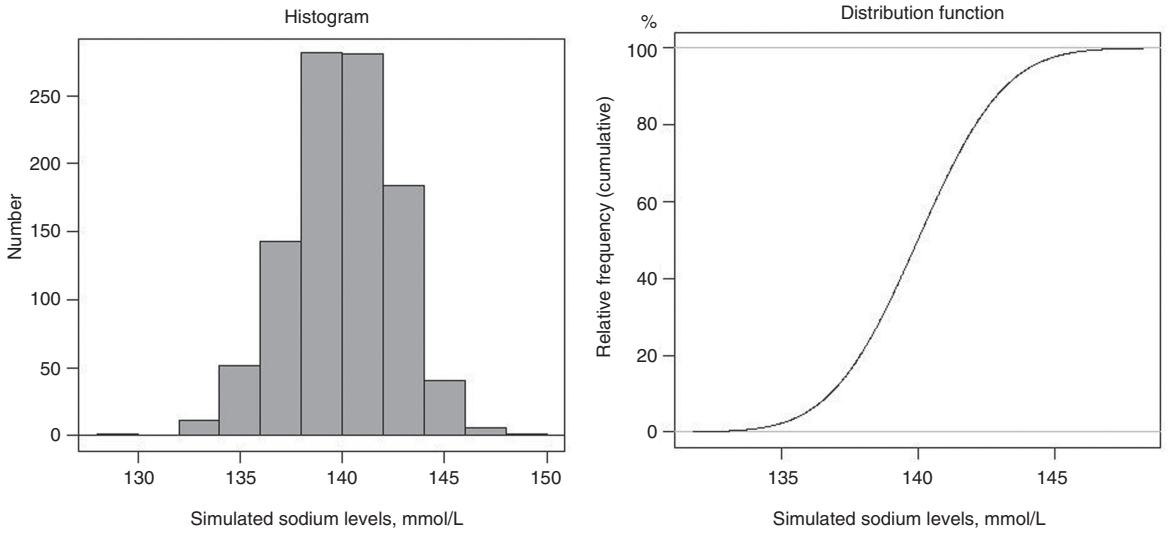

Figure 1: On the left, a histogram of normally distributed sodium levels (mean 140 mmol/L, standard deviation 2.5 mmol/L, total number 1000). On the right, an S-shaped distribution function Φ(x), indicating which theoretical proportion comprises all the values to the left of the indicated value x (up to mean 140, for example, 50%).
Relative frequency is the number in each column divided by the total number.

evaluations. The basis of the QQ-plot method is the NORM.INV function in Excel [18]. The name (norm=normal, inv=inverse) designates the inverse function of the s-shaped distribution function[1] of a normal distribution (Figure 1), and it has the following syntax:

$$x=\text{NORM.INV}(p, \mu, \sigma)$$

p=probability, μ=mean, σ=standard deviation.

In order to simulate, for example, a random sodium value from a normal distribution with the parameters μ=140 and σ=2.5, one enters the following formula in cell A1 of an Excel sheet:

$$=\text{NORM.INV}(\text{rand}();140;\ 2.5)$$

The "rand()" function in this formula yields equally distributed decimal numbers between 0 and 1 and thus covers the entire range of possible probabilities p.

In this present study, the simulation of such random numbers of known distribution plays a major role, because it yields predictable results on the basis of which the quality of a model-based statistical method can be assessed (for example, see Figure 3, Tables 3 and 6). If one takes the above formula in cell A1 and copies it to the 999 cells from A2 to A1000 below in the Excel sheet, one obtains a total of 1000 normally distributed random numbers that represent realistic sodium values with a typical frequency around the mean of 140: In the range between μ–1σ and μ+1σ, one expects 68% and between μ–2σ and μ+2σ 95%. This 2σ- or 95% range corresponds to the required reference interval.

If the Analysis ToolPak has been installed, one may simply select the "Data Analysis" submenu item in the "Data" menu item. Calling up the "Random Number Generation" function, a screen opens for the input of parameters (for example, for a variable with 1000 random numbers, "Standard" distribution, mean 140, standard deviation 2.5).

Even skewed distributions can be simulated using the NORM.INV function. For example, to obtain lognormally distributed random values for alanine aminotransferase (ALT) in men with a mode near 20 U/L ($e^{3.0}$) and a 97.5 percentile near 45 U/L ($e^{3.8}$), one generates in column A a normal distribution 3.0±0.4 and enters in cell B1 the formula

$$=\text{EXP}(A1).$$

This is, then, copied to the 999 cells A2 to A1000 below in the Excel sheet to obtain lognormally distributed values in column B (Figure 2).

Finally, using this method, it is possible to generate also mixed populations that approximate the real, clinical situation, such as 800 inconspicuous sodium levels (140±2.5) and 200 reduced levels (130±4.5). The result is shown in Table 2: The stated relative frequencies are obtained via the "Frequency" Excel function and/or the "Histogram" menu item in Analysis Tools.

1 An inverse function does not calculate y from x, but x from y – in the present case, the appropriate quantile and/or test result (x) from a given cumulative frequency distribution (y).

B1 =EXP(A1)

	A	B	C	D	E	F	G	H
1	2,88	18		Class	Frequency	Relative frequency	Cumulative	
2	2,49	12		5	0	0,000	0,000	
3	3,10	22		10	38	0,038	0,038	
4	3,51	33		15	210	0,210	0,248	
5	3,48	32		20	242	0,242	0,490	
6	3,69	40		25	197	0,197	0,687	
7	2,13	8		30	135	0,135	0,822	
8	2,91	18		35	83	0,083	0,905	
9	3,44	31		40	46	0,046	0,951	
10	2,57	13		45	28	0,028	0,979	
11	2,72	15		50	8	0,008	0,987	
12	2,32	10		55	8	0,008	0,995	
13	2,26	10		60	2	0,002	0,997	
14	2,61	14		65	2	0,002	0,999	
15	2,69	15		70	1	0,001	1,000	
16	2,15	9						
17	2,77	16						
18	2,84	17						
19	3,05	21						
20	2,85	17						
21	2,87	18						
22	2,85	17						
23	3,54	34						
24	2,97	19						
25	2,93	19						
26	2,79	16						
27	3,79	44						
28	3,35	28						
29	3,95	52						
30	2,74	15						
31	3,66	39						
32	2,36	11						
33	3,22	25						
34	3,36	29						

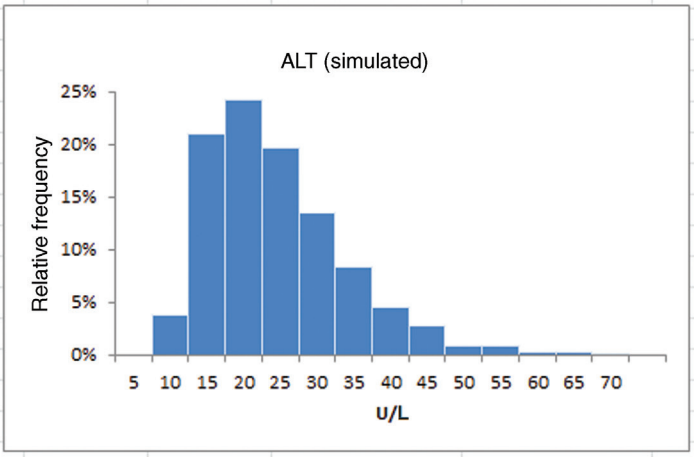

Figure 2: Simulation of lognormally distributed values using the example of ALT. (original representation in the German Excel version with deimal commas; the same applies to figures 4, 5, and 8).
In column A, normally distributed values (3.0±0.4) were generated, representing exponents to base e; in column B, these were delogarithmized with the exponential function EXP. The theoretical maximum of the histogram is therefore $e^{3.0}$=20, and of the 97.5 percentile $e^{3.8}$=45.

Visual check for normal distribution

The visual methods presented here are based on two-dimensional diagrams in which observed and theoretical probabilities are plotted against each other. At its simplest, this is done without a computer using a probability probability plot (PP-plot). Figure 3 demonstrates the approach using the simulated sodium levels from Table 2 as an example: If the points form a straight line, a normal distribution can be assumed (black symbols in Figure 3). The reference interval (μ±2σ) then lies between the intersections of the straight line with the 2.5 and/or 97.5 percentiles. Values deviating from this straight line (red symbols in Figure 3) represent "outliers" (in this case, hyponatremia).

QQ-plot method

We chose the quantile quantile plot (QQ-plot) to recreate the historical method in Excel. Unlike the PP-plot (Figure 3) described above, it is the inverse function (NORM.INV) that is used here: the x- and y-axes are swapped, so that the diagram appears rotated by 90 degrees, and the (new) x-axis can be scaled linearly by converting the probabilities to quantiles (Figure 4).

Table 2: Simulation of 1000 sodium levels by means of the NORM.INV function.

Sodium, mmol/L	Non-diseased subjects only		20% diseased subjects	
Class	Number	Cumulative	Number	Cumulative
117.5	0	0.0%	0	0.0%
120.0	0	0.0%	1	0.1%
122.5	0	0.0%	7	0.8%
125.0	0	0.0%	22	3.0%
127.5	0	0.0%	27	5.7%
130.0	0	0.0%	54	11.1%
132.5	0	0.0%	30	14.1%
135.0	19	1.9%	48	18.9%
137.5	142	16.1%	137	32.6%
140.0	335	49.6%	277	60.3%
142.5	325	82.1%	255	85.8%
145.0	158	97.9%	125	98.3%
147.5	20	99.9%	16	99.9%
150.0	1	100.0%	1	100.0%

The left side shows random values of a normal distribution (μ=140, σ=2.5), which represent the results of non-diseased subjects. On the right, the parameters for 20% of values were modified to simulate hyponatremia (μ=130, σ=4.5).

The practical approach in Excel is similar to that employed in the above simulations of sodium and ALT levels. The NORM.INV function is used to calculate theoretical quantile of a standard normal distribution (μ=0, σ=1), by entering in the formula a systematically ascending series of probabilities between 0 and 1 instead of random numbers (Figure 4). Thanks to computers, it is not necessary to limit oneself to a few points, as depicted in Figure 3. Instead, one can create a separate class for each measured value, and then compare its actual quantile with the theoretical one.

As for the above example of sodium, one thus obtains 1000 classes to which a 1000 theoretical quantiles are assigned in the diagram. The extreme probabilities 0 and 1 must be left aside, because they would correspond to minus and/or plus infinity quantiles that could not be represented (cf. y-axis in Figure 3)[2].

Results

Figure 5 illustrates the two examples described in the previous section for sodium (normally distributed) and ALT (lognormally distributed). The x-axis represents the theoretical quantile of the standard normal distribution (μ=0, σ=1); so, similar to the y-axis in Figure 3, it shows the deviations from the mean in multiples of the standard deviation, only this time in linear scaling. In the case of sodium, the QQ-plot produces the straight line, as expected. As for ALT, the result is a curved plot without a linear portion. But it can be transformed into a straight line by scaling the y-axis logarithmically (Figure 5, bottom). Alternatively, the ALT levels can be logarithmized prior to the analysis, and a linear y-axis can be used.

Figure 6 shows typical QQ-plots for mixed populations. Among the 1000 normal sodium levels (140±2.5 mmol/L), 100, 200 and/or 300 were replaced by reduced levels (130±4.5) to simulate 10%, 20% and 30% outliers. As one can see, the points are set off by a kink in the line. The mean and standard deviation of a (sub-)population were then calculated in the usual manner from the values that were on a straight line in the QQ-plot. In the present case, the kink was localized in the range between 136 and 137 mmol/L.

Table 3 shows that with the addition of reduced values, the standard deviations in the uncorrected datasets increase (row 1), while the mean values keep decreasing (row 3), thus producing a reference interval that is too wide and the mean value is too small (for example with a 30% addition of reduced values, 126–148 mmol/L). However, the elimination of the outliers on the basis of the QQ-plot yields a remarkably stable approximation: The mean value now remains constant (row 5), and the standard deviation decreases only slightly (row 6). Due to the somewhat smaller standard deviation, the lower reference limit changes by no more than 1 mmol/L, while the upper limit is not affected at all.

In a further series of simulation experiments, we wanted to find out about the effect of the subjective reading of the "kink" on the final result. We have been able to show that the mean values are virtually not affected by inexact readings of up to 5%, while standard deviations tend to be under- rather than overestimated. Figure 7 shows, by way of an example, that the standard deviation – as expected – keeps decreasing, the more measured values are cut off as outliers (in other words, the further the point of measurement is moved to the linear part). Meanwhile, the reference limits in this example, too, changed only slightly by a maximum of 1 mmol/L.

Real data from a study involving 246 clinically inconspicuous students (male and female) were available for the practical verification of the method [17]. As the test subjects were not standardized with respect to their fasting status, one had to expect mostly inconspicuous

2 This is why the formula in Figure 4 is divided by n+1 rather than n. This yields for the highest probability 1000/1001=0.9990, and not 1000/1000=1.

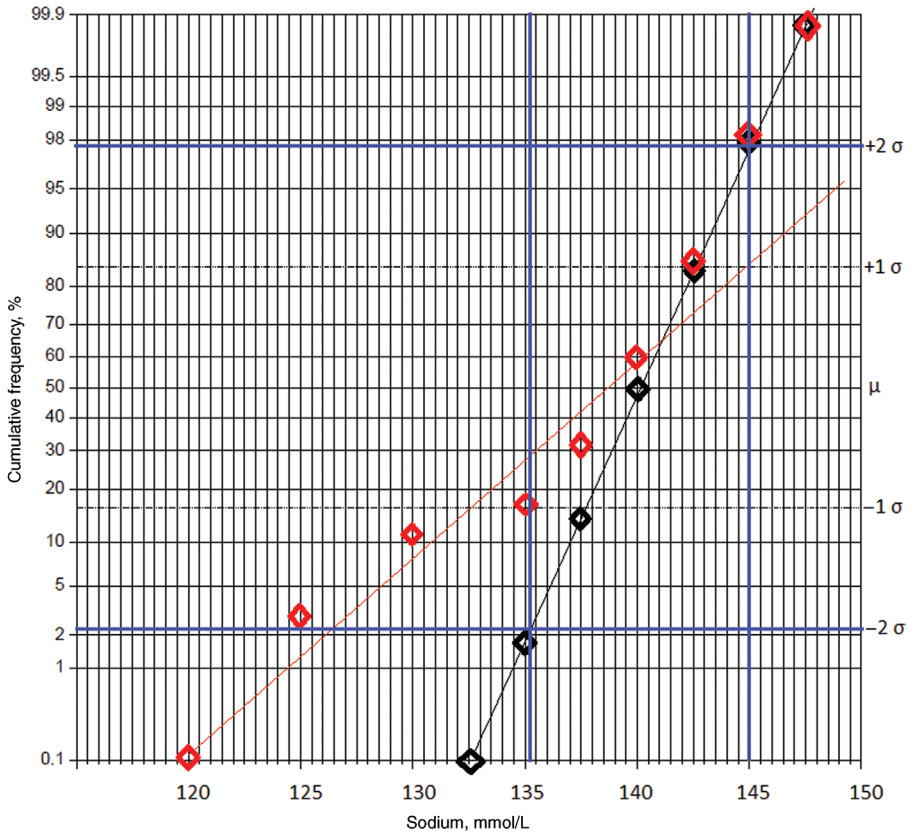

Figure 3: Visual check of the sodium levels in Table 2 for normal distribution using a probability probability plot.
Black: non-diseased, Red: mixed population with 20% hyponatremia. The blue lines represent the 2σ or 95% range that corresponds to the reference interval (here, about 135–145 mmol/L).

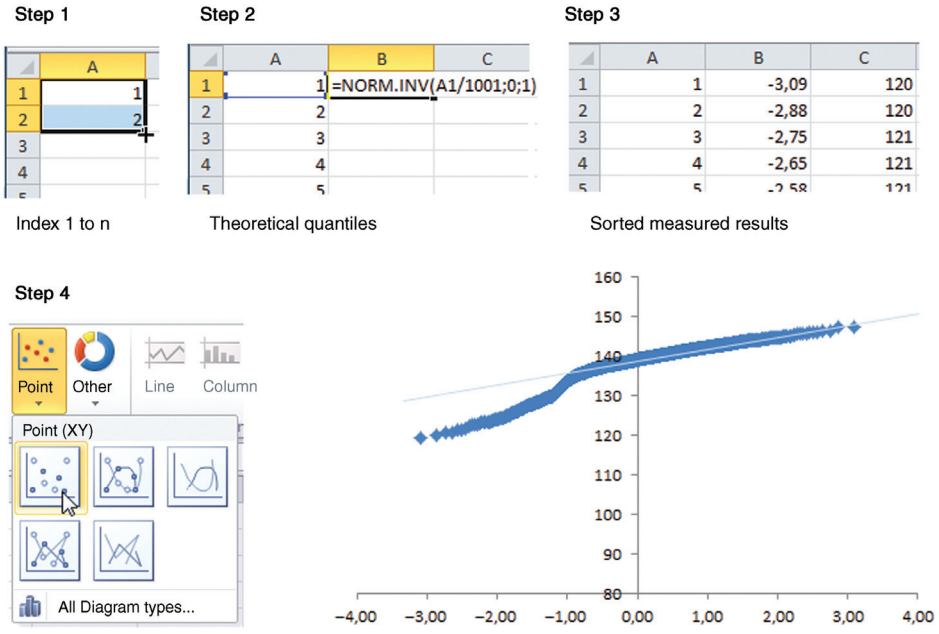

Figure 4: Generating a QQ-plot with Excel in four steps.
Step 1: In column A you create a series of numbers i from 1 to n. **Step. 2:** In column B, using the NORM.INV function (p, μ, σ), calculate the quantiles of a standard normal distribution. The probability p is calculated from the index i divided by n+i. For example, for n=1000, enter in cell B1: =NORM.INV(A1/1001); 0; 1). This formula is copied to all n cells of column B. **Step 3:** Copy the measurement results to column C and sort them in an ascending order. Step 4: In order to obtain the desired QQ-plot, highlight columns B and C, and insert a scatter chart.

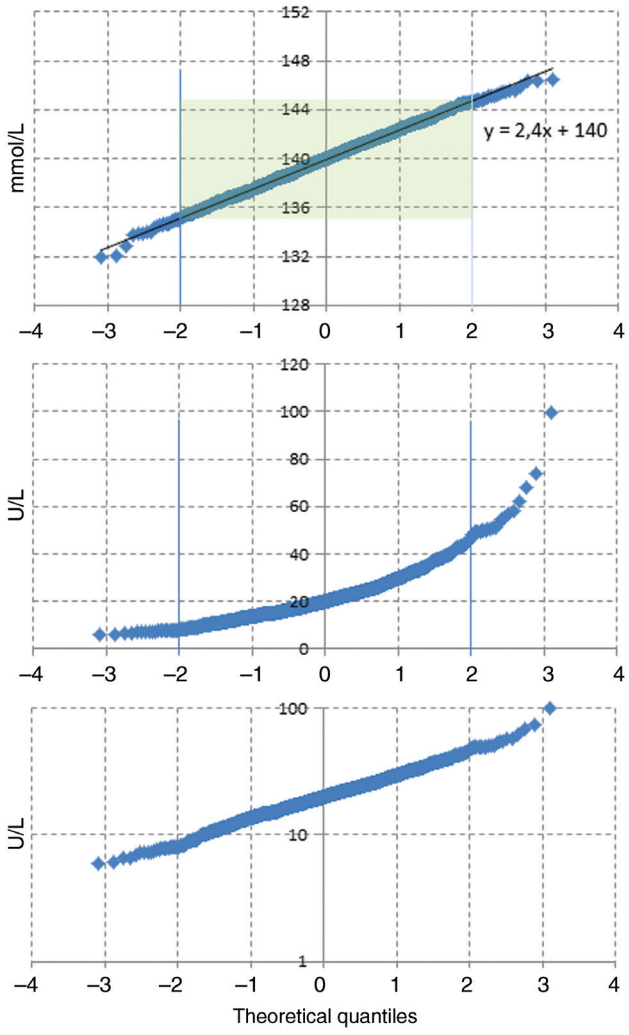

Figure 5: QQ-plots for one thousand simulated readings each. Normally distributed sodium values form a straight line (top); from their regression equation, one can estimate the mean μ and the standard deviation σ. The reference interval (135 mmol/L–145 mmol/L, μ±2σ) can be obtained between the –2 and +2 quantiles of the standard normal distribution (green area). Lognormally distributed ALT levels produce a curved plot (center) that becomes linear after logarithmic scaling of the y-axis (bottom).

results, as well as a limited number of pathological findings, in connection with nutrition-related factors.

Figure 8 illustrates the frequency distributions and QQ-plots of total protein, glucose and triglycerides. Total protein (top) represents non-disturbed measured values that appeared in the QQ-plot as almost normally distributed without any significant deviations. Albumin, sodium, potassium, calcium and creatinine exhibited patterns similar to total protein. Glucose readings (diagram in the middle) yielded in the central area a straight line as well, but some hypoglycemias (lowest value 2.2 mmol/L or 40 mg/dL) and hyperglycemias (highest value 17.5 mmol/L or 315 mg/dL) were found. Conspicuously elevated levels were also found for cholesterol, GGT, ALT and AST. Prior to the QQ-plot analysis, the original values for the three enzymes were logarithmized in order to transform the curved plot into a straight line (cf. Figure 5, bottom).

The triglyceride levels (lower diagram) exhibited a complex distribution consisting of subpopulations, which demonstrated the limits of the method presented herein. Two, about equally long, straight portions emerged in the QQ-plot: One marked a hypoglycemic range between 0.34 and 0.80 mmol/L (30–70 mg/dL), while the other one covered the range between 0.80 and 2.0 mmol/L (70–175 mg/dL). Hypertriglyceridemia was also observed up to 3.74 mmol/L (327 mg/dL). This is not surprising, however, because the test subjects were not examined under standardized conditions. The division into two linear portions with different slopes remained intact even after the values had been logarithmized (not shown). As explained in the more detailed discussion (Figure 10), such data cannot be evaluated with this method.

But apart from this example, all QQ-plots were linear over a large section (sometimes after the values had been logarithmized). Any outliers were set off clearly by way of a kink and were excluded from further calculations. The reference intervals obtained in this way corresponded well to the information established at the MHH despite the relative small number of cases (fewer than 250) (examples in Table 4).

Two extensive datasets from clinical practice were made available to us for a comparison of methods: 17,506 creatinine readings between 0.26 and 22.61 mg/dL, and 56,137 AST readings between 1 and 200 U/L. The analysis was done separately for men and women. Figure 9 shows the two graphics for creatinine levels in women (n=9.393), which were obtained using the two methods. In the histogram of the comparison method (left), one can see the symmetrically distributed population of healthy subjects, who are represented in the QQ-plot (right) by a straight line. When analyzing the AST data, we also tested the hypothesis that the values were lognormally distributed [19]. To do so, the original data were logarithmized prior to the analysis, and the reference limits derived from this were then delogarithmized.

Table 5 shows the agreement between the two methods. The difference for creatinine was a maximum of 0.1 mg/dL, and 3 U/L for AST. In the case of AST, the logarithmic transformation produced a slight widening and rightward shift of the reference interval.

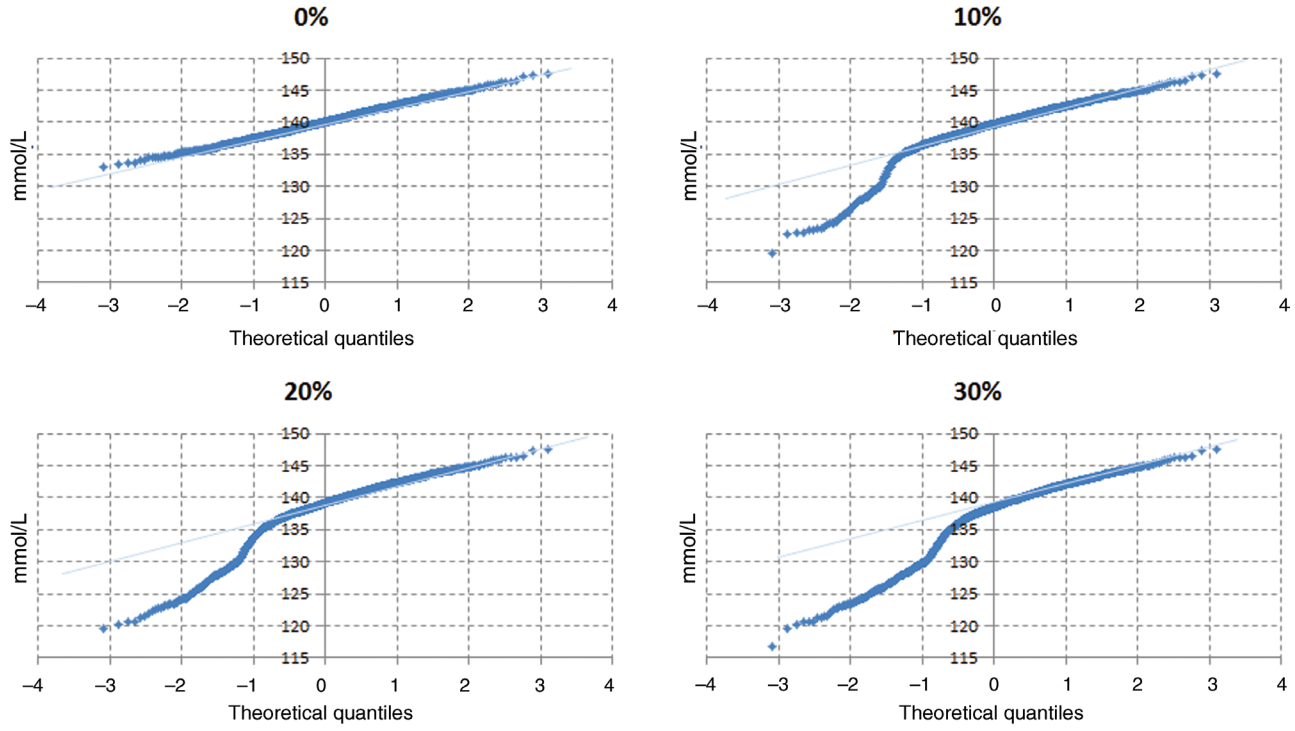

Figure 6: Simulation of one thousand sodium levels each with a variously-sized percentage of hyponatremia (color highlighted area). One can see a line segment against which the reduced values are offset by means of a kink.

Table 3: Estimation of reference intervals for sodium (target 135–145 mmol/L) from mixed populations with 0%–30% reduced levels.

	0%	10%	20%	30%
(A)				
Mean	140	139	138	137
Standard deviation	2.6	4.1	5.1	5.7
Lower limit	135	131	128	126
Upper limit	145	147	148	148
(B)				
Mean	140	140	140	140
Standard deviation	2.6	2.4	2.3	2.3
Lower limit	135	135	136	136
Upper limit	145	145	145	145

(A) Calculation from all one thousand values. (B) After visual elimination of values that are not on the straight line in the QQ-plot.

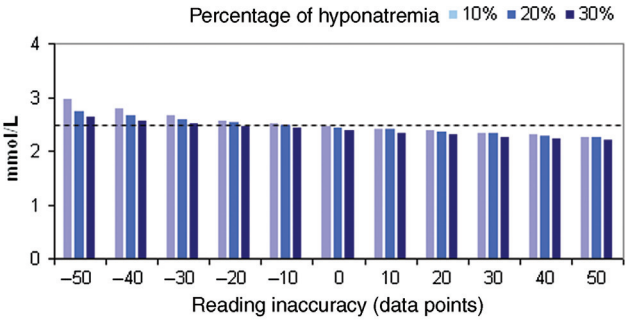

Figure 7: Identical simulation experiment as in Figure 6. As part of an error analysis, the localization of the kink, however, was shifted left by 10–50 measurement points (negative values) and/or right (positive values). The resulting flawed estimation of the standard deviations from the target (dashed line) was a maximum of 15%.

Discussion

The method presented herein is not fundamentally new, but only represents the last link (for the time being) in a chain of precursor versions, which can all be traced back to the American biostatistician Robert G. Hoffmann [14]. He was the first to propagate in 1963 the concept of "indirect estimation" of reference intervals from measurements taken from outpatients and inpatients, rather than reference subjects. Two conditions needed to be met for this: The measured values were to follow, more or less, a Gaussian normal distribution, and the majority of patients should be clinical inconspicuous with respect to the analyte examined. As proof of concept, he demonstrated his approach on the basis of glucose measurements in non-diabetics by sorting the measured values in an ascending order and marking the cumulative frequency of each value on probability paper (cf. Figure 3).

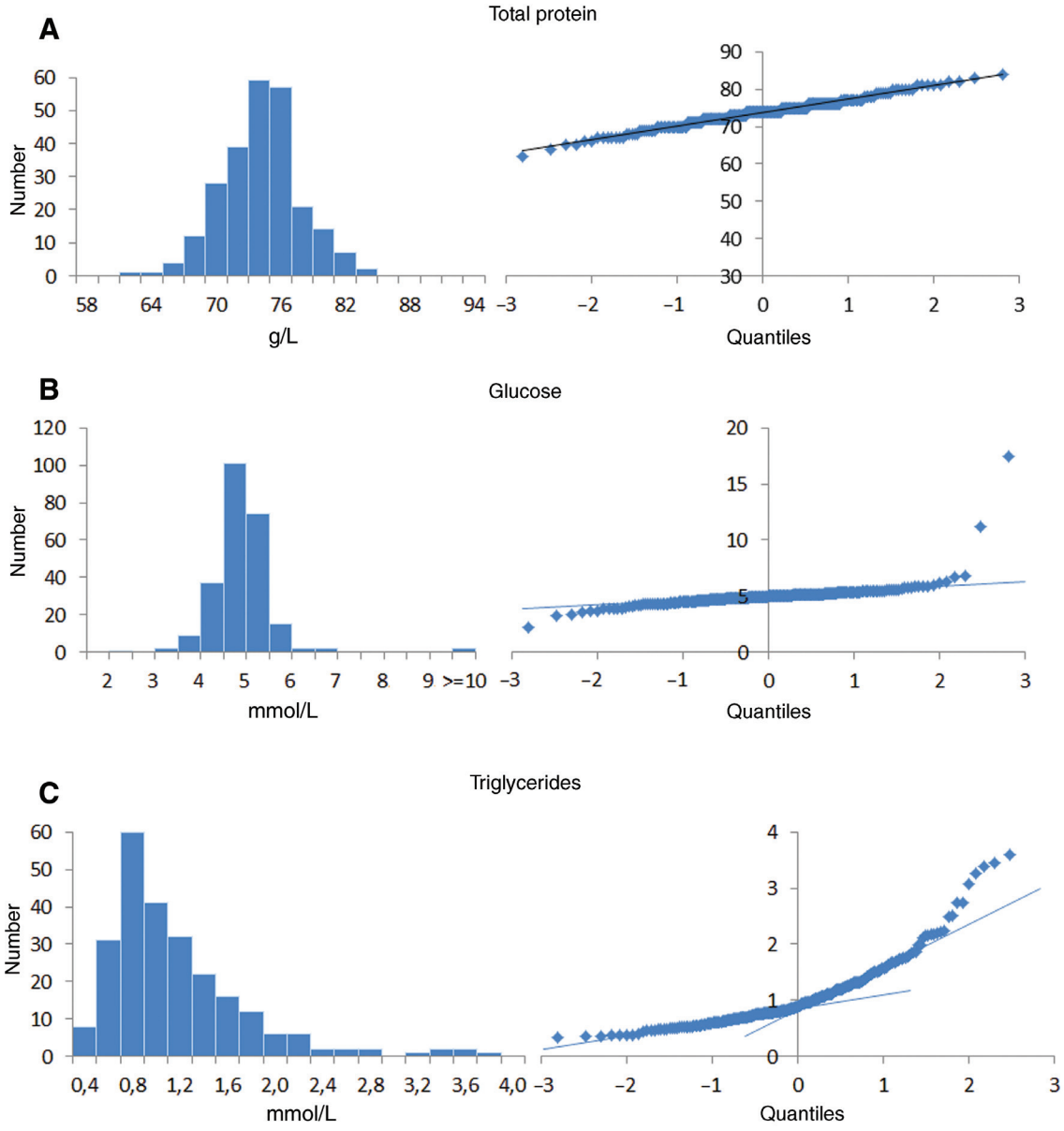

Figure 8: Selected application examples of the QQ-plot method.
(A) Linear QQ-plot without conspicuous values. (B) Linear QQ-plot with few pathological values on both sides. (C) Complex QQ-plot with two approximately equal straight segments and pathological values at the top.

Table 4: Estimation of reference intervals with the QQ-plot method for a population of clinically inconspicuous students whose blood was collected under non-standard conditions [17].

	MV	SD	Estimated reference interval	MHH specification
Total protein, g/L	73.8	3.7	66–81	65–80
Albumin, g/L	43.8	4.0	36–52	35–52
Sodium, mmol/L	140	1.8	136–143	135–145
Potassium, mmol/L	4.29	0.29	3.7–4.9	3.6–5.4
Calcium, mmol/L	2.38	0.09	2.2–2.6	2.15–2.6
Glucose, mmol/L	4.95	0.44	4.1–5.8	3.9–5.5

He determined the reference interval visually by placing a straight line through the central part of the points and identifying their intersections using the 2.5 and 97.5 percentiles. His approach has been adopted, sometimes in its original form, sometimes in a modified version, in several medical papers [20–22].

The method described herein combines various modifications of the original method so as to achieve estimations as solid as possible with the least possible effort. The basis for this is a “modified Hoffmann approach”, published in 2014 [4], in which the quantiles of a QQ-plot

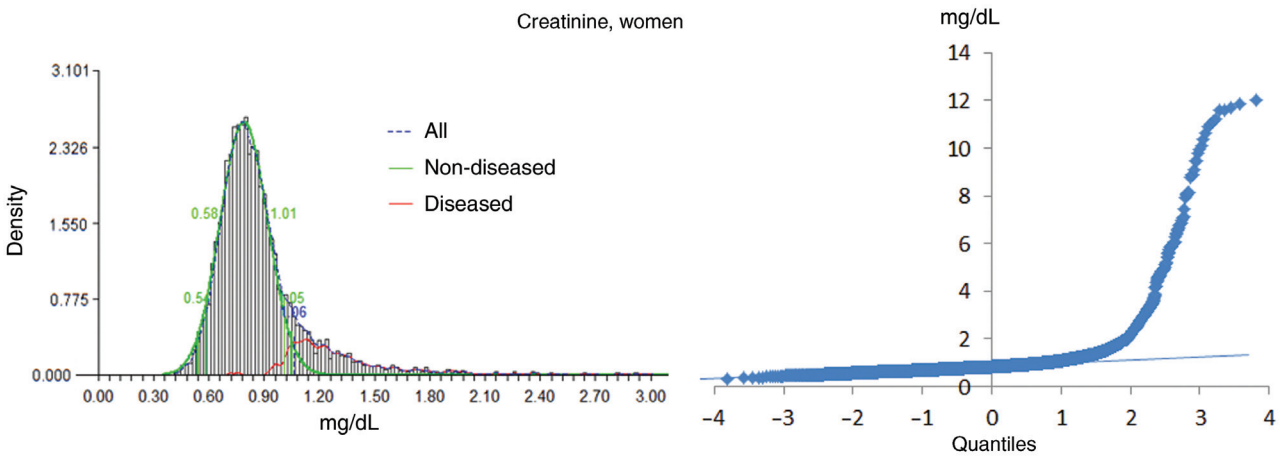

Figure 9: Method comparison of the reference interval estimation based on routine clinical data, with the example of creatinine in 9393 women.

On the left, the RLE method of the Reference Values Working Group (calculation via R); the green line indicates the density of the assumed normal distribution, and the red line shows the density of the elevated levels that do not fit the model (x-axis is cut off on the right). On the right, the corresponding QQ-plot (calculation in Excel); the data of the non-pathological primary population are represented as a straight line, and the outliers are clearly distinguishable due to the kink.

Table 5: Method comparison of the reference interval estimation based on routine clinical data.

Parameter and distribution model	Men		Women	
	Method 1	Method 2	Method 1	Method 2
Creatinine, mg/dL (N)	0.6–1.4	0.7–1.3	0.5–1.2	0.5–1.1
AST, U/L (N)	12–33	9–33	12–29	10–29
AST, U/L (LN)	13–36	13–38	13–32	13–33

Despite a high proportion of outliers, both methods provide quite similar estimations. Explanations on the model hypotheses can be found in the text. Method 1=QQ-plot method, Method 2=comparison method of Reference Values Working Group. N, Normal distribution model; LN, lognormal distribution model.

are calculated using Excel, as described in Figure 4. But the authors follow this up with a segmented regression analysis using the statistical software package R, which was not adopted in this context due to its complexity and susceptibility to errors. Instead, we determine the linear part of the QQ-plot visually and estimate the reference interval $\mu \pm 2\sigma$ [23] by eliminating potential outliers directly by way of calculating the mean and standard deviation.

This approach is not only much easier, but it also appears to produce more solid results. Our experiments with simulated data show that the readings from diseased subjects and subjective reading inaccuracies have only little effect on the estimation of reference intervals. By contrast, Shaw et al. [4], using the more complex regression-based method, obtained results some of which were extremely flawed – the authors attributed this to a high percentage of diseased test subjects.

Figure 10 provides an explanation for the error-proneness of the regression-based approach in connection with mixed populations: Three populations of different sizes (300 healthy, 200 moderately ill and 100 severely ill subjects) were simulated with normally distributed, arbitrary measured values (100±20, 150±40, and 250±50). In total, this produces a typical left-skewed distribution reminiscent of a lognormal distribution. In the QQ-plot, one can see three sections that are more or less linear, but their intersections with the y-axis do not match the expected mean values: Instead of 100, 150 and 250, one reads 120, 100 and 180. The slopes of the pitch lines in Figure 10 do not correspond to the expected standard deviations either.

To arrive at a correct calculation, one would have to isolate the data of the three line segments and plot them against the quantiles of matching standard normal distributions with n=300, 200 and 100, respectively. In a

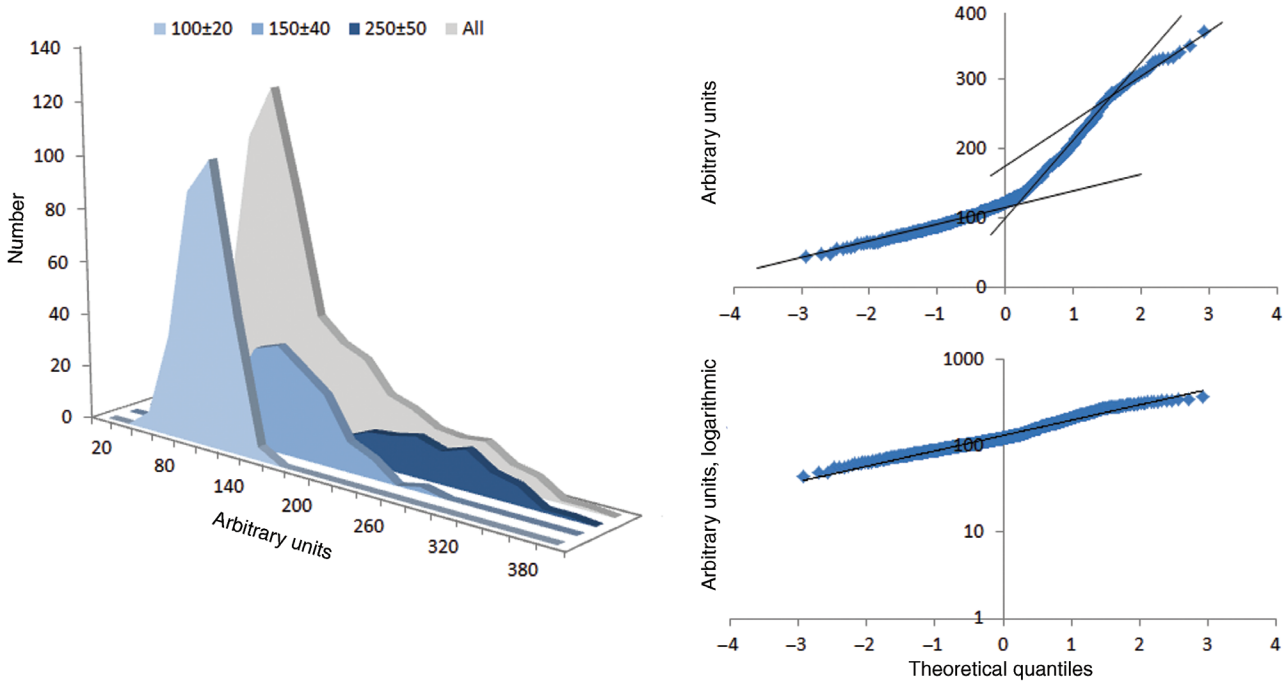

Figure 10: Histograms (left) and the QQ-plots for a mixed patient cohort.
Three normally distributed sets of values were generated using a simulation (blue areas), which produce in total a left-skewed distribution (gray area). The three subpopulations can be distinguished visually well in the QQ-plot (top, right), but the position of the straight sections (slope and intercept) does not allow for conclusions to be drawn about the underlying mean values and standard deviations. Even after a logarithmic transformation of the y-axis (bottom, right), the three linear segments are visible, but are substantially flattened by the automatic scaling in Excel.

Table 6: Estimation of reference intervals of a mixed population (Figure 10) with the method presented herein, as well as by way of the segmented regression calculation according to Shaw et al. [4].

	Target (simulation)	Current method	Shaw et al.
1 (n=300)	60–140	62–125	66–177
2 (n=200)	70–230	83–257	–124 to 324
3 (n=100)	150–350	262–350	45–305

The degree of red color symbolizes the extent of each flawed estimation; in extreme cases, the regression method yields a negative lower limit.

summary analysis of all 600 data points, the respective quantiles are shifted by a distance equivalent to the data from the other lines. As a result, the mean values tend to be over- and sometimes underestimated, while the standard deviations are overestimated consistently. In extreme cases, these errors actually produce negative lower limits (Table 6, right-hand column).

Five recommendations can be derived from what has been said up to this point:

– The QQ-plot method described herein is especially suited for a quick visual assessment as to whether patient readings are generally appropriate for calculating reference intervals.
– If the graph produces a continuous line (Figure 5, top), one can assume a normal distribution and estimate the reference interval from $\mu \pm 2\sigma$.
– Any outliers are usually easy to identify by deviations from the line shape (Figure 8, center, Figure 9, right) and can be eliminated from the calculation of μ and σ up to a proportion of 20%–30% (Table 3).
– Where the entire curve deviates from the line shape (Figure 5, center), the measured values can be logarithmized and the QQ-plot recreated. If this produces a straight line, points 2 and 3 are repeated with logarithmized values, and the results are then delogarithmized.
– Where even this fails to generate a straight line (Figure 10, bottom right), one must expect considerable mistakes in the estimation (Table 6). In this case, a more suitable patient cohort should be assembled – or better yet, a cohort of non-diseased reference subjects.

Aside from the simple logarithmic transformation, a number of other transformation procedures have been

recommended to approximate skewed distributions to a normal distribution [23]. For example, in the RLE method of the Reference Values Working Group, the Box-Cox transformation, part of the group of power transformations, was used by default [7]. Some authors propose that the original data should generally be transformed prior to the application of statistical models where the distribution function is unknown [19, 24]. But one must keep in mind that conspicuous data structures may be hidden by transformations (cf. Figure 10) and that one should always look for physiological or clinical causes in the event of skewed distributions before they are "normalized" on a purely statistical basis.

In a study published in 2015, Tate et al. [25] recommend the use of indirect methods particularly to adapt published reference intervals to local circumstances (analysis, pre-analysis, special patient populations, etc.). The authors demand that the majority of readings must be from clinically inconspicuous subjects and that the conspicuous readings be clearly distinguishable statistically. The present paper takes this demand into account in a straight-forward manner: As Figure 9 shows, the outliers are clearly visible to the naked eye because of the "kink" in the QQ-plot, while deviations from the bell-curve shape of the normal distributions are less clear, for example.

The method described herein does not compete with other methods, but represents a useful addition. Compared to the IFCC reference method [1], this method works without recruiting healthy test subjects, and compared to the RLE method of the Reference Values Working Group [7], it does not require installation of the statistics software R, while still delivering plausible results even with only around 200 measured values. The visual assessment of the QQ-plot may be seen as problematic, because it is subjective and cannot be automated. A way to automate it, however, is being worked on.

Still, visualization also provides some advantages over purely computational methods, because it offers at a glance a lot of information on data quality and the form of distribution, on any outliers and subpopulations, etc. Thus, measured values that could affect the estimation of reference intervals are identified with relative certainty. After all, even for hand-picked reference subjects, it is impossible to rule out that they suffer from clinically "silent" diseases (fatty liver, atherosclerosis) or take substances (alcohol, drugs) that could affect the outcome. This is particularly true of laboratory data routinely obtained from clinical populations. Therefore, the QQ-plot method should generally be employed as a type of pre-filter regardless of the direct or indirect method that is ultimately used.

Acknowledgments: The authors wish to thank the German Society for Clinical Chemistry and Laboratory Medicine (DGKL) for supporting the study as part of the Bioinformatics and Reference Values Working Groups. Special thanks go to Dr. Farhad Arzideh, University of Bremen, Dr. Norman Bitterlich, Medizin & Service in Chemnitz, and Prof. Dr. Frank Klawonn, Helmholtz-Zentrum in Braunschweig, for their valuable discussion contributions.

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

Research funding: 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

1. Clinical and Laboratory Institute Standards document C28-A3: Defining, establishing and verifying reference intervals in the clinical laboratory: approved guideline, 3rd ed., 2008.
2. DAkkS Checkliste zur DIN EN ISO 15189:2014 für medizinische Laboratorien, Abschnitt 5.5.2., available at www.dakks.de, retrieved on 05.10.2015.
3. Haeckel R, Wosniok W, Arzideh F. A plea for intra-laboratory reference limits. Part 1. General considerations and concepts for determination. Clin Chem Lab Med 2007;45:1033–42.
4. Shaw J, Cohen A, Konforte D, Bineh-Marvasti T, Colantonio D, Adeli K. Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clin Biochem 2014;47: 166–72.
5. Sonntag O. Ist das normal? – Das ist normal! Über die Bedeutung und Interpretation des sogenannten Normalwertes. J Lab Med 2003;27:302–10.
6. Glick J. Statistics of patient test values: application of indirect normal range and quality control. Clin Chem 1972;18:1504–13.
7. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, et al. A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-laboratory databases demonstrated by catalytic activity concentrations of enzymes. Clin Chem Lab Med 2007;45:1043–57.
8. Arzideh F, Brandhorst G, Gurr E, Hinsch W, Hoff T, Roggenbuck L, et al. An improved indirect approach for determining reference limits from intra-laboratory databases exemplified by concentrations of electrolytes. J Lab Med 2009;33:52–66.
9. Zierk J, Arzideh F, Haeckel R, Rascher W, Rauh M, Metzler M. Indirect determination of pediatric blood count reference intervals. Clin Chem Lab Med 2013;51:863–72.

10. Arzideh F, Wosniok W, Haeckel R. Reference limits of plasma and serum creatinine concentrations from intra-laboratory data bases of serveral German and Italian medical centres. Clin Chim Acta 2010;411:215–21.
11. Arzideh F, Wosniok W, Haeckel R. Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory data bases from several German and Italian medical centres. Clin Chem Lab Med 2011;49:659–64.
12. Hoffmann G. IT-Werkzeuge zur Auswertung großer labordiagnostischer Datensätze. Klin Chem Mitteilungen 2011;42:124–30.
13. Pryce J. Level of haemoglobin in whole blood and red blood cells, and proposed convention for defining normality. Lancet 1960;2:333–6.
14. Hoffmann R. Statistics in the practice of medicine. J Am Med Assoc 1963;185:864–73.
15. Cook M, Levell M, Payne R. A method for deriving normal ranges from laboratory specimens applied to uric acid in males. J Clin Path 1970;23:778–80.
16. Hoffmann G. Auflösung eines Dilemmas: Referenzintervalle zum Selbermachen. Trillium Diagnostik 2014;12:159–61.
17. Lichtinghagen R, Senkpiel-Jörns D, Brand K, Janzen N. Beurteilung des Einflusses verlängerter Stauzeiten auf nicht-normalisierte versus normalisierte klinisch-chemische Messgrößen. J Lab Med 2013;37:131–7.
18. http://office.microsoft.com/de-ch/excel-help/norm-inv-funktion-HP010335690.aspx, retrieved on 26.5.2015.
19. Haeckel R, Wosniok W. Observed, unknown distributions of clinical chemical quantities should be considered to be log-normal: a proposal. Clin Chem Lab Med 2010;48:1393–6.
20. Neumann G. Determination of normal ranges from routine laboratory data. Clin Chem 1968;14:979–88.
21. Reed A, Cannon D, Winkelman J, Bhasin Y, Henry R, Pileggi V. Estimation of normal ranges from a controlled sample survey. Clin Chem 1972;18:57–66.
22. Soldin O, Hoffmann E, Waring M, Soldin S. Serum iron, ferritin, transferrin, total iron binding capacity, hs-CRP, LDL, cholesterol and magnesium in children; new reference intervals using the Dade Dimension Clinical Chemistry System. Clin Chim Acta 2004;342:211–7.
23. Reed A, Wu G. Evaluation of a transformation method for estimation of normal range. Clin Chem 1974;20:576–81.
24. Harris E, DeMets D. Estimation of normal ranges and cumulative proportions by transforming observed distributions to Gaussian form. Clin Chem 1972;18:605–12.
25. Tate J, Yen T, Jones G. Transference and validation of reference intervals. Clin Chem 2015;61:1012–5.

Article note: Original German online version at: http://www.degruyter.com/view/j/labm.2015.39.issue-6/labmed-2015-0082/labmed-2015-0082.xml?format=INT. The German article was translated by Compuscript Ltd. and authorized by the authors.