Frank Klawonn, Georg Hoffmann and Matthias Orth*

Quantitative laboratory results: normal or lognormal distribution?

<https://doi.org/10.1515/labmed-2020-0005>

Received January 17, 2020; accepted April 27, 2020; previously published online June 4, 2020

Abstract: The identification of a suitable distribution model is a prerequisite for the parametric estimation of reference intervals and other statistical laboratory tasks. Classification of normal vs. lognormal distributions from healthy populations is easy, but from mixed populations, containing unknown proportions of abnormal results, it is challenging. We demonstrate that Bowley's skewness coefficient differentiates between normal and lognormal distributions. This classifier is robust and easy to calculate from the quartiles Q_1 – Q_3 according to the formula $(Q_1 - 2 \cdot Q_2 + Q_3)/(Q_3 - Q_1)$. We validate our algorithm with a more complex procedure, which optimizes the exponent λ of a power transformation. As a practical application, we show that Bowley's skewness coefficient is suited selecting the adequate distribution model for the estimation of reference limits according to a recent International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) recommendation, especially if the data is right-skewed.

Keywords: ALAT; Bowley's skewness; creatinine; hemoglobin; leukocytes; lognormal distribution; normal distribution; potassium; power transformation; quartile skewness; sodium.

Background

Quantitative laboratory results of healthy individuals usually show either a symmetric or a right-skewed

histogram [1, 2]. The former may be described by a normal and the latter by a lognormal distribution, although real data probably never follow exactly ideal distributions in the form of normal, lognormal or other simple types of distributions. Nevertheless, in the spirit of the saying that “all models are wrong, but some are useful” attributed to the famous statistician George Box, it is reasonable to work with such idealized model assumptions. In the context of quantitative laboratory results, one might argue that it is worthwhile to consider more than just two options – normal vs. lognormal – for instance in the form of Box-Cox transformations. However, we demonstrate here that it is extremely difficult – if not impossible – to estimate the correct parameters for a Box-Cox transformation from mixed populations. This is the reason why we focus here on the decision between a normal and lognormal distribution. Evaluating the skewness – and consequently the distribution type – is easy, if the data set represents results of a homogenous cohort of healthy individuals [1]. It is, however, challenging under the conditions of a recent IFCC recommendation [3], where reference intervals are estimated from mixed populations.

The correct selection of the distribution type has an impact on the parametric estimation of reference intervals [1, 3, 4], as well as on other statistical laboratory tasks such as the calculation of permissible uncertainties [5] or the standardization of laboratory results [6].

The decision whether the data of a mixed population should be transformed or not, will always end in a vicious circle: Pathological values can be misleading for the choice of the transformation, whereas removing outliers before transformation can be too strict because in a skewed, e.g. lognormal, healthy population, non-pathological values could be classified as outliers. We believe that deciding on the transformation based on a method which is robust against outliers is a way to break the vicious circle. The focus of this paper is on the decision for the transformation, not the steps after the transformation, i.e. removing outliers and estimation reference limits. Nevertheless, we apply simple methods for reference interval estimation to generally demonstrate the usefulness of our approach but without the intention to discuss and evaluate different

*Correspondence: PD Dr. med. Matthias Orth, Vinzenz von Paul Kliniken gGmbH, Stuttgart, Germany; Medizinische Fakultät Mannheim, Ruprecht Karls Universität, Mannheim, Germany; and Institut für Laboratoriumsmedizin, Adlerstr. 7, 70199 Stuttgart, Germany, Phone: 0711 6489 2760, E-Mail: Matthias.orth@vinzenz.de. <https://orcid.org/0000-0003-2881-8384>

Frank Klawonn: Helmholtz Centre for Infection Research, Braunschweig, Germany; and Ostfalia University, Wolfenbüttel, Germany

Georg Hoffmann: German Heart Center, Munich, Germany; and Trillium GmbH Medizinischer Fachverlag, Grafrath, Germany

methods for reference interval estimation once the data have been transformed properly.

Unknown proportion of abnormal results [3, 4, 7]. To overcome the problem of the unknown proportion of abnormal results, a common suggestion is to apply a power transformation of the form $x' = k \cdot x^\lambda$ (with $x' = \log(x)$ for $\lambda = 0$) making a skewed distribution more symmetric [3, 7]. The challenge here is, however, choosing the optimal exponent λ [7, 8].

Materials and methods

Patient test results

Routine laboratory results were queried from the database of the Vinzenz von Paul Kliniken, Stuttgart, Germany, and random samples with 1000 males and 1000 females, aged 18–50 years, were drawn from the first values after admission. Blood samples were collected from hospitalized patients and immediately sent to the central laboratory on an automatic track system. Ethylenediaminetetraacetic acid (EDTA) blood was used for blood counts on Sysmex XN10/20 analyzers (Sysmex, Kobe, Japan) and serum for the biochemical analyses on Abbott ci8200 Architect systems (Abbott, Chicago, IL, USA). Standard methods were applied for hemoglobin (photometric), leukocytes (conductometric), sodium and potassium (potentiometric), creatinine (photometric Jaffé kinetic), and alanine aminotransferase (ALAT) (photometric with pyridoxal phosphate).

Calculation of Pearson's moment coefficient of skewness and of Bowley's skewness

The classical Pearson's moment coefficient of skewness (PS) was calculated from eq. 1 and Bowley's skewness BS from Eq. 2:

$$PS = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (1)$$

$$BS = \frac{(Q3 - Q2) - (Q2 - Q1)}{Q3 - Q1} = \frac{Q1 - 2 \cdot Q2 + Q3}{Q3 - Q1} \quad (2)$$

where μ and σ represent the mean and the standard deviation, respectively, while $Q1$, $Q2$ and $Q3$ are the three quartiles partitioning the ordered values into four subsets of equal size. Assuming normality, the 95%

upper confidence limit for PS was calculated according to [9]:

$$CI_{PS} = 1.96 \cdot \sqrt{\frac{6 \cdot n \cdot (n-1)}{(n+1) \cdot (n-2) \cdot (n+3)}} \text{ which is about } \frac{4.8}{\sqrt{n}} \quad (3)$$

for large samples.

For BS, we obtained the following empirical function, using 10,000 Monte Carlo simulations:

$$CI_{BS} = \frac{2.6}{\sqrt{n}} \quad (4)$$

Calculation of transformation plot for symmetry developed by Emerson and Soto (ES)

For a method comparison, we investigated the transformation plot for symmetry developed by Emerson and Soto [10] as another quantile-based approach. This procedure (ES) aims to optimize the exponent λ for the aforementioned power transformation using a graphical, regression-based format:

$$\frac{q_u + q_l}{2} - q_{0.5} \approx (1 - \lambda) \cdot \frac{[(q_u - q_{0.5})^2 + (q_{0.5} - q_l)^2]}{4 \cdot q_{0.5}} \quad (5)$$

In this equation, $q_{0.5}$ is the median; q_l and q_u represent the lower and upper “letter values” [8], respectively, for quantiles with probabilities $F = 0.25$, $E = 0.125$, $D = 0.0625$ and so on. At least five pairs of such numbers $x = \frac{[(q_u - q_{0.5})^2 + (q_{0.5} - q_l)^2]}{4 \cdot q_{0.5}}$ and $y = \frac{q_u + q_l}{2} - q_{0.5}$ are calcu-

lated and plotted. Ideally, the points form a straight line with intercept 0 and slope $1 - \lambda$. If λ is close to 1 or 0, a normal, respectively lognormal, distribution can be assumed, where we use the natural logarithm in all our experiments here. Changing the base of the logarithm corresponds to a scaling and would not change the results in principle.

Estimation of reference intervals from routine laboratory data

Finally, we used the Bowley method as an upstream step for estimating reference intervals from routine laboratory results according to the following scheme:

1. Define distribution type (based on Bowley's skewness) and take logarithms if needed.
2. Remove abnormal outliers beyond the whiskers of the boxplot [8, 11].

3. Select algorithm to calculate μ and σ of the underlying normal distribution [4, 12].
4. Estimate the quantiles 0.025 and 0.975 (reference limits) from $\mu \pm 1.96 \cdot \sigma$ [4].

Assuming a nearly normal distribution for “non-diseased” values after step 1, step 2 can be expected to return mainly “normal” results without gross pathological outliers [11]. Numerous algorithms have been described for step 3, many of which are based on Robert G Hoffmann’s probability plot [13]. Here we use the more recent modification of Hoffmann et al. [4], which is based on a quantile plot, as well as a maximum likelihood estimator (EM algorithm for Gaussian mixture models), which is included in the *mclust* R package [12]. If the original data have been logaritized, the result must be antilogaritized.

Annex: R code and sample data

Essential functions for simulated and real data are included as R code in the “Supplementary material” (<https://doi.org/10.1515/labmed-2020-0005>) together with the example data used in this article.

Results

Figure 1 illustrates the principle of our method, using simulated Na and ALAT concentrations as typical examples

[14]: For normally distributed Na data, both skewness measures were about 0, whereas lognormally distributed ALAT data exhibited a Pearson skewness of 1.30 and a Bowley skewness of 0.17.

In order to demonstrate the robustness of Bowley’s skewness coefficient as compared to the classical Pearson skewness, we added one to 10 pathological outliers of 152 mmol/L to the simulated sodium data (Figure 1). Figure 2 shows that Pearson’s skewness exceeded 0.15 (red line), which is the upper limit of the 95% confidence interval for symmetry after addition of just three such outliers (0.3% of all values), thus leading to the erroneous assumption of a right-skewed distribution. In contrast, the quartile skewness remained at a level near zero and never exceeded the respective confidence interval of 0.08 for Bowley’s skewness.

To check the Bowley approach with real data, we applied the algorithm to routine blood counts, electrolyte concentrations and enzyme activities as shown in Table 1, and compared the results with the optimization of λ according to Emerson and Stoto (ES).

Taking a 95% confidence limit of $+\frac{2.6}{\sqrt{1000}} \approx 0.08$ for

BS as a cut-off for the detection of right-skewed data, we obtained the expected distribution types for Hb, Na (i.e. normal distributions) as well as for white blood cells (WBC) and ALAT (i.e. lognormal distributions). Surprisingly, K and Crea came out “intermediate” with respect to both sexes (Table 1, rows 4 and 5).

Figure 1: Boxplots and histograms of 1000 simulated sodium ($\mu=140$, $\sigma=2.5$) and ALAT values ($\mu \log=2.6$, $\sigma \log=0.49$, shift=2.5). The boxes extend from the 25th to 75th percentile (Q1, Q3) and the thick vertical line represents the median (Q2). If the distribution of the central 50% of the values is symmetric, the two halves of the box ($Q2 - Q1$ and $Q3 - Q2$) are equal and their difference – the numerator of the fraction – becomes 0. If the distribution is right-skewed, the right half will be larger than the left one, and the difference will become positive. The denominator $Q3 - Q1$ (interquartile range, IQR) serves to standardize BS to a range from -1 to $+1$.

Figure 2: Addition of an increasing number of pathological results (152 mmol/L) to the sodium data of Figure 1. The red line indicates the upper limit of the 95% confidence interval for the classical skewness of a normal distribution at $n=1000$.

Figure 3 shows that the boxplots and density curves actually reflect the results of this analysis quite nicely: Na and Hb appeared symmetric or left-skewed for both sexes, whereas WBC and ALAT were clearly right-skewed. The boxplots for K in males and Crea in females, however, were not quite symmetric, and the corresponding density curves exhibited shoulders on the right side, which obviously raised BS above the limit of 0.08.

Looking at these results in more detail, we observed in accordance with a publication of Haeckel and Wosniok [15] that taking logarithms of the original data had only a minor effect on Bowley's skewness of most analytes in Table 1 except for WBC and ALAT. These latter analytes were the only two, for which we definitely expected a lognormal distribution [14, 16]. Therefore, we would like to introduce a slight modification to the aforementioned algorithm. Bowley's skewness should be calculated from the original and the log-transformed data, and the difference between both skewness measures should be taken as a criterion with a threshold of 0.05 (Figure 4).

As to the method comparison depicted in Table 1, the proposed exponents of the ES method were close to 0 for WBC in both sexes, and close to 1 for Na and K in females, indicating the expected lognormal and normal distributions, respectively. Some λ values fell surprisingly far outside the interval of 0–1, reaching almost 20 for Na in men (as compared to 1 in women). If we assume that any markedly left-skewed distribution is due to pathological outliers on the left, we may set $\lambda > 1$ to $\lambda = 1$, in order to improve the agreement between the two methods. However, differences worth discussing remain for K in males and Crea in both sexes.

To investigate the practical application of our algorithm, we calculated reference intervals from original and log-transformed real data as described in the “Materials and methods” section. In a series of preparatory experiments, we determined the following optimal experimental conditions. Step 2 of the algorithm was repeated until no further outliers were detected. For the QQ plots in step 3, we calculated 39 quantiles with equidistant probabilities between 0.025 and 0.975. Linear regression lines were constructed from the central 27 dots. The intercepts were set to μ and the slopes to σ . From the results of the mclust function [12], we selected μ and σ from the subpopulation that made up the highest proportion of the Gaussian mixture. Figure 5 shows the results for Na (a typical normal distribution) and ALAT (a typical lognormal distribution).

The upper half of Figure 5 shows that, with regard to reference interval estimation, normally distributed analytes like sodium yield very robust results, irrespective of the distribution model and the method to fit that model. In contrast, the lower part demonstrates that lognormally distributed analytes like ALAT are susceptible to the distribution model chosen. The choice of the wrong model leads to a curved QQ plot and yields a very narrow reference interval with a too low upper limit.

Table 1: Bowley skewness (left) and proposed λ for power transformation (right) of routine laboratory results.

	Bowley's skewness				Distribution type		Proposed λ (ES)	
	Female	Female log	Male	Male log	Female	Male	Female	Male
Hb	0.00	−0.04	0.00	−0.03	Normal	Normal	3.71	3.45
WBC	0.11	−0.01	0.16	0.06	Lognormal	Lognormal	0.09	0.00
Na	−0.33	−0.34	−0.33	−0.34	Normal	Normal	1.00	19.67
K	0.20	0.17	0.00	−0.02	Lognormal	Normal	1.00	−1.61
Crea	0.09	0.05	0.07	0.02	Lognormal	Normal	0.42	−0.98
ALAT	0.27	0.12	0.21	0.01	Lognormal	Lognormal	0.30	0.34

Bold numbers indicate results to be discussed in detail.

Figure 3: Density and boxplots for routine laboratory results in women (pink) and men (blue).

Figure 4: Improved criterion using the difference of Bowley's skewness for original and log-transformed data.

We chose a cut-off at 0.05 by visual inspection to ensure that Na, K, Hb and Crea come out as normal distributions across both sexes.

Discussion

Our study shows that Bowley's quartile skewness is a simple and robust method to classify normality vs. lognormality in mixed populations (Table 1). In the simplest version, the 95% confidence interval for the skewness of normally distributed data may be used to make a safe classification of normally

distributed sodium [2] and hemoglobin [17] as well as lognormally distributed WBC [16] and ALAT [14] test results.

For K and Crea, the basic algorithm predicted different distributions for women and men, which were due to irregularities in the shapes of the density curves (Figure 3). Interesting enough, the distribution of potassium has been a matter of unresolved debate since the

Figure 5: Comparison of distribution models in the context of reference interval estimation.

Green lines indicate the confirmation of expected results, red lines indicate unexpected density curves and QQ plots. Left column normal distribution, right column lognormal distribution. Upper panel (A) sodium, lower panel (B) ALAT.

1950s [2, 18]. This finding underscores the statement of Ralph Graesbeck [1] – one of the fathers of the reference value concept – that “laboratory results distribute as ‘nature feels fit’ and that parametric (curve-constructing)

statistics only imitate the distribution with a function that allows calculation of desired informative indices”.

Nevertheless, we would like to suggest a slightly more sophisticated approach in order to avoid the

aforementioned discrepancies. Our results confirm an earlier observation [15] that the shape of the density curves does not change very much upon log transformation if the data is normally distributed with a relatively small biological variation. Consequently, the Bowley skewness will not change very much either, and thus the differences should be close to 0. Choosing a maximal difference of 0.05 as a cut-off will lead to homogenous classification results for K and Crea (Figure 4).

Power transformations using optimized exponents have been suggested by some authors as alternatives to model such intermediate distributions more correctly [3, 7]. From our experience, these approaches can be helpful but can also be misleading: Especially with regard to the ES method [8, 10] tested here, some results do in fact fit while others are not plausible at all. This observation confirms an earlier finding that the ES method may behave poorly with skewed data [19]. So, given the fact that log transformation may be applied anyway without a great risk of false results (Figure 5 and ref [15]), the question should be asked whether it is worthwhile to determine a λ between 0 and 1 with complex and error-prone methods.

In a final series of experiments, we tested whether choosing an appropriate distribution model had an influence on the indirect estimation of reference limits from routine data. Investigating two representative examples (i.e. ALAT and Na) in detail (Figure 5), we could show that this is indeed the case for the log-normally distributed transaminase ALAT, whereas no difference was observed for the normally distributed electrolyte sodium. This again confirms the suggestion of Haeckel and Wosniok that “unknown distributions of clinical chemical quantities should be considered to be log-normal” [15].

It is noteworthy that the robustness against outliers of Bowley's quartile skewness is essential to decide for the right data transformation. Because we assume that the decision for the transformation should be carried out before the removal of outliers, classical non-robust hypothesis tests for normality such as the Shapiro-Wilk or the Kolmogorov-Smirnov test will necessarily fail. For our examples, these tests for normality reported very small p-values (all of them except for WBC were much smaller than 0.007) both for the untransformed and the transformed data, leading to the rejection of the null hypotheses of normally and of lognormally distributed data.

As a side observation, our literature research revealed that the statistical distribution of laboratory results was an active research topic in the past century [e.g. 13, 16–18], while it is currently not in the focus of laboratory

medicine. In the future, we expect an increasing interest in this topic again, e.g. in the context of big data applications and standardized storage of results in electronic patient records [6, 20, 21]. Our method opens the possibility to easily review the distribution of laboratory values on a large scale.

Research funding: None declared.

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

Competing interests: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: The local Institutional Review Board (Heidelberg University, Medical Faculty of Mannheim) deemed the study exempt from review. Ethical approval has been obtained by the Ethical Committee of the Medical Faculty at the Ruprecht-Karls-Universität, Medical Faculty of Mannheim, Germany.

References

1. Graesbeck R. The evolution of the reference value concept. *Clin Chem Lab Med* 2004;42:692–7.
2. Feldman M, Dickson B. Plasma electrolyte distributions in humans – normal or skewed. *Am J Med Sci* 2017;354:453–7.
3. Jones G, Haeckel R, Loh T, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination – review and recommendations. *Clin Chem Lab Med* 2018;57:20–9.
4. Hoffmann G, Lichtinghagen R, Wosniok W. Simple estimation of reference intervals from routine laboratory data. *J Lab Med* 2015;39:389–402.
5. Haeckel R, Wosniok W, Gurr E, Peil B. Permissible limits for uncertainty of measurement in laboratory medicine. *Clin Chem Lab Med* 2015;53:1161–71.
6. Hoffmann G, Klawonn F, Lichtinghagen R, Orth M. The zlog value as a basis for standardization of laboratory results. *J Lab Med* 2018;41:23–32.
7. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, et al. A plea for intra-laboratory reference limits. Part 2. *Clin Chem Lab Med* 2007;45:1043–57.
8. Hoaglin D, Mosteller F, Tukey J. Understanding robust and explanatory data analysis. Wiley Classics Library Edition. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2000, ISBN 0-471-38491-7.
9. Kendall M, Stuart A. The advanced theory of statistics, volume 1: distribution theory, 3rd ed. Griffin 1969. ISBN 0-85264-141-9.
10. Emerson J, Stoto M. Explanatory methods for choosing power transformations. *J Am Stat Assoc* 1982;77:103–8.
11. Horn P, Feng L, Li Y, Pesce A. Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem* 2001;47:2137–45.

12. Scrucca L, Fop M, Murphy B, Raftery A. Clustering, classification and density estimation using Gaussian finite mixture models. *R J* 2016;8:205–33.
13. Hoffmann R. Statistics in the practice of medicine. *J Am Med Assoc* 1963;185:150–9.
14. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. *Clin Chem Lab Med* 2019;57:1933–47.
15. 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.
16. Nieto F, Szklo M, Folsom A, Rock R, Mercuri M. Leukocyte count correlates in middle-aged adults. *Am J Epidemiol* 1992;136:525–37.
17. Meyers L, Habicht J, Johnson C. Components of the difference in hemoglobin concentrations in blood between black and white women in the united states. *Am J Epidemiol* 1979;109:539–49.
18. Fawcett J, Wynn V. Variation of plasma electrolyte and total protein levels in the individual. *British Med J* 1956;2:582–5.
19. Cameron M. Choosing a symmetrizing power transformation. *J Am Statist Assoc* 1984;79:385, 107–8.
20. Orth M, Aufenanger J, Hoffmann G, Lichtinghagen R, Stiegler Y, Peetz D. Possibilities and risks of e-health in laboratory medicine. *J Lab Med* 2016;40:227–37.
21. Shaw J, Cohen A, Konforte D, Binesh-Marvasti T, Colantonio DA, 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.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/labmed-2020-0005>).