Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression

Steffen Martens[1], Katy Klauenberg[1], Jörg Neukammer[1], Simon Cowen[2], Stephen L. R. Ellison[2], and Clemens Elster[1]

[1]PTB, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany
[2]LGC, Laboratory of the Government Chemist, Queens Road, Teddington, TW11 0LY, UK

16 June 2020

1 Summary

In metrology, often two methods measuring the same quantity are to be judged whether or not they are in agreement. For measurements across a whole range of values, this can be done by comparing their straight-line fit to the identity line. Such a comparison is only meaningful, when uncertainties are available. Furthermore, the estimates of the straight-line fit and their uncertainties are only reliable when all sources of uncertainty have been accounted for. In particular, the measurements of both methods in a comparison are usually uncertain, and common instruments or standards cause correlation among or between them.

When fitting a straight-line relation, the weighted total least-squares (WTLS) method accounts for correlation and uncertainties in both variables. This example focuses on WTLS and defines a measurement model from it to propagate all uncertainties and correlations through to the estimate of the slope and intercept, and associate uncertainties with them according to the GUM. Using the example of two high accuracy methods measuring the total haemoglobin concentration in blood, i.e. the cyanmethaemoglobin and alkaline haematin method, we indicate how correlations can be inferred, demonstrate how they can be accounted for and show their impact on the regression. The results are discussed and recommendations are given.

2 Introduction of the application

The total haemoglobin (Hb) concentration in blood is one of the most frequently measured analytes in clinical medicine because of its significance for evaluating the state of health of a human. The medical need for this analyte and the different spectrophotometric methods applied are summarized in Appendix A. For external quality assurance of routine laboratories, interlaboratory comparisons are performed in which the deviation from the reference value may not exceed 6 % [1]. To evaluate such round robin tests, ideally reference or "higher order" measurement procedures allowing for standard uncertainties smaller than 0.6 % (an order of magnitude below the allowable deviations)

are required. The cyanmethaemoglobin (HiCN) method is the internationally accepted, spectrophotometric reference method [2–4] to determine the total Hb concentration. Critical issues of the HiCN method are the toxicity of the potassium cyanide involved and that it is not traceable to the International System of Units. An alternative spectrophotometric procedure for the determination of reference values for this quantity is the non-cyanide, alkaline haematin (AHD) method. Among other advantages, the AHD method has the potential as a primary method [5, 6] since a primary calibrator exists.

Previous comparisons of the HiCN and the AHD method with high-accuracy procedures [5, 7] demonstrate a good agreement, but are limited to only one blood sample with a Hb concentration in the normal range, i.e. a healthy person. Studies based on protocols for routine diagnostics[1] also show a good agreement between both methods (see [8–10] and references therein) and rely, among others, on the regression of a straight-line relationship. However, these comparisons do not consider the uncertainty of measurements. Estimates of regression parameters will usually differ when all uncertainties are accounted for. In addition, these comparison studies do not provide an uncertainty for the regression estimates. It is thus difficult to compare the results of these studies and to quantitatively judge the agreement between the reference and the alternative AHD method.

This example demonstrates how the uncertainties of HiCN and AHD measurements, including correlation, can be propagated to give the uncertainty of their straight-line relation. The total Hb concentrations are used, which PTB measured with both the HiCN and the AHD method for $P = 104$ blood samples over the past 10 years. The data cover the whole range from 60 $\mathrm{g\,L^{-1}}$ to 190 $\mathrm{g\,L^{-1}}$ relevant in clinical diagnosis and include pathologically low as well as pathologically high Hb concentrations. These measurements and their associated uncertainties, say x_p, $u(x_p)$ and y_p, $u(y_p)$, are displayed in figure 1 and can be found online in repository [11]. Derivation of the total Hb concentration involves quantities common to both methods and all samples (cf. Appendix B for background information). Some of these common quantities contribute significantly to the uncertainty of the Hb concentration [5, 7]. Therefore, it is reasonable to suspect significant correlation among the HiCN as well as among the AHD method (cf. clause 5.2.4 in [12]).

Also beyond method comparison, uncertainty in all variables of a regression and correlation among or between them is prevalent in metrology. For example in calibrations, the reference and the device under test usually both display uncertainty. Additionally, measurements over the range of use are often performed with the same measuring instrument or physical standard which often contribute a considerable amount of uncertainty.

This example focusses on a measurement model that is based on the weighted total least-squares (WTLS) method. The measurement model allows for uncertainty evaluation following the GUM. The WTLS method accounts for uncertainties in both variables of a regression, as well as, for correlation among and between them. WTLS is recommended by multiple standards [13, 14] and applied in metrology (e.g. Refs. [15–17]).

3 Specification of the measurand

Let X denote the total Hb concentrations obtained by HiCN and Y the corresponding quantity measured by the AHD method. The straight-line relation

$$Y = \beta_0 + \beta_1 X \tag{1}$$

[1] In routine applications only one value for the absorbance is measured, while reference procedures include dilution series, repeat measurements and centrifugation to reduce uncertainties.

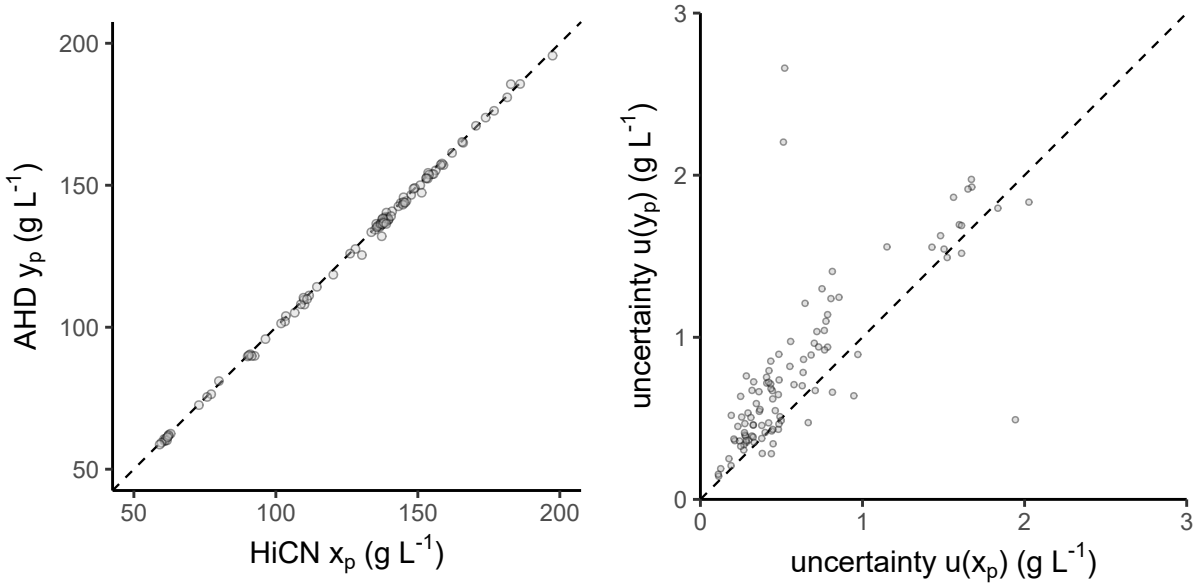

Figure 1: Left: Visualization of haemoglobin concentration measurements x_p, y_p performed at PTB on $P = 104$ blood samples by the two methods HiCN and AHD.
Right: Standard uncertainties $u(x_p)$ and $u(y_p)$ for both methods and all samples. These measurement results are available online in repository [11]. In both panels, the dashed line represents the identity $y = x$ and the markers are drawn as transparent; thus, overlayed markers appear darker.

is assumed to model the relationship between the measured values of both methods, and is supported by previous studies comparing the HiCN and the AHD method (see [8–10] and references therein). The measurands are the intercept parameter β_0 and slope parameter β_1 of the straight-line model (1). If both methods measure the same, uniquely defined quantity, one usually obtains estimates close to $\widehat{\beta}_0 = 0$ and $\widehat{\beta}_1 = 1$.

The input quantities influencing the measurands are the P pairs (X_p, Y_p). Estimates of these inputs are the Hb concentration measurements of each method, x_p and y_p. Standard uncertainties $u(x_p)$ and $u(y_p)$ of these inputs are of the same magnitude (cf. figure 1). In addition, any two inputs X_p, X_q are correlated due to the use of common standards in their measurement, especially of the same molar extinction coefficient ϵ and corrections C_0, C_1 (as detailed in Appendix B). The covariance matrix $\mathbf{U}_x$ shall contain these correlations as well as the standard uncertainties $u(x_p)$. Likewise, the covariance matrix $\mathbf{U}_y$ contains the correlations and standard uncertainties $u(y_p)$ for the inputs Y_p. (For the definition of a covariance matrix, we refer to clause 3.11 in the supplement 1 to the GUM [18].)

4 Measurement model

The measurement model for straight-line regression can be constructed from the appropriate least-squares method. The frequently applied ordinary and weighted least-squares method are inappropriate here because they assume that the measured values of one method are exact. Notably, regressing one method over the other will generally result in different estimates than the other way around; especially, when the uncertainties of both methods are similar and non-negligible – as for HiCN and

AHD. The measurand would thus be ambiguous. Also Deming regression [19] and Passing-Bablok regression [20], two common methods for method comparison, are not appropriate for this data set. First, the uncertainties $u(y_p)$ cannot be expressed as a common multiple of $u(x_p)$ as Deming regression requires (see right panel in figure 1); second, it is important to take account of applicable uncertainty and covariance information where possible and Passing-Bablok regression does not use information on uncertainties.

Weighted total least-squares (WTLS) is the method recommended by multiple standards [13, 14] when the uncertainty associated with the measured values x_p and y_p are both non-negligible. It also addresses correlation. The WTLS method is based on minimizing the generalized sum of squares

$$Q = \left(\mathbf{x} - \widetilde{\boldsymbol{\xi}}\right)^\top \mathbf{U}_x^{-1} \left(\mathbf{x} - \widetilde{\boldsymbol{\xi}}\right) + \left(\mathbf{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right)^\top \mathbf{U}_y^{-1} \left(\mathbf{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right) \tag{2}$$

with respect to $\widetilde{\beta}_0$, $\widetilde{\beta}_1$ and the unknown, "true" values of $\mathbf{x}$ called $\widetilde{\boldsymbol{\xi}}$. Here, the vector $\mathbf{x}$ contains the elements $\mathbf{x} = (x_1, \ldots, x_P)^\top$ and the vectors $\mathbf{y}$ and $\widetilde{\boldsymbol{\xi}}$ are likewise defined. The minimizer of (2) defines the solution $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\xi}}^\top)$ of the WTLS method.

The measurement model is then defined by replacing the estimates $\mathbf{x}$ and $\mathbf{y}$ in the minimization of Q by the underlying quantities $\mathbf{X} = (X_1, \ldots, X_P)^\top$ and $\mathbf{Y} = (Y_1, \ldots, Y_P)^\top$, respectively. That is,

$$(\beta_0, \beta_1, \boldsymbol{\xi}^\top)^\top = \arg \min_{\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\boldsymbol{\xi}}} \left\{ \left(\mathbf{X} - \widetilde{\boldsymbol{\xi}}\right)^\top \mathbf{U}_x^{-1} \left(\mathbf{X} - \widetilde{\boldsymbol{\xi}}\right) + \left(\mathbf{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right)^\top \mathbf{U}_y^{-1} \left(\mathbf{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right) \right\}, \tag{3}$$

where only (β_0, β_1) define the measurand.

5 Estimation and uncertainty evaluation

Following the GUM [12, 21], estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the measurands are obtained by evaluating measurement model (3) at the estimates $\mathbf{x}$ and $\mathbf{y}$ of the input quantities $\mathbf{X}$ and $\mathbf{Y}$. The uncertainties associated with $(\widehat{\beta}_0, \widehat{\beta}_1)$ result from propagating the uncertainties in $\mathbf{U}_x$ and $\mathbf{U}_y$ associated with the estimates of the input quantities through this measurement model.

Measurement model (3) is implicit, multivariate, non-linear and usually no closed form is available for its solution. An iterative scheme for deriving estimates and their associated uncertainties is described in clause 10 of the standard [13]. This simple scheme also provides correlations between β_0 and β_1, and is valid for any covariance matrices $\mathbf{U}_x$ and $\mathbf{U}_y$ whose eigenvalues are all positive.

Assuming a Gaussian distribution[2], a 95 % coverage interval for each measurand β_i with $i = 0, 1$ is given by

$$\left[\widehat{\beta}_i - 1.96\, u\left(\widehat{\beta}_i\right), \widehat{\beta}_i + 1.96\, u\left(\widehat{\beta}_i\right)\right].$$

A two-dimensional, joint 95 % coverage region can be calculated following clause 6.5.2 in [21].

In order to estimate the slope and intercept of a straight-line relation as well as valid uncertainties and/or coverage intervals, the full covariance matrices $\mathbf{U}_x, \mathbf{U}_y$ and possible cross-correlation between $\mathbf{X}$ and $\mathbf{Y}$ need to be known. Annex D in [13] describes how these covariances can be calculated for common, simple measurement models. For more involved measurement models, like for HiCN and AHD measurements, we recommend the Monte Carlo method [21], where distributions for all input quantities are propagated through a joint measurement model to arrive at the $2P$-dimensional, joint distribution for the outputs $\mathbf{X}$ and $\mathbf{Y}$.

[2] Cf. section 10.2.3 in [13] for the approximate validity of this Normality assumption.

The uncertainty in HiCN and in AHD measurements is dominated by a common quantity, namely the molar extinction coefficient ϵ (see [5, 7]). We thus suspect that the covariance matrices $\mathbf{U}_x$ and $\mathbf{U}_y$ are governed by a common correlation coefficient ρ. That is, we set their elements $\mathbf{U}_{x,pq} = \rho u(x_p)u(x_q)$ and $\mathbf{U}_{y,pq} = \rho u(y_p)u(y_q)$ for all $p \neq q$. The diagonal elements contain the variances, i.e. $\mathbf{U}_{x,pp} = u^2(x_p)$ and $\mathbf{U}_{y,pp} = u^2(y_p)$. Further details are given in Appendix C. First Monte Carlo evaluations of the joint uncertainty budget showed that correlation coefficients up to $\rho = 0.8$ may be realistic. Details on how to jointly evaluate the correlation, uncertainties and estimates for the input quantities of least-squares methods applying the Monte Carlo method are illustrated in [22]. The correlation between HiCN and AHD is dominated by two common quantities, viz., the cuvettes' absorption length d and the mean molar mass M(Hb). According to [5], the amount of cross-correlation is much smaller compared to correlation between the estimates x_p and x_q as well as between the estimates y_p and y_q. We assume zero cross-correlation throughout this example. Note that the results reported below are conditional on the plausibility of this correlation structure. The real correlation structure and amount could be different and is to be inferred from the quite complex measurement model described in Appendix B.

6 Reporting the result

Let us now apply the measurement model (3) to the estimates and uncertainties presented in figure 1 and to the above covariance structures $\mathbf{U}_x$ and $\mathbf{U}_y$. For selected correlation coefficients ρ, the results are listed in table 1. The estimate, associated standard uncertainty and the covariance for the measurands β_0 and β_1 are obtained by the algorithm in clause 10 of [13] and application of the law of propagation of uncertainty [12]. R Markdown [23] code for this algorithm is available online in repository [11]. Figure 2 depicts the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and the corresponding 95 % coverage interval.
Nearly identical results have been obtained by applying the Monte Carlo method [21] to the measurement model (3) and the algorithm in Ref. [13]. The non-linearity of (3) could cause differences; however, this was not observed. Software is available that implements WTLS and propagates uncertainties. For example, the CALIBRATION CURVE COMPUTING Software provided by INRIM [16] also produces the results in table 1, although a slightly different algorithm is implemented (which relies on an implicit set of normal equations).

Before interpreting the results of a regression, the data as well as the assumptions contributing to the analysis should be assessed critically. For instance, graphically analysing the (weighted) residuals did not indicate a violation of the straight-line assumption (1), since no systematic behaviour of these

Table 1: Results obtained by weighted total least-squares with uncertainty evaluation according to the GUM for varying correlation coefficients ρ. Listed are the estimates and uncertainties for slope and intercept.

Correlation a.u.	$\widehat{\beta}_0$ $\mathrm{g\,L^{-1}}$	$u(\widehat{\beta}_0)$ $\mathrm{g\,L^{-1}}$	$\widehat{\beta}_1$ a.u.	$u(\widehat{\beta}_1)$ a.u.	$\mathrm{cov}(\widehat{\beta}_0,\widehat{\beta}_1)$ 10^{-3} a.u.
$\rho = 0.0$	-0.488 6	0.166 7	0.998 4	0.001 6	-0.24
$\rho = 0.6$	-0.489 4	0.105 5	0.998 6	0.001 2	-0.10
$\rho = 0.8$	-0.489 4	0.074 6	0.998 6	0.000 9	-0.05

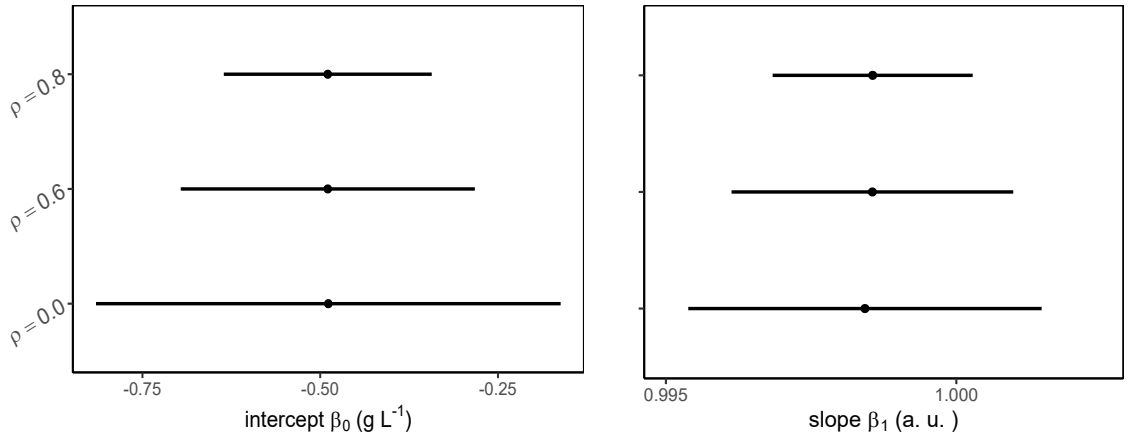

Figure 2: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (dots) and their 95 % coverage intervals for the weighted total least-squares regression results listed in table 1.

residuals were observed. A significant outcome of the χ^2 test, whose application is recommended in standard [13], does not necessarily indicate departures from the linearity assumption. The χ^2 test assesses, whether the (weighted) residuals are independently normally distributed – an assumption which is not required for WTLS estimation and measurement model (3). Any observed test statistics which exceed the 95 % quantile of the χ^2-distribution, are suspected to be due to non-normally distributed residuals rather than a violation of the straight-line assumption (1). The former does not contradict the assumptions of our analysis.

7 Discussion and conclusion

This example demonstrates how two measurement methods can be compared to judge whether both measure the same quantity over a defined measurement range. If the uncertainties of both methods are non-negligible, ordinary and weighted least-squares methods are inappropriate. Instead, weighted total least-squares is a suitable method which allows for an uncertainty propagation when embedded in a measurement model in line with the GUM.

Using the example of measuring the total haemoglobin concentration in blood, it is reasoned that correlation among and possibly between two measurement methods is not unusual, and likely to be rather frequent in metrology in general. We indicate how these correlations can be inferred and select a common correlation structure for this example.

The reader observes a small but significant offset between the HiCN and AHD method for measuring haemoglobin – irrespective of the amount of correlation. The slope of the linear relation between both methods is compatible with unity for all reasonable values of correlation, but would be significantly smaller than one for higher correlations $\rho \geq 0.9$. For the assumed correlation structure, the estimates of the linear relation vary little with the amount of correlation. However, their uncertainty changes by the factor $\sqrt{1-\rho}$, i.e. it reduces to two thirds for a correlation coefficient of $\rho = 0.6$ and to a half for $\rho = 0.8$, compared to WTLS estimation without correlation. Also the covariance between $\widehat{\beta}_0$ and $\widehat{\beta}_1$ scales with $1-\rho$. In addition, the estimates change with varying correlation coefficient when fewer observations are available. These relationships are detailed in Appendix C. Other correlation structures, for instance when the correlation within one method is much larger

than within the other method, will also change the estimates.

Our analyses show that the HiCN method leads to slightly higher Hb concentrations than the AHD method, if the correlation structure and the amount of correlation are realistic. This has been observed before ([24] and references therein) and may be caused by a background due to bilirubin. However, the differences between the HiCN and the AHD method are sufficiently small. If the correlation assumptions can be confirmed in future, both methods could be applied to determine higher order measurement values to evaluate round robin tests for external quality assurance in laboratory medicine.

We conclude that only stating the uncertainty of a fitted (linear) relation allows for a quantitative comparison of two methods over their measurement range. To derive these uncertainties reliably and to give valid estimates, it is important to account for correlation among and between the measurement methods. Otherwise, the conclusions drawn from such a comparison study could differ and become unreliable.

Acknowledgment

This work is part of the project 17NRM05 “EMUE” which has received funding from the EMPIR programme co-financed by the Participating States and from the European Unions’ Horizon 2020 research and innovation programme. Part of the experiments for high accuracy Hb concentration determination were supported by the funding programme “Messen, Normen, Prüfen und Qualitätssicherung” (MNPQ) of the Federal Ministry for Economic Affairs and Energy and by the European Metrology Research Programme (EMRP) HLT-05 2012 “Metallomics”.

References

[1] *Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen*. Deutsches Ärzteblatt, 2019. See also the unauthorised translation of the previous version of the Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK. J. Lab. Med., 39(1): 26–69, 2019.

[2] S. M. Lewis and S. Kumari. Chapter 7: Haemoglobinometry. In *Guidelines on standard operating procedures for haematology*. WHO, 1999.

[3] BS 3985:2003 Haemoglobincyanide (cyanmethhaemoglobin) preparation as a standard for spectrometric haemoglobin. BS, British Standards Institution, BSI Group, UK, 2003.

[4] DIN 58931:2010 Haematology – determination of haemoglobin concentration in blood – reference method. DIN, German Standards Institution, Beuth-Verlag, Berlin, Germany, 2010.

[5] C. Frank, C. Brauckmann, M. Palos, C. G Arsene, J. Neukammer, M. E. del Castillo Busto, S. Zakel, C. Swart, B. Güttler, and R. Stosch. Comparison of potential higher order reference methods for total haemoglobin quantification – an interlaboratory study. *Anal. Bioanal. Chem.*, 409(9):2341–2351, 2017.

[6] ISO 17511 In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials. ISO, International Organization for Standardization, Geneva, Switzerland, 2003.

[7] K. Witt, H. U. Wolf, C. Heuck, M. Kammel, A. Kummrow, and J. Neukammer. Establishing traceability of photometric absorbance values for accurate measurements of the haemoglobin concentration in blood. *Metrologia*, 50(5):539–548, 2013.

[8] N. M. M. Moharram, R. El Aouad, S. Al Busaidy, A. Fabricius, S. Heller, W. G. Wood, H. U. Wolf, and C. C. Heuck. International collaborative assessment study of the AHD[575] method for the measurement of blood haemoglobin. *E. Mediterr. Health J.*, 12(5):522–534, 2006.

[9] S. Parikh, B. Parikh, C. Shah, P. Bhansali, J. Patel, and D. Joshi. Haemoglobinometry by a novel alkaline haematin detergent-575 method. *Gujarat Medical Journal*, 65(1):14–19, 2010.

[10] V. T. Anchinmane and S. V. Sankhe. Evaluation of hemoglobin estimation with non-cyanide alkaline haematin D-575 method. *Int. J. Res. Med. Sci.*, 4(10):4297–4299, 2016.

[11] S. Martens, K. Klauenberg, J. Neukammer, S. Cowen, S. L. R. Ellison, and C. Elster. 2020. Available at https://zenodo.org/communities/emue.

[12] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections*. BIPM, 2008.

[13] ISO/TS 28037 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010.

[14] ISO/TS 28038 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018.

[15] M. Krystek and M. Anton. A least-squares algorithm for fitting data points with mutually correlated coordinates to a straight line. *Meas. Sci. Technol.*, 22(3):035101, 2011.

[16] A. Malengo and F. Pennecchi. A weighted total least-squares algorithm for any fitting model with correlated variables. *Metrologia*, 50(6):654, 2013.

[17] R Feistel, J. W Lovell-Smith, P Saunders, and S Seitz. Uncertainty of empirical correlation equations. *Metrologia*, 53(4):1079, 2016.

[18] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement' – Propagation of distributions using a Monte Carlo method, JCGM 101:2008*. BIPM, 2008.

[19] W. E. Deming. *Statistical adjustment of data*. New York: Wiley, 1943.

[20] H. Passing and W. Bablok. A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I. *J. Clin. Chem. Clin. Biochem.*, 21:709–720, 1983.

[21] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement' – Extension to any number of output quantities, JCGM 102:2011*. BIPM, 2011.

[22] S. Martens, K. Klauenberg, B. Mickan, C. Yardin, N. Fischer, and C. Elster. 2020. Available at https://zenodo.org/communities/emue.

[23] J. J Allaire, Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. *rmarkdown: Dynamic Documents for R*, 2020. R package version 2.1.

[24] R. Zander, W. Lang, and H. U. Wolf. Alkaline haematin D-575, a new tool for the determination of haemoglobin as an alternative to the cyanhaemiglobin method. *Clinica chimica acta*, 136(1):83–93, 1984.

[25] R. Chaudhary, A. Dubey, and A. Sonker. Techniques used for the screening of haemoglobin levels in blood donors: current insights and future directions. *J. Blood. Med.*, 8:75–88, 2017.

[26] CIBA-Geigy AG. *Wissenschaftliche Tabellen Geigy: Teilband Hämatologie und Humangenetik*. Basel, 8th edition, 1979. 4. Nachdruck 1985.

[27] R. Enns. Capsule endoscopy in patients with iron deficiency. *J. Gastroenterol. Hepatol.*, 8:847–849, 2012.

[28] N. J. White. Anaemia and malaria. *Malaria Journal*, 17:371–387, 2018.

[29] M. Usman, M. Moinuddin, and S.A. Ahmed. Role of iron deficiency anemia in the propagation of beta thalassemia gene. *Korean J. Hematol.*, 46:41–44, 2011.

[30] T. Harvey, A. Zkik, M. Auges, and T. Clavel. Assessment of iron deficiency and anemia in pregnant women: an observational french study. *Women's Health*, 12:95–102, 2016.

[31] T. Srivastava, H. Negandhi, S.B. Neogi, J. Sharma, and R. Saxena. Methods for haemoglobin estimation: A Review of "What Works". *J. Hematol. Transfus.*, 2:1028–1034, 2014.

[32] A. da Silva Pereira, I. R. Ribeiro da Castro, F. F. Bezerra, J. F. N. Neto, and A. C. Feldenheimer da Silva. Reproducibility and validity of portable haemoglobinometer for the diagnosis of anaemia in children under the age of 5 years. *J. Nutr. Sci.*, 9:e3, 2020.

[33] D. Drabkin and J.H. Austin. Spectrophotometric studies: I. Spectrophotometric constants for common haemoglobin derivatives in human, dog, and rabbit blood. *J. Biol. Chem.*, 98:719–733, 1932.

[34] I. Oshiro, T. Takenaka, and J. Maeda. New method for haemoglobin determination by using sodium lauryl sulfate (SLS). *Clin. Biochem.*, 15:83–88, 1982.

[35] A. Karsan, I. Maclaren, D. Conn, and L. Wadsworth. An evaluation of haemoglobin determination using sodium lauryl sulfate. *Am. J. Clin. Pathol.*, 100:123–126, 1993.

A Haemoglobin concentration: Importance and determination

The total haemoglobin (Hb) concentration in blood is part of the complete blood count, which is one of the most frequently measured analytes in clinical medicine. For example, Hb concentrations are needed for screening blood donors to protect their health and to guarantee the quality of the blood product [25]. Deviations of the Hb concentration from the normal range (137 $\mathrm{g\,L^{-1}}$ – 162 $\mathrm{g\,L^{-1}}$ for men and 123 $\mathrm{g\,L^{-1}}$ – 145 $\mathrm{g\,L^{-1}}$ for women; c.f. [26, table 4, p. 190]) are observed for various diseases. Further diagnostics are initiated to identify the origin of such an anomaly. Iron deficiency

could be caused by bleeding in the gastrointestinal tract [27], malaria [28] or thalassemia, the most common genetic disorder worldwide [29]. In addition, haemoglobin concentration is relevant to manage iron deficiency in pregnant women [30].

Total haemoglobin concentration is determined by a variety of methods [31], depending on the specific medical application. In countries where anaemia is widespread, portable instruments are used to estimate haemoglobin concentration using capillary blood for analysis [32]. Measurements with higher precision and accuracy compared to such point-of-care instruments are routinely performed in laboratory medicine and require venous blood and chemical conversion of the different haemoglobin variants to a stable end product, which is subsequently spectrophotometrically analysed. Conversion to cyanmethaemoglobin (HiCN), first applied by Drabkin and Austin [33], has been considered as a gold standard for routine applications [31] and is also internationally accepted as higher-order method [2–4] to determine reference measurement values in external quality assurance of medical laboratories [1]. However, because of the toxicity of the potassium cyanide involved, the HiCN method is not allowed in most countries and has been replaced by the sodium lauryl sulfate (SLS) procedure [34, 35].

Typically, in laboratory medicine accuracies below 6 % shall be reached for Hb concentration measurements. This value is stated in the guideline of the German Medical Association for Quality Assurance in Medical Laboratory Examinations [1] and indicates the maximum allowable deviation to pass the ring trials mandatory in Germany. To evaluate such external quality assurance schemes, so-called "higher-order measurement methods" or reference procedures are required providing results with expanded uncertainties (95 % confidence level) possibly smaller than 1.5 %. This requirement is specified in DIN 58931 [4, p. 18] and was met in comparison experiments [5, 7]. For such higher-order procedures the same reagents may be used to convert the different Hb variants to a stable end product. Lower uncertainties are achieved by gravimetrical preparation of dilution series and centrifugation to suppress the scattering of residual white blood cells or agglomerates of membranes of erythrocytes. In addition, high-accuracy absorbance measurements are required, traceable to a national standard [4]. Although the HiCN method is frequently used as a reference method for comparison when evaluating new procedures for the determination of the total Hb concentration, it is presently not traceable to the International System of Units. In particular, material suited as primary calibrator is not available, it is known that verdoglobin is not converted to HiCN and that background due to bilirubin can cause systematic deviations towards higher concentrations. It follows that according to the ISO standard 17511 on metrological traceability [6] the HiCN method can be characterised as an international conventional reference measurement procedure. An alternative spectrophotometric procedure for the determination of total Hb concentration is the non-cyanide, alkaline haematin (AHD) method. In contrast to the HiCN procedure, when applying the AHD method verdoglobin is converted to the end product chlorohaemin and the sensitivity against bilirubin perturbations is much smaller. In addition, the globin protein is destructed and solutions of the end product, the well-defined molecule chlorohaemin, might serve as primary calibrator. Hence, the AHD method may have the potential as a primary method [5, 6].

B Details of the measurement methods for haemoglobin concentration

The HiCN and the AHD method both rely on the measurement of the spectral absorbance. The photometrical traceability is established by correcting the measured absorbance values[3] $a_{i,p}^{k}$ using

[3]Each absorbance value $a_{i,p}^{k}$ in turn is based on a series of repeated measurements and its uncertainty is evaluated following the GUM.

the linear relationship $C_{0,k} + C_{1,k} a_{i,p}^k$ for $k \in \{\text{HiCN}, \text{AHD}\}$, blood sample $p = 1, \ldots, P$ and dilution i (cf. [7]). As recommended in DIN 58931:2010 [4], at least four dilutions ϕ_i of each blood sample are prepared and the associated Hb mass fractions $w_{i,p}^k$ are calculated according to

$$w_{i,p}^k = \frac{\left(C_0^k + C_1^k a_{i,p}^k\right) M(\text{Hb})}{d \epsilon^k \phi_i}. \tag{4}$$

Here, d represents the absorption length of the rectangular spectrophotometric cuvette, ϵ^k is the molar decadic absorption coefficient of the reaction product and $M(\text{Hb})$ is the mean molar mass of one Hb subunit. The estimates and associated uncertainties for the input quantities in (4) can be found in [7]. The final reported total Hb concentration for each sample and method, x_p and y_p, are determined by a weighted average of the Hb mass fractions $w_{i,p}^k$ over the dilutions i. The associated uncertainties are discussed in detail in Ref. [7].

C Influence of correlation for a common structure

If the covariance matrix for the HiCN method is given by

$$\mathbf{U}_x = (1-\rho)\text{diag}(\mathbf{u}_x^2) + \rho \mathbf{u}_x \mathbf{u}_x^\top$$

with $\mathbf{u}_x = (u(x_1), \ldots, u(x_P))^\top$, as described in section 5, the inverse of $\mathbf{U}_x$ is determined by

$$\mathbf{U}_x^{-1} = \frac{1}{1-\rho}\left[\text{diag}\left(\frac{1}{u^2(x_1)}, \ldots, \frac{1}{u^2(x_P)}\right) - \frac{1}{P-1+1/\rho}\left(\frac{1}{\mathbf{u}_x}\right)\left(\frac{1}{\mathbf{u}_x}\right)^T\right].$$

The inverse $\mathbf{U}_y^{-1}$ can be determined by analogy. Then, the generalized sum of squares (2) simplifies to

$$Q = \frac{1}{1-\rho}\left(\sum_{p=1}^{P} \frac{(x_p - \xi_p)^2}{u^2(x_p)} + \frac{(y_p - \beta_0 - \beta_1 \xi_p)^2}{u^2(y_p)} - \frac{1}{P-1+1/\rho}\left(\sum_{p,q=1}^{P} \frac{(x_p - \xi_p)(x_q - \xi_q)}{u(x_p)u(x_q)} + \sum_{p,q=1}^{P} \frac{(y_p - \beta_0 - \beta_1 \xi_p)(y_q - \beta_0 - \beta_1 \xi_q)}{u(y_p)u(y_q)}\right)\right).$$

The factor $1/(1-\rho)$ is irrelevant for the optimization of Q and thus does not influence the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$. At the same time it influences the uncertainties $u(\widehat{\beta}_i)$ and the covariance $\text{cov}(\widehat{\beta}_0, \widehat{\beta}_1)$ when the number of observations P is large, which change approximately by the factor $\sqrt{1-\rho}$ and $1-\rho$, respectively, compared to a correlation coefficient of $\rho = 0$.

For a small number of observations, table 2 shows the influence of the correlation coefficient on the estimates (assuming the same, above correlation structure). In particular, the table lists for a subset of size $P = 20$ of the data in figure 1 the estimates and uncertainties for $\widehat{\beta}_0$, $\widehat{\beta}_1$ and for $\rho = 0$, $\rho = 0.8$. The reader observes, that compared to no correlation, the estimate for the slope changes by almost half of the uncertainty (i.e. $\widehat{\beta}_1^{\text{corr}} - \widehat{\beta}_1 \approx u(\widehat{\beta}_1)/2$) and at the same time the uncertainty reduces considerably.

Table 2: Results obtained by weighted total least-squares with uncertainty evaluation according to the GUM for a subset of size $P = 20$ of the data in figure 1. Listed are the estimates and uncertainties for slope and intercept.

Correlation a.u.	$\widehat{\beta}_0$ $\mathrm{g\,L^{-1}}$	$u(\widehat{\beta}_0)$ $\mathrm{g\,L^{-1}}$	$\widehat{\beta}_1$ a.u.	$u(\widehat{\beta}_1)$ a.u.
$\rho = 0.0$	0.153 3	0.458 8	0.994 5	0.004 2
$\rho = 0.8$	0.176 1	0.206 1	0.996 5	0.002 4