Letter to the Editor

Marith van Schrojenstein Lantman* and Marc H. M. Thelen

The impact of measurement uncertainty on the uncertainty of ordinal medical scores based on continuous quantitative laboratory results

<https://doi.org/10.1515/cclm-2020-1890>

Received December 31, 2020; accepted March 17, 2021;

published online March 25, 2021

Keywords: clinical-decision making; measurement uncertainty; medical algorithm.

To the Editor,

The publication of the ISO/TS 20914:2019 practical guidance on measurement uncertainty (MU) has triggered a discussion on the calculation of MU related to derived biological quantities [1–4]. These papers postulate that the combined uncertainty of a derived biological quantity can be estimated by combining the individual measurand MU's [1–5]. To obtain an accurate MU, it is important to correct for correlations between the measurands [1, 5]. These correlations may cause interactions between the various measurands that can either increase or decrease the total combined uncertainty [1, 6].

However, neither the ISO20914:2019 nor the papers mentioned above suggest approaches for ordinal medical algorithms based on continuously expressed measurands with known uncertainty. Here we present how the uncertainty of continuous measurements impacts the uncertainty

of ordinal medical algorithms, and we provide guidance on the calculation in extension of ISO/TS 20914:2019 examples.

An ordinal medical algorithm is a score-based system where several continuously expressed measurands are translated into individual ordinal scores as a result of the application of absolute thresholds on continuous original results. The scores for the individual elements are then summed into an overall combined score, which is used to make medical decisions. When exceeding a predefined diagnosis or treatment limit, the patient is diagnosed or treated, respectively. As the individual continuous measurands have an uncertainty, so will the total end score of the medical algorithm.

However, the impact of MU of the measurands on the end score will differ per patient, as the proximity of the measurands to their respective cut-off scoring values is different. Therefore, the uncertainty approach for ordinal measurands (e.g., by Bashkansky et al. [7]) does not apply, since it does no right to the continuous base of the data where some values that are close to a threshold are more prone to uncertainty than those that are more remote from thresholds.

It is important to have knowledge of the score-uncertainty, as patients with additional risks or benefits could become apparent and currently used scoring systems could perhaps be improved.

In line with other calculations in ISO 20914:2019 we propose that the determination of the score-uncertainty is done by the following steps:

- (1) Calculate the relative expanded uncertainty %U(y) of the involved measurand y in the medical score in line with ISO20914:2019. We have used examples A3 and A4. To calculate %U(y), we used the following data:
  - u_{rw} : Calculate the uncertainty under intermediate precision conditions by using internal quality control data from routinely used commercial quality control materials of the past 2 years for every level used.
  - u_{cal} : Obtain the uncertainty of the calibrator(s) from the manufacturer. Note that manufacturers

*Corresponding author: Marith van Schrojenstein Lantman, MSc, Department of Laboratory Medicine, Radboud University Medical Centre, Geert Grooteplein Zuid 10, 6500HB, Nijmegen, The Netherlands; Laboratory for Clinical Chemistry, Part of Result Laboratorium, Amphia Hospital, Breda, The Netherlands; and SKML, Organisation for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands, E-mail: marith.vanschrojensteinlantman@radboudumc.nl

Marc H. M. Thelen, Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands; Laboratory for Clinical Chemistry, Part of Result Laboratorium, Amphia Hospital, Breda, The Netherlands; and SKML, Organisation for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands. <https://orcid.org/0000-0003-1771-669X>

may report the expanded uncertainty rather than the uncertainty.

- u_{bias} : Since none of the measurands involved had an unacceptable bias that needed immediate correction by the laboratory during the period involved, no uncertainty for bias correction u_{bias} needed to be applied.
  - k : to determine $\%U(y)$ we opted for a 95% confidence, which results in a coverage factor of 2.
- (2) Multiply individual patient results with the $\%U(y)$ of the most relevant concentration level to obtain half the width of the 95% interval of each measurand.
  - (3) Add (2) to and subtract (2) from the individual patient result to obtain the 2.5 and 97.5% percentile of outcome.
  - (4) Using the ordinal medical scoring algorithm on the patient results, 2.5% result and 97.5% result to obtain the ordinal scores of each measurand involved.
  - (5) Compute the highest and lowest obtainable medical algorithm score by using the lowest and highest sub scores as computed in (4).
  - (6) Compare the obtained scores at (5) with the medical decision-limit, as well as the original score of the patient.

A graphical representation of this theory is shown in Figure 1. The examples make clear that the same uncertainty in the continuous results can result in a different score uncertainty between patients based on the closeness of the individual measurands to scoring thresholds. Therefore, the ordinal medical scoring uncertainty is not only a magnitude that depends on the measurement uncertainty of the underlying measurements, but also on the distribution of the measurement results over the measurement range.

To demonstrate the impact of MU on scoring uncertainty we have applied our approach to two commonly used medical scoring algorithms using real patient data from our hospital with the application of MU data of the laboratory tests involved from our laboratory.

The Child-Pugh score is used to decide whether to operate on liver patients, and provides a mortality prognosis for patients [8]. The difference in Child-Pugh classification A, B and C is the mortality associated with surgery, respectively 10, 30 and 70–80%. The overall 1-year survival differs between Child-Pugh A, B and C patients as well: 100, 80 and 45%, respectively. Because of these major mortality differences and altered clinical decisions based on this classification, it is important to incorporate the MU in to the clinical decision limit. We included patients if values of albumin, bilirubin and INR were recorded in the laboratory information system (LIS). The used u_{rw} and, if

Figure 1: Figurative representation of effects of measurement uncertainty on a hypothetical score system of two patients. Patient 1 (yellow) has a continuous result that translates into a score of 3 on item 1 and a continuous result that translates into a score of 1 on item 2. The 95% interval (shown as horizontal bars) are fully within the region of these scores and therefore, inclusion of measurement uncertainty does not lead to a different score. However, for patient 2 identical uncertainties in the continuous results over span more than one scoring category for both item 1 and item 2. Inclusion of the measurement uncertainty does lead to a different score, as in item 1 the horizontal bar crosses the cut-off point between score 1 and 2, whereas in item 2 the horizontal bar crosses the cut-off point between score 3 and 4. Thus, in patient 2, the score is 5, with a 95% interval of 4–6.

available, u_{cal} are shown in Table 1. As the encephalopathy and ascites scores were not available, they were scored randomly and MU was not calculated for these factors. Of the 6,558 observations in the dataset, inclusion of the MU

Table 1: used data for the calculation of the impact of uncertainty on clinical outcome.

Measurand	Relative standard uncertainty of calibrator $\%u_{\text{cal}}$ (concentration level)	Relative standard uncertainty under intermediate precision conditions $\%u_{\text{rw}}$ (concentration level)	Relative total uncertainty of measurands $\%u(y)$
Albumin	0.853% (34.8 g/L)	3.78% (27.85 g/L) 2.67% (46.125 g/L)	3.88% 2.80%
Bilirubin	0.422% (74.1 $\mu\text{mol/L}$)	8.71% (6 $\mu\text{mol/L}$) 3.54% (116.8 $\mu\text{mol/L}$)	8.72% 3.57%
Haematocrit	0.902% (0.36 L/L)	3.51% (0.174 L/L) 1.92% (0.364 L/L)	3.62% 2.12%

Table 1: (continued)

Measurand	Relative standard uncertainty of calibrator % u_{cal} (concentration level)	Relative standard uncertainty under intermediate precision conditions % u_{rw} (concentration level)	Relative total uncertainty of measurands % $u(y)$
INR	NA	1.09% (1.005)	1.09%
		0.973% (2.005)	0.973%
Potassium	0.333% (3 mmol/L)	1.57% (2.63 mmol/L)	1.60%
	0.357% (7 mmol/L)	0.811% (7.42 mmol/L)	0.886%
Creatinine	0.429% (6,720 μ mol/L)	3.21% (77 μ mol/L)	3.24%
		2.04% (584 μ mol/L)	2.08%
Leukocytes	0.901% (7.1×10^9 /L)	3.11% (2.93×10^9 /L)	3.24%
		1.71% (6.83×10^9 /L)	1.93%
Sodium	0.208% (120 mmol/L)	0.734% (112.8 mmol/L)	0.763%
	0.25% (160 mmol/L)	0.663% (142.3 mmol/L)	0.709%
pH	NA	0.0888% (7.148)	0.089%
		0.107% (7.352)	0.107%
		0.134% (7.551)	0.134%
pO ₂	NA	3.95% (8.928 kPa)	3.95%
		2.65% (13.51 kPa)	2.65%
		1.87% (20.09 kPa)	1.87%

Per measurand, the relative standard uncertainty of the calibrator % u_{cal} and, if available, % u_{cal} of a second concentration is displayed. Also, the relative imprecision of the measurement procedure % u_{rw} at different concentration levels are shown. The relative total uncertainty % $u(y)$ of the measurands is shown in the last column that combines the information in the u_{cal} and u_{rw} column at particular concentration levels.

resulted in a potential change of Child-Pugh classification in 769 patients (11.72%).

The APACHE II score is an often-used score in many clinical decisions at intensive care units. The APACHE II score is based on age, Glasgow coma score (GCS), pO₂, temperature, mean arterial pressure, arterial pH, sodium, potassium, creatinine, haematocrit and white blood cell

count [9]. We included patients into the analysis if values of the APACHE II measurands were recorded in the laboratory information system (LIS) and assessed whether inclusion of the MU would result in a different clinical decision-limit (APACHE score > 13.5 [10]). The used u_{rw} and, if available, u_{cal} are shown in Table 1. The markers GCS and temperature were scored randomly and MU was not calculated for these factors. Of the 509 patients in the dataset, inclusion of the MU resulted in a potential different APACHE II medical decision in 41 patients (8.05%).

These data are in conjunction with recent publications and an editorial regarding effects of MU on calculated biological variables, and research is important to make further advances in clinical chemistry. Yet, no consideration has been given to medical algorithms, of which, due to their mathematical construct, a standardised uncertainty cannot be calculated. On a patient level, one can assess the individual score uncertainty, which can aid clinicians in their decision-making process. The examples of the Child–Pugh score and APACHE II score effects shown above indicate that the impact of MU on the clinical decision may be substantial, which warrants future efforts to conduct research on the uncertainty of other medical scoring systems.

Even when MU of the measurands involved is identical between institutions, the impact of the scoring uncertainty on the associated clinical decision may differ between institutions if their patients differ in their pre-test probability for the condition assessed by the algorithm. Therefore, every healthcare institution should calculate the effects of MU on their particulate patients to assess the possible effects.

This method of establishing the effects of MU on ordinal medical algorithm scores is based on the assumption that scores are summed. When scores are multiplied, attention has to be given to the effect of correlation, as described in earlier mentioned papers [1, 4–6].

Our approach to scoring uncertainty can promote the undertaking of new studies assessing whether the MU should be within, without or crossing the medical decision limit. In this way, the acceptable decision uncertainty can set the goal for the allowable MU for the measurands involved in its calculation. Due to the compensating effect of correlation between the different elements in a scoring algorithm, certain measurands might require stricter APS when used as standalone markers as opposed to when they're used in medical algorithms [2]. We hope this will give an impulse to new research into more APSs that are based on medical need as formulated in the first EFLM strategic meeting in Milan 2014 as the Milan 1a and 1b and will improve patient outcome and optimize healthcare that is already in place [11].

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.

References

1. Rigo-Bonnin R, Canalias F. Measurement uncertainty estimation for derived biological quantities. *Clin Chem Lab Med* 2021;59: e1–7.
2. Farrance I, Frenkel R, Badrick T. ISO/TS 20914:2019 – a critical commentary. *Clin Chem Lab Med* 2020;58:1182–90.
3. International Organization for Standardization. Medical laboratories – practical guidance for the estimation of measurement uncertainty. ISO/TS 20914:2019.
4. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. *Clin Chem Lab Med* 2020;58: 1407–13.
5. Farrance I, Badrick T, Frenkel R. Uncertainty in measurement: a review of the procedures for determining uncertainty in measurement and its use in deriving the biological variation of the estimated glomerular filtration rate. *Pract Lab Med* 2018;12:e00097.
6. Farrance I, Frenkel R. Measurement uncertainty and the importance of correlation. *Clin Chem Lab Med* 2021;59:7–9.
7. Bashkansky E, Gadrich T. Some metrological aspects of ordinal measurements. *Accred Qual Assur* 2010;15:331–6.
8. Tsois A, Marlar CA. Use of the child Pugh score in liver disease. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2020. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK542308/>.
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13: 818–29.
10. Fadaizadeh L, Tamadon R, Saeedfar K, Jamaati HR. Performance assessment of Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II in a referral respiratory intensive care unit in Iran. *Acta Anaesthesiol Taiwan* 2012;50: 59–62.
11. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med* 2015;53:833–5.