Testing the limits: the diagnostic accuracy of reference change values

Arne Åsberg, Ingrid Alsos Lian, Ingrid Hov Odsæter & Gustav Mikkelsen

Published online: 31 Mar 2021.

Submit your article to this journal ↗

Article views: 272

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS   Check for updates

Testing the limits: the diagnostic accuracy of reference change values

Arne Åsberg, Ingrid Alsos Lian, Ingrid Hov Odsæter and Gustav Mikkelsen

Department of Clinical Chemistry, Trondheim University Hospital, Trondheim, Norway

ABSTRACT

Reference change values (RCVs) are used by the physician to judge whether a change in analyte concentration from one sample to the next may represent a clinically significant change. Published RCVs are usually given as fixed percentages of the analyte concentration in the first sample. The accuracy of published RCVs is not well known. We obtained public-use data from the US National Health and Nutrition Examination Survey (NHANES) 2001–2002 to study the distribution of changes in the concentration of eight commonly used analytes. Specimens were obtained on two occasions 7–47 days apart from 279 to 411 individuals with an analyte concentration within the reference interval in both samples. The analytes were albumin, calcium, cholesterol, phosphate, potassium, sodium, hemoglobin and thrombocytes. For each analyte, normal within-subject biological coefficient of variation from the EFLM Working Group on Biological Variation and the NHANES analytical coefficient of variation were used to calculate the 5 and 95 percentile RCVs. These RCVs were calculated as fixed percentages of the analyte concentrations in the first sample and compared to the empirical 5 and 95 percentiles. The sensitivity of the RCVs in detecting changes outside the empirical percentiles ranged from 0.35 for sodium to 0.80 for albumin. The specificity of the RCVs in detecting changes inside the empirical percentiles ranged from 0.85 for potassium to 0.97 for thrombocytes. Calculating RCVs as fixed percentages of the analyte concentration in the first sample lessened the diagnostic accuracy. RCVs given as a function of the first result would perform better.

Introduction

According to Fraser [1], the term reference change value (RCV) was introduced by Harris and Yasaka [2]. RCVs may help the physician determine whether a change in analyte concentration from one sample to the next represents a clinically significant change. The idea is that changes between the limits of $\pm$ RCV may stem from random, normal biological and analytical variation and may not represent a change in the clinical condition. Traditionally, an RCV is calculated as $2^{0.5} \times Z \times (\mathrm{CV_A}^2 + \mathrm{CV_I}^2)^{0.5}$, where Z represents the number of standard deviations from the mean in the standard normal distribution, and $\mathrm{CV_A}$ and $\mathrm{CV_I}$ represent the ordinary analytical and normal within-subject biological coefficients of variation, respectively [1]. Each laboratory has to know its ordinary $\mathrm{CV_A}$. The magnitude of the normal $\mathrm{CV_I}$ is not so easily obtained and may have been wrongly estimated for several years [3]; however, new estimates have now been published [4–6]. According to Røraas et al. [3], the limits of RCV are not necessarily symmetrical, and Aarsand et al. [4] published asymmetrical RCVs for several commonly used analytes, the numerical values of percent decrease being lower than the numerical value of percent increase. RCVs were given as one pair of percentage for each analyte, for instance −8.8% and 9.6% for serum potassium [4], meaning that 'the reference interval of change' of

serum potassium from one day to some other day is from a decrease of 8.8% to an increase of 9.6%. Implicitly, this goes for any serum potassium starting value, and Fraser stated that a 'change' is the change from the first to the second sample [7]. Whether the percentage change in concentration from the first to the second sample really is independent of the concentration in the first sample was not mentioned by Fraser in his review article from 2011, nor was it mentioned in newer [4] or older literature [2]. However, an individual detected with an analyte concentration in the low end of the reference range on the first sampling day might tend to have a higher value on the next sampling day (and vice versa for those with the first value in the high end of the reference range). This would create a regression towards the mean and a correlation between the two measurements. If both measurements were high or both were low, they would still be correlated. Generally, if two measurements, say r_1 and r_2, are correlated and their relation can be expressed as $r_2 = a \times r_1 + b$, then the difference between r_2 and r_1 is: $r_2 - r_1 = a \times r_1 + b - r_1 = (a - 1) \times r_1 + b$, i.e. the change from r_1 to r_2, depends on r_1 if $a \neq 1$. Calculating the change as a percentage of r_1 may not help, because $100 \times [(a - 1) \times r_1 + b]/r_1 = 100 \times (a - 1 + b/r_1)$ is also dependent on r_1 if $b \neq 0$. In such cases, the limits of RCV calculated as a fixed percentage of the concentration in the first sample might

not equal the corresponding, real percentiles in the distribution of changes in the population and might impair the diagnostic accuracy of the RCVs in detecting changes outside these percentiles. We aimed to test this hypothesis.

Methods

Population

We used data from the U.S. Department of Health & Human Services, Centers for Disease Control and Prevention, the National Health and Nutrition Examination Survey (NHANES) 2001–2002 survey [8], which represented 'the total noninstitutionalized civilian population residing in the 50 states and District of Columbia' [9]. About 5% of the participants were recruited for a 'second day laboratory exam' [10]. In doing so, the NHANES staff aimed for an approximately uniform age distribution for participants aged 16–69 years, about an equal number of women and men, and an approximately equal number of non-Hispanic blacks and whites, and Mexican Americans. The blood samples could be taken at different times on the two days (no matching to morning, afternoon or evening). From the two days of sampling, we selected data on serum albumin, calcium, cholesterol, phosphate, potassium, sodium, and blood hemoglobin and thrombocytes, as these are commonly used to monitor health conditions. In addition, we selected data on serum C-reactive protein (CRP) to exclude individuals with inflammation. Gender, age and time between the two sampling days were also registered. We included individuals that were at least 18 years of age (at least 20 years in case of cholesterol) and with CRP values within 0–10 mg/L (the stated reference range for CRP) on both sampling days. Also, for each analyte, the concentration in the first and the second sample had to be within the normal reference range as given by NHANES in separate documents at [11], because we wanted to compare the real and expected (from RCVs) changes in clinically stable individuals.

Laboratory methods

According to NHANES, the laboratory methods for the second day examinations were the same as for the first day examinations [10]. Laboratory analyses were performed at Coulston Foundation using a Hitachi Model 704 analyzer (Chiyoda City, Japan) and at Collaborative Laboratory Service, using a Beckman Synchron LX20 analyzer (Fullerton, CA). Procedures were taken to assure equal mean analyte levels in the two laboratories [10]. For cholesterol, we used data from 'the reference analytic method', as recommended by NHANES. Hemoglobin and thrombocytes were analyzed on Beckman Coulter MAXM (Fullerton, CA), and CRP was measured with a nephelometric method on a BN2 instrument from Dade Behring (now Siemens, Munich, Germany). Details on the analytical methods are given in separate documents [11].

Statistical methods

We defined a 'change' in analyte concentration as the difference between concentrations in the second and the first sample [7], and calculated changes accordingly. Outliers in the distribution of changes were identified by the generalized extreme studentized deviate method of Rosner [12], using an alpha value of 0.01, and excluded. Next, we used quantile regression to model the empirical 5 and 95 percentiles in the distributions of changes as functions of the analyte concentration in the first sample, age, gender and the time (in days) between the two samplings. The statistical significance of these four variables was tested for two percentiles in eight analytes, comprising $4 \times 2 \times 8 = 64$ significance tests, so for these analyses we defined p values less than $.05/64 = 0.0008$ as statistically significant. From quantile regression models of the 5 and 95 percentiles as functions of the concentration in the first sample only, we plotted the percentiles and their 95% confidence areas, along with the individual observations, against the concentration in the first sample. In these plots, we added lines corresponding to the RCVs calculated from CV_I published by the EFLM Working Group on Biological Variation [4–6] and CV_A published by NHANES [13, 14]. RCVs in percent were calculated as $100\% \times (\exp(\pm Z \times 2^{0.5} \times (CV_{lnA}^2 + CV_{lnI}^2)^{0.5}) - 1)$ [15], where $Z = 1.65$ [4], $CV_{ln} = (\ln(1 + CV^2))^{0.5}$ [3], and where CV is a fraction. These RCV lines correspond to the 5 and 95 percentiles in the distribution of changes. Sensitivity of the RCVs in detecting changes outside the empirical percentiles was calculated as the fraction of changes outside the empirical 5 or 95 percentiles that also was outside the corresponding RCVs. Specificity was calculated as the fraction of changes within the range of the empirical 5 to the 95 percentiles that also was inside the corresponding RCVs. The area under the receiver operating characteristic (ROC) curve was calculated as (sensitivity + specificity)/2 [16]. Spearman's rank correlation coefficient was used to study the correlation between the first and second analytical result, and the correlation between the change in analyte concentration and change in CRP.

The Stata software, version 16 (StataCorp, College Station, TX) was used for all statistical analyses, except for outlier detection, where we used MedCalc, version 19 (MedCalc Software, Ostend, Belgium).

Results

The total number of individuals at least 18 years of age with analytical results from two separate days are shown in Table 1. For cholesterol we had to select data from individuals at least 20 years of age, because NHANES did not state any reference limits for younger individuals. Between 17% and 40% of the individuals were excluded due to a CRP result above 10 mg/L and/or because at least one result was outside the reference limits for the analyte (Table 1). After excluding those, another four individuals were excluded because the difference between the two analytical results was an outlier. The age ranged from 18 to 69 (20–69 for cholesterol) for both women and men. Median ages are given in

Table 1. For each analyte, the total number of individuals with two analytical results and age at least 18 years (at least 20 for cholesterol) are given, along with the numbers the exclusion criteria CRP above 10 mg/L or missing, and the concentration of at least one analytical result being outside the reference range. Next, the numbers of outliers are given, then the total number of included individuals with their fraction of women and the median age of women and men.

Analyte	Total number of individuals	CRP high or missing	Outside reference range	Outlier	Number of included individuals	Fraction of women	Median age (years) Women	Men
Albumin	496	73	26	1	402	0.48	37	43
Calcium	496	73	17	0	411	0.46	37	41
Cholesterol	440	67	128	0	262	0.42	43	46
Phosphate	496	73	103	0	335	0.44	37	44
Potassium	496	73	33	0	396	0.45	35	40
Sodium	496	73	79	0	366	0.44	41	41
Hemoglobin	508	84	40	0	393	0.47	35	41
Thrombocytes	508	84	26	3	401	0.46	40	42

Table 2. Results of the quantile regression analyses showing the empirical 5 and 95 percentiles (95% confidence intervals) as functions of the first analytical result.

Analyte	5 percentile Regression coefficient	Constant	95 percentile Regression coefficient	Constant
Albumin	−0.17 (−0.35 to 0.02)	3.5 (−4.5 to 12)	−0.40 (−0.56 to −0.24)	20 (13–26)
Calcium	−0.50 (−0.66 to −0.34)	1.05 (0.66–1.44)	−0.40 (−0.52 to −0.28)	1.05 (0.76–1.34)
Cholesterol	−0.11 (−0.34 to 0.11))	−0.21 (−1.31 to 0.90)	−0.29 (−0.41 to −0.16)	1.94 (1.32–2.55)
Phosphate	−0.46 (−0.70 to −0.22)	0.36 (0.08–0.64)	−0.68 (−0.81 to −0.56)	1.00 (0.85–1.15)
Potassium	−0.75 (−0.91 to −0.59)	2.6 (2.0–3.2)	−0.30 (−0.46 to −0.14)	1.7 (1.0–2.3)
Sodium	−0.75 (−0.89 to −0.61)	102 (82–121)	−0.80 (−1.01 to −0.59)	114 (85–144)
Hemoglobin	−0.13 (−0.25 to −0.01)	0.71 (−0.99 to 2.42)	−0.14 (−0.24 to −0.04)	2.63 (1.21–4.05)
Thrombocytes	−0.32 (−0.48 to −0.16)	40 (−4 to 84)	−0.03 (−0.17 to 0.10)	63 (25–100)

Table 1. The number of days between the first and second sampling ranged between 7 and 47, median 17–18 days for all analytes. The correlation between the analyte concentration in the second and first sample, as measured with the Spearman rank correlation coefficient, varied from 0.35 for sodium to 0.92 for hemoglobin. The p values of correlation coefficients were less than .0001 for all analytes. The median difference between the analyte concentration in the second and first sample was zero, or the 95% confidence interval of the median included zero, for all analytes except hemoglobin, where the median difference was −0.2 g/dL (95% confidence interval −0.3 to −0.2 g/dL), and albumin, where the median difference was −1.0 g/L (95% confidence interval −1.0 to −0.9 g/L). The change in analyte concentration did not significantly correlate with the change in CRP, except for the change in albumin, where the Spearman rank correlation coefficient was −0.129 ($p = .010$).

In the quantile regression models of the 5 and 95 percentiles, the coefficient of the analyte concentration in the first sample was statistically significant ($p < .0008$) for at least one of the percentiles for all analytes. Age, gender and time between the days of sampling were not statistically significant for any analyte percentile. The parameters of the 5 and 95 percentiles estimated from models with only the analyte concentration in the first sample as predictor are given in Table 2. The percentiles are illustrated in Figure 1, along with the limits of RCVs calculated as fixed percentages of the analyte concentration in the first sample. The percentage values of RCVs are given in Table 3, and their accuracies in finding changes outside or inside the empirical 5 and 95 percentile are given in Table 4.

Discussion

For eight commonly used analytes, we have shown that changes in the concentration from the first to the second sample obtained 7–47 days apart depended on the concentration in the first sample. This may be due to the expected mathematical association, but also to a biological regression towards the mean. The variables age, gender and time between sampling did not reach statistical significance for any analyte. Most importantly, these results show that the 5 and 95 percentiles in distribution of changes did not depend on the time lap between samplings, at least over a time span of 7–47 days. Time dependency would severely complicate the clinical implementation of RCVs.

However, when we calculated limits of RCV as fixed percentages of the concentration in the first sample, as we are taught to do [7], those limits did not show the same dependency on the concentration in the first sample as the empirical 5 and 95 percentiles (Figure 1). We quantitated this discrepancy by calculating the accuracy of the RCV limits in detecting the same 'abnormal' changes as the empirical 5 and 95 percentiles. If these were real diagnostic tests, we would not be very impressed by the performance of the sodium RCVs with a ROC area of 0.65, while the ROC areas were in the range of 0.81–0.87 for the RCVs of the other analytes (Table 4). The best performance was shown by the RCVs of thrombocytes, with a ROC area of 0.87. Comparing the sodium and thrombocytes graphics in Figure 1, one difference is the steeper slope of the sodium percentile lines compared to those of the thrombocyte percentiles. Presumably sodium has a stronger homeostatic regulation than thrombocytes: a low concentration in the first sample
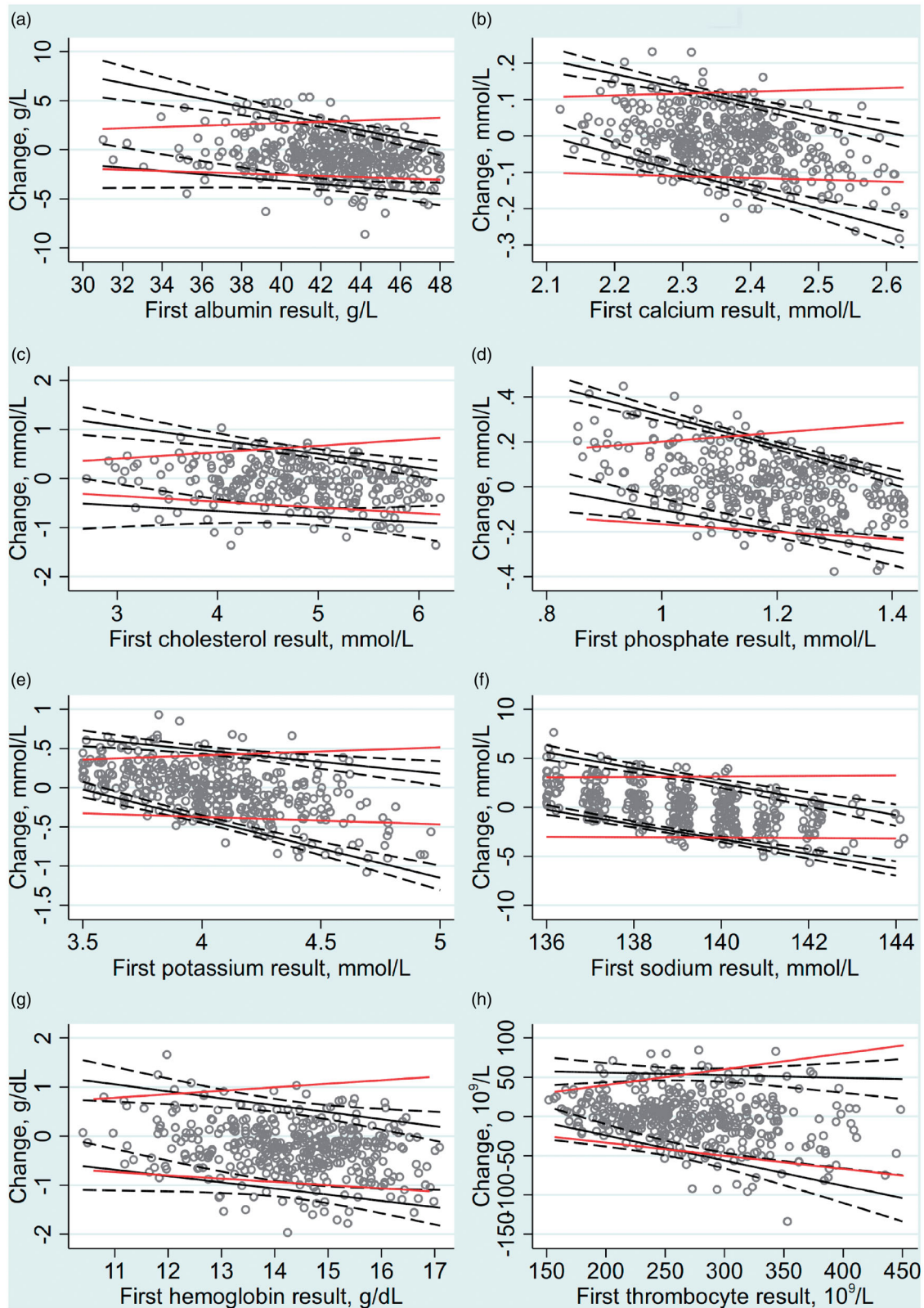
Figure 1. Individual changes in analyte concentration from the first to the second sample plotted against the concentration in the first sample for albumin (a), calcium (b), cholesterol (c), phosphate (d), potassium (e), sodium (f), hemoglobin (g) and thrombocytes (h). In order to reduce the problem of overlapping points, 2% random noise was added. Also shown are the empirical 5 and 95 percentiles with their 95% confidence areas, in addition to the limits of reference change values (red lines).

(r_1) is likely to be followed by a higher concentration in the second sample (r_2) and thus the difference $r_2 - r_1$ is higher at low concentrations, and lower at high concentrations.

Another obvious finding is that the distances between the two percentile lines and between the two RCV lines are not very different for sodium. If the RCV lines were tilted as the

Table 3. For each analyte, the analytical coefficient of variation (CV$_A$) from NHANES [13,14] and normal within-subject biological coefficients of variation (CV$_I$) from the EFLM Working Group on Biological Variation [4–6] are listed, in addition to the limits of low (5%) and high (95%) reference change values (RCV). All figures are percentages.

Analyte	CV$_A$	CV$_I$	RCV Low limit	RCV High limit
Albumin	1.3	2.50	−6.36	6.80
Calcium	1.1	1.81	−4.82	5.07
Cholesterol	1.5	5.18	−11.8	13.4
Phosphate	1.7	7.67	−16.7	20.1
Potassium	1.5	3.92	−9.33	10.3
Sodium	0.8	0.53	−2.21	2.26
Hemoglobin	1.1	2.74	−6.66	7.13
Thrombocytes	3.1	7.22	−16.7	20.1

Table 4. The diagnostic accuracy of the 5% and 95% RCVs in detecting changes from the first to the second sample outside or inside the 5 and the 95 percentiles in the distributions of observed changes. Point estimates with 95% confidence intervals in parenthesis are given.

Analyte	Sensitivity	Specificity	Area under the ROC curve
Albumin	0.80 (0.63–0.92)	0.88 (0.84–0.91)	0.84 (0.77–0.91)
Calcium	0.71 (0.54–0.85)	0.92 (0.89–0.95)	0.82 (0.74–0.89)
Cholesterol	0.75 (0.53–0.90)	0.90 (0.85–0.93)	0.82 (0.73–0.91)
Phosphate	0.71 (0.51–0.87)	0.92 (0.89–0.95)	0.82 (0.73–0.91)
Potassium	0.79 (0.61–0–91)	0.85 (0.81–0.88)	0.82 (0.75–0.89)
Sodium	0.35 (0.22–0.51)	0.95 (0.92–0.97)	0.65 (0.58–0.72)
Hemoglobin	0.66 (0.49–0.80)	0.96 (0.93–0.97)	0.81 (0.73–0.88)
Thrombocytes	0.76 (0.60–0.89)	0.97 (0.94–0.98)	0.87 (0.80–0.93)

percentile lines, the diagnostic accuracy would improve. Constructing the RCV lines slightly asymmetrical about zero as done in Figure 1, in accordance with Røraas et al. [3], did not help. In fact, the regression coefficients of the empirical 5 and 95 percentiles showed a negative slope for all analytes (Table 2), more pronounced for those analytes considered to be more strictly regulated, i.e. sodium, potassium, calcium and phosphate.

We used the ROC area as a summary measure of diagnostic accuracy. Sensitivity and specificity could have been combined into other measures, as Youden's index or the diagnostic odds ratio (DOR) [17]. Sensitivity and specificity are equally weighted in all these measures, which may not be optimal; however, we had no data for a different weighing of sensitivity and specificity. We think the ROC area is more easily interpreted than Youden's index and DOR.

A potential weakness of this work is our lack of knowledge of possible *changes* in the health condition of the participants. If a change in their clinical condition occurred between the two sampling days, the 5 and 95 percentiles constructed in Figure 1 would not represent the percentiles in a clinically stable population and the diagnostic accuracy of the RCVs might be wrongly estimated. However, the participants represented a noninstitutionalized population, where the second day examinations were conducted for quality control and for research purposes [10]. NHANES states that measures of intra-individual variation can be evaluated from these data [10], and so have been done [13, 14]. To secure including data from only clinically stable individuals, we excluded individuals with CRP above 10 mg/L in the first or the second sample, or with analyte

concentration outside the reference limits in the first or the second sample. Except for albumin, the change in analyte concentration did not correlate with the change in CRP, indicating that any small inflammation had nothing to do with the change in analyte concentration. For albumin, the correlation was weak. Furthermore, except for albumin and thrombocytes, no outliers were detected in the distribution of changes.

We limited the study to eight commonly used analytes. Whether similar findings are to be found with other analytes, remains to be seen. However, we expect that for strictly regulated analytes the concentration change will depend on the concentration in the first sample.

Another possible weakness might be any systematic differences between the analyte concentrations in the first and second sample due to a shift or drift of the analytical methods. NHANES did not mention whether the primary and secondary samples were analyzed in the same analytical run; however, the median difference was not significantly different from zero for six of the eight analytes. For hemoglobin, the median difference was −0.2 g/dL, and for albumin −1 g/L, indicating just small analytical shifts.

Taken together, we think the distribution of changes in the concentration from one sample to the next as selected from this NHANES population may be representative of changes seen in ambulant patients of stable health.

Conclusions

The diagnostic accuracy of published RCVs was less than optimal, mostly due to calculating RCVs as fixed percentages of the analyte concentration in the first sample. RCVs given as a function of the first result would perform better.

Disclosure statement

References

[1] Fraser CG. Reference change values. Clin Chem Lab Med. 2011;50(5):807–812.

[2] Harris EK, Yasaka T. On the calculation of a "reference change" for comparing two consecutive measurements. Clin Chem. 1983;29(1):25–30.

[3] Røraas T, Støve B, Petersen PH, et al. Biological variation: the effect of different distributions on estimated within-person variation and reference change values. Clin Chem. 2016;62(5): 725–736.

[4] Aarsand AK, Díaz-Garzón J, Fernandez-Calle P, et al. The EuBIVAS: within- and between-subject biological variation data for electrolytes, lipids, urea, uric acid, total protein, total bilirubin, direct bilirubin, and glucose. Clin Chem. 2018;64(9): 1380–1393.

[5] Carobene A, Aarsand AK, Guerra E, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation data for 15 frequently measured proteins. Clin Chem. 2019;65(8):1031–1041.

[6] Coşkun A, Carobene A, Kilercik M, et al. Within-subject and between-subject biological variation estimates of 21 hematological parameters in 30 healthy subjects. Clin Chem Lab Med. 2018;56(8):1309–1318.

[7] Fraser CG. Biological variation: from principles to practice. Washington (DC): AACC Press; 2001.

[8] NHANES 2001–2002 laboratory data; [Internet]. National Health and Nutrition Examination Survey; 2020; [cited 2020 Jun 15]. Available from: https://wwwn.cdc.gov/nchs/nhanes/Search/DataPage.aspx?Component = Laboratory&CycleBegin Year = 2001

[9] Curtin LR, Mohadjer LK, Dohrmann SM, et al. The National Health and Nutrition Examination Survey: sample design, 1999–2006. Vital Health Stat 2. 2012;155:1–39.

[10] NHANES 2001–2002 second day laboratory exam; 2020; [Internet]. Available from: https://www.cdc.gov/nchs/data/nhanes/nhanes_01_02/seconddaylabdoc_b.pdf

[11] NHANES 2001–2002 laboratory methods; [Internet]. National Health and Nutrition Examination Survey; 2020; [cited 2020 Jun 25]. Available from: https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/LabMethods.aspx?BeginYear = 2001

[12] Rosner B. Percentage points for a generalized ESD many-outlier procedure. Technometrics. 1983;25(2):165–172.

[13] Lacher DA, Hughes JP, Carroll MD. Biological variation of laboratory analytes based on the 1999–2002 National Health and Nutrition Examination Survey. Natl Health Stat Rep. 2010; 21:1–8.

[14] Lacher DA, Barletta J, Hughes JP. Biological variation of Hematology Tests based on the 1999–2002 National Health and Nutrition Examination Survey. Natl Health Stat Rep. 2012;54: 1–12.

[15] Carobene A, Marino I, Coşkun A, et al. The EuBIVAS Project: within- and between-subject biological variation data for serum creatinine using enzymatic and alkaline picrate methods and implications for monitoring. Clin Chem. 2017;63(9):1527–1536.

[16] Newman TB, Kohn MA. Evidence-based diagnosis: an introduction to clinical epidemiology. 2nd ed. Cambridge: Cambridge University Press; 2020.

[17] Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003; 56(11):1129–1135.