Rui Zhen Tan, Corey Markus, Samuel Vasikaran and Tze Ping Loh* for the APFCB Harmonization of Reference Intervals Working Group

Comparison of four indirect (data mining) approaches to derive within-subject biological variation

<https://doi.org/10.1515/cclm-2021-0442>

Received April 14, 2021; accepted January 21, 2022;

published online February 3, 2022

Abstract

Objectives: Within-subject biological variation (CV_i) is a fundamental aspect of laboratory medicine, from interpretation of serial results, partitioning of reference intervals and setting analytical performance specifications. Four indirect (data mining) approaches in determination of CV_i were directly compared.

Methods: Paired serial laboratory results for 5,000 patients was simulated using four parameters, d the percentage difference in the means between the pathological and non-pathological populations, CV_i the within-subject coefficient of variation for non-pathological values, f the fraction of pathological values, and e the relative increase in CV_i of the pathological distribution. These parameters resulted in a total of 128 permutations. Performance of the Expected Mean Squares method (EMS), the median method, a result ratio method with Tukey's outlier exclusion method and a modified result ratio method with Tukey's outlier exclusion were compared.

Results: Within the 128 permutations examined in this study, the EMS method performed the best with 101/128 permutations falling within ± 0.20 fractional error of the

'true' simulated CV_i , followed by the result ratio method with Tukey's exclusion method for 78/128 permutations. The median method grossly under-estimated the CV_i . The modified result ratio with Tukey's rule performed best overall with 114/128 permutations within allowable error.

Conclusions: This simulation study demonstrates that with careful selection of the statistical approach the influence of outliers from pathological populations can be minimised, and it is possible to recover CV_i values close to the 'true' underlying non-pathological population. This finding provides further evidence for use of routine laboratory databases in derivation of biological variation components.

Keywords: biological variation; data mining; indirect approach; outlier.

Introduction

Within-subject biological variation (CV_i) describes the day-to-day fluctuation of a biological parameter around a physiological set-point within an individual. In laboratory medicine, biological variation components are used in various ways, interpretation of the significance of a change in serial laboratory results, in determining the need for partitioning of reference intervals, standardised sample collection procedures and setting analytical performance specifications [1]. Traditionally, CV_i is determined by the direct approach involving repeated sampling of a group of subjects over time. Many studies have been published using the direct approach in determining CV_i , however, differing study designs and inclusion of a small number of subjects often limits the broader application of these findings. The Biological Variation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) was formed to develop and curate existing published biological variation data (<https://biologicalvariation.eu/>) and has developed a critical appraisal checklist for this purpose [2]. Direct biological variation studies can be operationally and financially demanding and a recent flagship of

***Corresponding author: Tze Ping Loh**, Department of Laboratory Medicine, National University Hospital, 5 Lower Kent Ridge Road, 119074, Singapore, Singapore, Phone: (+65) 67724345, Fax: (+65) 67771613, E-mail: tploh@hotmail.com

Rui Zhen Tan, Engineering Cluster, Singapore Institute of Technology, Singapore, Singapore, E-mail: RuiZhen.Tan@singaporetech.edu.sg

Corey Markus, Flinders University International Centre for Point-of-Care Testing, Flinders Health and Medical Research Institute, Flinders University, Rundle Mall, South Australia, Australia, E-mail: corey.markus@flinders.edu.au. <https://orcid.org/0000-0002-5594-9737>

Samuel Vasikaran, Department of Clinical Biochemistry, PathWest-Royal Perth Hospital, Perth, Western Australia, Australia, E-mail: Samuel.Vasikaran@health.wa.gov.au

such studies is the European Biological Variation Study (EuBIVAS), which has generated much high-quality biological variation data [3–13].

More recently, the use of routine longitudinal results available in laboratory information system databases has been exploited as a potential source for deriving reference intervals [14] and estimation of biological variation components. This approach requires careful application of clinical and statistical considerations to capture data from non-pathological populations before further statistical modelling to estimate the biological variation component [14–19]. Besides the reduced costs involved in performing such indirect/data mining approaches by leveraging laboratory data already obtained from routine clinical investigations, a greater number of subjects can be included overcoming the limitation seen in some direct studies. The indirect approach also facilitates estimation of biological variation components in special populations such as children [15–17]. Additionally, the indirect approach can also be applied to measurands that may be difficult to conduct using direct studies, such as therapeutic agents [18, 19].

In this simulation study, we systematically compare the performance of four indirect approaches for derivation of CV_i , three of which have been previously published and proposed a modification for one of the previously describe approaches. All approaches are compared under the same set of differing mixed populations consisting of simulated pathological and non-pathological values.

Methods

Statistical techniques

Three univariate techniques and one modified technique for indirect determination of CV_i were investigated. A brief overview of these approaches is presented below.

1) Expected Mean Squares method (EMS)

In this method, the mean of the sum of squares is used to estimate the expected mean squares within-patient, EMS_i , to yield and estimate for the within-patient standard deviation, SD_i , as follows:

$$EMS_i = SD_A^2 + SD_i^2 \quad (1)$$

EMS_i is obtained as follows:

$$EMS_i = \frac{1}{N} \sum_{j=1}^N \frac{(x_{ij} - \bar{x}_i)^2}{\bar{x}^2} \quad (2)$$

where N is the total number of patients, x_{ij} is the j -th measurement for patient i , $\bar{x}_i$ is the mean of the measurements for patient i , and $\bar{x}$ is the mean of all the measurements.

Previously, this method was applied on a dataset from non-pathological patients only to study the effect of number of replicates, samples, and individuals in determination of CV_i [20]. Applications to mixtures of non-pathological and pathological results requires exclusion of pathological values prior to the calculation of within-patient EMS [15]. To exclude patients with pathological phenotypes, Tukey's method for outlier detection is first applied on all the data (i.e. non-pathological and pathological) to exclude values below the lower boundary of the first quartile minus 1.5 times the interquartile range and values greater than the upper boundary of the third quartile plus 1.5 times the interquartile range. Patients with two remaining consecutive measurements after this exclusion process were used for calculation of the within-patient EMS.

2) Median method

The median method, first proposed by Loh et al. is similar to the EMS method above, except that the median, rather than the mean, is used for the estimation of EMS_i . The rationale for using the median over mean, is to reduce the influence of outliers and producing robust estimates [17].

3) Result ratio method with Tukey exclusion

In this approach, the second result is divided by the first result for each patient to yield a ratio [19]. In the original paper, a frequency distribution of the results ratio is then subjected to Bhattacharya's analysis for outlier exclusion. However, the manual application of Bhattacharya's method on the 12,800 simulated data sets is unfeasible. As an alternative, Tukey's method for outlier exclusion is applied. Use of a common outlier exclusion technique facilitates an equitable comparison across all indirect CV_i methods. Moreover, a ratio distribution (also known as a quotient distribution) is a probability distribution obtained from the ratio of random variables having two other known distributions. In particular, the ratio of two normal distribution follows a Cauchy distribution, rather than a normal distribution [21]. As such, the application of any outlier exclusion method on the ratio distribution is to remove data that are very unlikely to belong to the set due to its extremely low or high values, rather than identifying a central subset of Gaussian distributed data, as in the case of Bhattacharya's method. Tukey serves this purpose well. Since a different outlier exclusion technique has been applied, this method may not reflect the actual performance of the original method using Bhattacharya's method.

The coefficient of variation for the remaining data, CV_r , is calculated and from this the CV_i is obtained as follows:

$$CV_i^2 = \frac{CV_r^2}{\sqrt{2}} - CV_a^2 \quad (3)$$

4) Modified result ratio method

Here, a modification of the result ratio method [19] with more stringent outlier exclusion is examined. In the original paper, outlier exclusion is only applied once on the ratio that is obtained by dividing the second result by the first result of each patient.

In this method, Tukey's rule is applied three times for outlier exclusion. First it is used at the population level on all the data. Next, like the original paper, outlier exclusion is applied at the individual level, on the derived ratios obtained by dividing the second result by the first, and subsequently finally, on the ratio obtained by dividing the first result by the second. The application of outlier exclusion on both derivations of ratios ensures that patients with pathological

measurements in either the first or second measurements are treated similarly. Finally, CV_{tot} was calculated from individuals with two consecutive values remaining to estimate the CV_i .

$$CV_i^2 = \frac{CV_{tot}^2}{\sqrt{2}} - CV_a^2 \quad (4)$$

Simulations

The simulation approach used in this study was modelled after that described by Røraas and colleagues [20]. The simulated data used in this study were drawn from three interrelated parameters: (1) the mean (physiological set point) of a patient from the population distribution, (2) the ‘true’ biological variation of a sample drawn from the intra-individual distribution of the patient, (3) the observed laboratory measurement drawn inclusive of the analytical imprecision that is used to calculate the CV_i . These simulation steps are represented graphically in Figure 1, and the statistical definitions are detailed below.

- (1) The individual patient’s mean, μ_i , is first determined by sampling from the non-pathological population distribution defined as,

$$\mu_i \sim N(\mu_g, CV_g^2 \mu_g^2) \quad (5)$$

where μ_g and CV_g are the population’s mean and between-person coefficient of variance respectively, with μ_g and CV_g fixed at 100 and 0.2 for all simulations.

- (2) For measurements from the non-pathological distribution, the ‘true’ biological value of the patient sample, μ_{ij} , is determined by sampling from a normal distribution centred at the mean of the patient as follows:

$$\mu_{ij} \sim N(\mu_i, CV_i^2 \mu_i^2) \quad (6)$$

On the other hand, for measurements from the pathological distribution, the ‘true’ biological value of the patient sample, μ_{ij} , is determined by sampling from a normal distribution centered at the mean of the patient translated by $d\mu_g$ to account for the shift in the mean of the patient when in a pathological state, with a CV_i that is $1 + e$ times the non-pathological state CV_i .

Hence the pathological state is defined as:

$$\mu_{ij} \sim N(\mu_i + d\mu_g, CV_i^2 (1 + e)^2 (\mu_i + d\mu_g)^2) \quad (7)$$

For each of the samples, the observed laboratory results is simulated after incorporating the analytical variation (CV_a). The paired observed laboratory results for each patient is then subjected to statistical analysis to derive the within-subject biological variation (CV_i)

Figure 1: Illustrated simulation steps for the consecutive measurement results used in this study.

In the simulation, two results representing four different clinical scenarios were sampled, for a total of 5,000 simulated patients. The samples collected followed the within-subject biological variation of the simulated subjects. Only one laboratory measurement was made on each sample, which incorporated the analytical variation.

- (3) Finally, the observed laboratory measurement obtained for the patient sample μ_{ijm} , is determined by sampling from a normal distribution centred at the 'true' biological value,

$$\mu_{ijm} \sim N(\mu_{ij}, CV_a^2 \mu_{ij}^2) \quad (8)$$

where CV_a represents the analytical coefficient of variance. The number of replicate measurements, m is set as 1 since patient samples are typically measured only once in routine laboratory testing.

To compare the performance of the four indirect methods in determining the CV_i , simulations of mixed populations from representative pathological and non-pathological distributions were performed. In each simulation, two random serial measurements were simulated for 5,000 simulated patients. Simulations were characterized by four parameters, d the percentage difference in the mean between the non-pathological and pathological populations, CV_i the within-subject coefficient of variation of the non-pathological distribution, f the fraction of pathological values, and e the relative increase in CV_i of the pathological distribution. Additionally, control experiments with a single non-pathological distribution whereby d , e and f were set as 0, were also examined. The effects of the simulation parameters on the data distribution are illustrated in Supplementary Figure 1.

For the $1 - f$ fraction of patients (which represents the non-pathological population), both serial measurements were drawn from the non-pathological distribution. For the remaining fraction f which represents the pathological population, three separate clinical scenarios of consecutive measurements were considered: (1) progression of a patient from non-pathological to a pathological state (either acute or chronic); (2) patients who recovered from a pathological state; (3) patients who remain in a pathological state. Accordingly, it was arbitrarily determined that of the pathological fraction f , 40% of the patients have their first measurement obtained from the non-pathological distribution and the second measurement obtained from the pathological distribution (scenario 1); 40% of the patients have their first measurement obtained from the pathological distribution and second measurement obtained from the non-pathological distribution (scenario 2); and the remaining 20% of the patients have both their measurements obtained from the pathological distribution (scenario 3).

A total of 128 conditions were simulated corresponding to all permutations of $d=0.05, 0.1, 0.2, 0.4$, $CV_i=0.05, 0.1, 0.2, 0.4$, $f=0.05, 0.1, 0.2, 0.4$ and $e=0.5$. The selected CV_i parameters approximated the 25th, 50th, 75th and 95th percentile values of the EFLM biological variation database (<https://biologicalvariation.eu/>). For each condition, 100 rounds of simulations were performed. These simulated consecutive patient results were then subjected to statistical analysis to derive the CV_i .

Implementation of the simulation

Customized Python code was used for generation of all data sets in the 128 conditions (provided as Supplementary material). For each condition, 100 rounds of simulation were performed. The average of the CV_i obtained from the 100 iterations was reported for each condition. Implementation of the EMS method, median method and the result ratio methods were performed entirely within the Python environment.

Evaluation of the performance of the techniques

To compare the performance of different techniques, the fractional error of the derived CV_i is determined as below.

$$error = \frac{CV_{derived} - CV_{actual}}{CV_{actual}}$$

where $CV_{derived}$ and CV_{actual} are the derived and actual CV_i respectively.

The performances of the different methods are assessed at three different levels of fractional errors, ± 0.1 , ± 0.2 and ± 0.4 , or 10, 20 and 40% respectively on each side of the true target CV_i .

Results

Simulations of conditions

To compare the performance of the different methods, 128 conditions varying the distances between the populations, the true underlying CV_i , fraction of pathological values and the relative increase in CV_i of the pathological distribution were examined in 100 rounds of simulation at each combination, respectively.

A high CV_i may lead to the production of negative values in the simulation due to the large spread of the distribution. The negative value is converted into zero for the purpose of this study. The fraction of zero values obtained for each condition are shown in Supplementary Table 1, with the fraction of zero values was negligible for $CV_i < 0.4$ and remained below 0.02 or 2% for CV_i at 0.4. This suggests censorship of negative values by conversion to zero in our analysis would not have significantly impact our findings.

Next, the fractions of non-pathological patients and pathological patients included in the determination of CV_i were averaged over the 100 rounds of simulations and determined for each of the exclusion methods. As shown in Supplementary Table 2, both the EMS and medians approaches applied outlier exclusion at the population level and the proportion of patients included in the two approaches were identical, with pathological patient results in scenario 3 excluded more often than in scenarios 1 and 2. The result ratio method with Tukey's rule applied pathological result exclusion at the individual level on the ratio obtained when individual's second result is divided by the first and performed considerably better than the EMS and median methods for patient results in scenario 1.

This likely resulted from the greater difference between the two measurements at the individual level since they were drawn from the two different distributions.

Table 1: Number of conditions in where the average CV_i lies within ± 0.1 , ± 0.2 and ± 0.4 error from the actual value for each method.

	<i>e</i>				<i>CVi</i>				<i>f</i>				<i>d</i>			
Error = 0.1	Total	0	0.5	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	
EMS	79	46	33	9	16	25	29	26	24	18	11	28	23	17	11	
Median	6	3	3	4	2	0	0	0	0	0	6	0	0	2	4	
Ratio	104	55	49	20	24	29	31	32	32	24	16	31	28	24	21	
Modified	115	58	57	24	28	31	32	32	32	30	21	32	30	26	27	
Error = 0.2	Total	0	0.5	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	
EMS	99	51	48	14	23	30	32	30	26	25	18	31	28	24	16	
Median	6	3	3	4	2	0	0	0	0	0	6	0	0	2	4	
Ratio	111	56	55	22	26	31	32	32	32	26	21	32	30	26	23	
Modified	121	61	60	27	30	32	32	32	32	32	25	32	31	30	28	
Error = 0.4	Total	0	0.5	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	0.05	0.1	0.2	0.4	
EMS	111	56	55	19	28	32	32	30	30	26	25	32	31	28	20	
Median	107	53	54	11	32	32	32	24	24	28	31	25	26	28	28	
Ratio	122	61	61	28	30	32	32	32	32	32	26	32	32	30	28	
Modified	123	61	62	28	31	32	32	32	32	32	27	32	32	30	29	

Number of conditions are further grouped by *e*, CV_i , *d* and *f* values. There are in total 128 conditions, with 64 conditions for each value of *e* and 32 conditions for each value of the CV_i , *d* and *f*.

Furthermore, since the ratio is asymmetric, the exclusion of patients progressing from non-pathological state to pathological state (scenario 1) is easily more identified as an outlier than the exclusion of patients recovering from a pathological state (scenario 2). This was because the average value of the non-pathological state was lower in the simulations, hence the ratio would have a lower value for the denominator, making this approach more sensitive for outlier detection in scenario 1. Additionally, from Supplementary Table 2, the fraction of pathological patients included from scenarios 1 through to 3 for determination of CV_i was lowest for the modified result ratio method compared to any of the other approaches. Since the fraction of pathological patient results retained in scenarios 1 and 2 was similar for the modified result ratio method, suggests that this approach excludes pathological values at the individual level for patients progressing from pathological states to normal and conversely entering pathological states equally well. These findings suggest that estimation of CV_i incorporating more stringent strategies to reduce the fraction of pathological values, can produce accurately derived CV_i irrespective of other population parameters examined in this study.

On the other hand, application of outlier exclusion in the result ratio method was better at excluding patients with one measurement from the pathological state (scenario 1 and 2). This likely resulted from the greater difference between the two measurements at the individual level since they were drawn from two different distributions. Furthermore, since the ratio is asymmetric, the exclusion of patients progressing from non-pathological state to pathological state (scenario 1) is easily more identified as an outlier than

the exclusion of patients recovering from a pathological state (scenario 2). This was because the average value of the non-pathological state was lower in the simulations, hence the ratio would have a lower value for the denominator, making the ratio more sensitive for outlier detection in scenario 1.

Comparison of standard methods

The average CV_i obtained using the different methods for the 128 conditions is shown in Table 1. The average fractional differences for CV_i in all the conditions are displayed graphically in Figure 2 and numerically in Supplementary Table 3. First, we compared the three existing standard methods, with the result ratio approach demonstrating the best overall performance with 104, 111 and 122 conditions out of a total of 128 conditions, within ± 0.1 , ± 0.2 and ± 0.4 fractional error of the true underlying CV_i respectively. The EMS method was in second place with 79, 99 and 111 conditions fulfilling the three performance criteria, respectively. Finally, the median method has the worst performance with only six conditions within the ± 0.1 and ± 0.2 fractional errors and grossly under-estimate the CV_i . The median method was originally proposed to minimize the effects of outliers on within-patient coefficient of variations measured from individual patients, however, it tends to underestimate the CV_i . At larger fractional errors of ± 0.4 an acceptable performance of 107 conditions fulfilling the performance measure was observed.

From Table 1, the performance of each method was similar at different values of *e*, the relative increase in CV_i of

Figure 2: Results of average CV_i fractional error obtained from the different methods.

The fractional errors are firstly arranged by e (left $e=0$, right $e=0.5$). Within each block of e , CV_i is arranged by increasing order (from left to right $CV_i=0.05$, $CV_i=0.1$, $CV_i=0.2$ and $CV_i=0.4$). Shape and colour of the points represent f and d respectively. Circle, triangle, diamond and square are used to represent $f=0.05$, $f=0.1$, $f=0.2$ and $f=0.4$ respectively, whereas white, light grey, dark grey and black are used to represent $d=0.05$, $d=0.1$, $d=0.2$ and $d=0.4$ respectively. Solid grey line is 0.0, dotted grey line is ± 0.2 .

the pathological distribution as compared to the non-pathological distribution. For the EMS method, the average CV_i was closer the true underlying CV_i at higher values. At relatively high CV_i values of 0.2 and 0.4, a high fraction of conditions had mean estimates of CV_i within the ± 0.20 fractional error. However, at low CV_i values of 0.05, only 14 out of 32 conditions had estimates of CV_i within ± 0.20 fractional error. Performances of the EMS method degraded with increasing values of d and f . At high d and f values and low CV_i this performance degradation is

especially concerning, where the fractional error of the derived mean CV_i varied between 1 and 2 or alternatively 100–200%. This poor performance at higher values of d and f , is most likely due to the modest performance of the Tukey's method applied at the population level in identification of outliers at high values of d and f .

For the result ratio method, estimation of CV_i also improved with the increasing true underlying CV_i . Similar to the EMS method, there was also a generalized degradation of performance at high values of d and f , however to a lesser extent. This better performance with the result ratio method at high d and f values, is likely due to the Tukey outlier method being applied at the individual level, thus ameliorating the influence of outliers in comparison to being applied at the population level in the EMS method.

Performance of the modified result ratio method

The performance of this modified method is reported in Figure 2 and Table 1. This approach had overall superior performance when compared to the other three methods, with 115, 121 and 123 conditions out of a total of 128 conditions, within ± 0.1 , ± 0.2 and ± 0.4 fractional error of the true underlying CV_i respectively. Nonetheless, performance was similar to the result ratio with Tukey's rule and EMS method at $CV_i=0.2$ and 0.4. At lower values of $CV_i=0.05$ and 0.1, better performance was observed with 27 and 30 out of 32 conditions within ± 0.20 fractional error, respectively. At higher values of f and d , superior performance was also observed over the result ratio method.

Examination of poor performance at low CV_i

To understand the poor performances of the various algorithms at low CV_i , histograms of non-pathological measurements, pathological measurements, all the measurements and various stages of Tukey outlier exclusion were plotted in Supplementary Figure 2 (a–g) for $CV_i=0.05$, $d=0.4$, $f=0.4$ and $e=0$. As observed from Supplementary Figure 2 (a) and (b), there was significant overlap between the non-pathological and pathological measurements. This is due to the moderate CV_g value of 0.2, resulting in large variation in individual patient mean, μ_i . The long tail of the overall distribution shown in Supplementary Figure 2 (c) is reduced by Tukey outlier exclusion on all the measurements (Supplementary Figure 2 (d)). This distribution has many measurements

greater than 150, which are not observed in the non-pathological distribution Supplementary Figure 2 (a) and likely explain the poor performance of the EMS method.

In our modified result ratio method, Tukey outlier exclusion is performed at the ratio level and helps to remove more outlier data. This explains the improvement in the modified result ratio method over the EMS method. However, there are still significant number of measurements greater than 150. This is consistent with the poor estimate of CV_i obtained. In contrast at $CV_i=0.4$, the distribution obtained after the modified result ratio method (Supplementary Figure 2 (n)) is similar to the non-pathological distribution (Supplementary Figure 2 (h)). This results in a better estimation of the CV_i values.

Determination of performance on data without diseased patients

Lastly, as a negative control, the CV_i and fractional differences were determined for the datasets in the absence of pathological subpopulations (Supplementary Table 4). In general, the EMS, result ratio and modified result ratio yielded estimates of CV_i within ± 0.1 fractional error. The performance of EMS was superior to result ratio and modified result ratio as its outlier exclusion method was less stringent than the latter yielding better estimates in the absence of pathological measurements. Consistent to that observed in the 128 conditions above, the median method grossly underestimated the CV_i .

Discussion

In this study of performance characteristics for different approaches in indirect derivation of CV_i , we used a range of acceptable performance criterion which were much more stringent than ± 0.50 used in other published studies [15]. The use of more stringent performance criteria allows better discrimination of the different approaches for indirect estimation of CV_i .

Comparing the three original methods, we found that EMS demonstrated the best generalised performance in the presence of predominantly normal populations, and one possible reason for this performance may be the use of Tukey's rule for outlier exclusion. From an unpublished work by the authors, it was reported that Tukey is one of the better performing methods for outlier exclusion in determination of the lower and upper percentile reference limits of a non-pathological population in the setting of mixed population distributions. Furthermore, the parameters or

Figure 3: The distribution of standard deviation follows a Chi-square distribution in which the median is lower than mean.

conditions where the EMS method does not perform well, are those where Tukey's rule also does not perform well for outlier exclusion. While the EMS method performed well at high CV_i values of 0.2 and 0.4, this is not so at lower CV_i of 0.05 and 0.1. In our simulations, the CV_g was held constant at 0.2, which suggested that EMS performance is acceptable in the setting of $CV_i \geq CV_g$, but less so for $CV_i < CV_g$. This is likely due to the difficulty of resolving values from the pathological distribution when $CV_i < CV_g$, as the shift in the pathological distribution at the individual level was not sufficiently large enough to exceed the lower and upper limits of the non-pathological population distribution.

The median method grossly and consistently underestimated the CV_i , as the standard deviation follows a positively skewed Chi-square distribution. Consequently, using the median approach for derivation of CV_i in this situation will produce a lower values in comparison to using the mean and is highlighted in Figure 3.

Unlike the EMS method, the result ratio method first takes the ratio of an individual's two consecutive measurements before outlier exclusion. Using the ratio of the two measurements, this method reduces the effect of CV_g . For example, two individuals could be at either end of a pathological population distribution in the setting of a high CV_g and have very different mean values. By taking the ratio of the two measurements, the second measurements are effectively normalized with the first, negating the effect of their different individual means. Subsequent application of the Tukey's method to remove outliers in the ratio values is also important. For an individual with one measurement from the pathological distribution and one from the non-pathological distribution (irrespective of order), this ratio would be expected to fall outside the range of values for individuals with both measurements originating from the same distribution.

The importance of outlier exclusion is not limited to biological variation data and has been noted by Harris and Fraser earlier [1]. In their proposed direct sampling approach, multiple rounds of outlier detection were applied to data collected from repeated sampling of participants. They included applying outlier detection on the overall data (i.e. individual measurements), the within-subject variance and the mean values of the subjects. In our proposed modified method, we sought to combine the advantages of the EMS and the result ratio methods. In this hybrid approach, Tukey's rule is first used on all the data for outlier exclusion. This accommodates identification of pathological values in the case of $CV_i \geq CV_g$, and pathological values from individuals where both measurements are from the pathological distribution. Following this, the result ratio is calculated for the remaining paired values, which minimises the impact of CV_g . This followed by a further round of Tukey's rule outlier exclusion to further exclude the impact of outlying individuals, whereby results ratios arise in the presence of one non-pathological and one pathological result. As shown in Supplementary Figure 2 and Supplementary Table 2, a large fraction of pathological results remained even after the three rounds of Tukey's exclusion. This highlights the difficulty in outlier exclusion, due to the overlapping observed in the non-pathological and pathological measurements. However, it also demonstrates that estimates of CV_i are robust to the presence of some pathological measurements so long as gross pathological results are removed.

Consequently, the proposed modified result ratio incorporating Tukey's rule is more robust against various factors that may influence the estimation of CV_i examined in this study, and as originally noted by Fraser and Harris [1]. A major advantage of using Tukey's method is in its simplicity and the avoidance of subjective optimization for outlier exclusion as required for Bhattacharya. Despite the apparent advantageous performance of the proposed hybrid method, it is still important to use defined *a priori* clinical criteria such as requesting location, clinical history, and other related laboratory results, to exclude as many pathological values as possible. Within the 128 conditions simulated in this study, it is possible to derive CV_i values close to the 'true' underlying non-pathological population following careful selection of the statistical approaches to minimise the influence of results from pathological populations.

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: This study involves only numerical simulation and does not require informed consent.

Ethical approval: This study involves only numerical simulation and does not require local Institutional Review Board.

References

1. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989;27:409–37.
2. Aarsand AK, Røraas T, Fernandez-Calle P, Ricos C, Díaz-Garzón J, Jonker N, et al. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. *Clin Chem* 2018;64:501–14.
3. Aarsand AK, Díaz-Garzón J, Fernandez-Calle P, Guerra E, Locatelli M, Bartlett WA, 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;64:1380–93.
4. Carobene A, Marino I, Coşkun A, Serteser M, Unsal I, Guerra E, 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;63:1527–36.
5. Carobene A, Aarsand AK, Guerra E, Bartlett WA, Coşkun A, Díaz-Garzón J, et al. European biological variation study (EuBIVAS): within- and between-subject biological variation data for 15 frequently measured proteins. *Clin Chem* 2019;65:1031–41.
6. Ceriotti F, Díaz-Garzón Marco J, Fernández-Calle P, Maregnani A, Aarsand AK, Coskun A, et al. The European Biological Variation Study (EuBIVAS): weekly biological variation of cardiac troponin I estimated by the use of two different high-sensitivity cardiac troponin I assays. *Clin Chem Lab Med* 2020;58:1741–7.
7. Cavalier E, Lukas P, Bottani M, Aarsand AK, Ceriotti F, Coşkun A, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates of beta-isomerized C-terminal telopeptide of type I collagen (beta-CTX), N-terminal propeptide of type I collagen (PINP), osteocalcin, intact fibroblast growth factor 23 and uncarboxylated-unphosphorylated matrix-Gla protein—a cooperation between the EFLM Working Group on Biological Variation and the International Osteoporosis Foundation-International Federation of Clinical Chemistry Committee on Bone Metabolism. *Osteoporos Int* 2020;31:1461–70.
8. Bottani M, Banfi G, Guerra E, Locatelli M, Aarsand AK, Coşkun A, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates for serum bioactive parathyroid hormone based on weekly samplings from 91 healthy participants. *Ann Transl Med* 2020;8:855.
9. Carobene A, Guerra E, Marqués-García F, Boned B, Locatelli M, Coşkun A, et al. Biological variation of morning serum cortisol: updated estimates from the European biological variation study (EuBIVAS) and meta-analysis. *Clin Chim Acta* 2020;509:268–72.

10. Clouet-Foraison N, Marcovina SM, Guerra E, Aarsand AK, Coşkun A, Díaz-Garzón J, et al. Analytical performance specifications for lipoprotein(a), apolipoprotein B-100, and apolipoprotein A-I using the biological variation model in the EuBIVAS population. *Clin Chem* 2020;66:727–36.
11. Carobene A, Guerra E, Locatelli M, Cucchiara V, Briganti A, Aarsand AK, et al. Biological variation estimates for prostate specific antigen from the European Biological Variation Study; consequences for diagnosis and monitoring of prostate cancer. *Clin Chim Acta* 2018;486:185–91.
12. Carobene A, Lao EG, Simon M, Locatelli M, Coşkun A, Díaz-Garzón J, et al. Biological variation of serum insulin: updated estimates from the European Biological Variation Study (EuBIVAS) and meta-analysis. *Clin Chem Lab Med* 2022;60: 518–22.
13. Bottani M, Aarsand AK, Banfi G, Locatelli M, Coşkun A, Díaz-Garzón J, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates for serum thyroid biomarkers based on weekly samplings from 91 healthy participants. *Clin Chem Lab Med* 2022;60:523–32.
14. Jones GRD, Haeckel R, Loh TP, 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.
15. Loh TP, Ranieri E, Metz MP. Derivation of pediatric within-individual biological variation by indirect sampling method: an LMS approach. *Am J Clin Pathol* 2014;142:657–63.
16. Loh TP, Sethi SK, Metz MP. Paediatric reference interval and biological variation trends of thyrotropin (TSH) and free thyroxine (T4) in an Asian population. *J Clin Pathol* 2015;68: 642–7.
17. Loh TP, Metz MP. Indirect estimation of pediatric between-individual biological variation data for 22 common serum biochemistries. *Am J Clin Pathol* 2015;143:683–93.
18. Chai JH, Flatman R, Teis B, Sethi SK, Badrick T, Loh TP. Indirect derivation of biological variation data and analytical performance specifications for therapeutic drug monitoring activities. *Pathology* 2019;51:281–5.
19. Jones GRD. Estimates of within-subject biological variation derived from pathology databases: an approach to allow assessment of the effects of age, sex, time between sample collections, and analyte concentration on reference change values. *Clin Chem* 2019;65:579–88.
20. Røraas T, Petersen PH, Sandberg S. Confidence intervals and power calculations for within-person biological variation: effect of analytical imprecision, number of replicates, number of samples, and number of individuals. *Clin Chem* 2012;58: 1306–13.
21. Hinkley DV. On the ratio of two correlated normal random variables. *Biometrika* 1969;56:635–9.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2021-0442>).