

Research Article

Selçuk Korkmaz*, Gökmen Zarasız, Dinçer Göksülük, Mehmet Senes, Cem Sönmez and Dogan Yucel



BioVar: an online biological variation analysis tool

BioVar: çevrimiçi bir biyolojik varyasyon analizi aracı.

<https://doi.org/10.1515/tjb-2019-0437>

Received October 17, 2019; accepted June 1, 2020; published online July 27, 2020

Abstract

Objectives: Biological variation (BV) analysis of laboratory tests gets increased attention due to its practical applications. These applications include correct interpretation of laboratory tests, the decision on the availability of reference intervals, contributions to clinical decision-making. It is critical to derive the BV information accurately and reliably. Another crucial step is to perform the statistical analysis of the BV data. Although there are updated and comprehensive guidelines, there is no reliable and comprehensive tool to perform statistical analysis of BV data.

Methods: We presented BioVar, an online tool for statistical analysis of the BV data based on available and updated guidelines.

Results: This tool can be used (i) to detect outliers, (ii) to control normality assumption, (iii) to check steady-state condition, (iv) to test homogeneity assumptions, (v) to

perform subset analysis for genders, (vi) to perform analysis of variance to estimate components of variation and (vii) to identify analytical performance specifications of laboratory tests. Moreover, plots can be created at each step of outlier detection to inspect outliers and compare gender groups visually. An automatic report can be generated and downloaded.

Conclusion: The tool is freely available through [turcosa.shinyapps.io/biovar/](https://shinyapps.io/biovar/), and source code is available on the Github: github.com/selcukkorkmaz/BioVar.

Keywords: Biological variation; Laboratory tests; Nested ANOVA; Variance components; Web tool.

Öz

Amaç: Laboratuvar testleri için uygulanan biyolojik varyasyon (BV) analizi, pratik uygulamaları nedeniyle oldukça ilgi görmektedir. Bu uygulamalar arasında laboratuvar testlerinin doğru yorumlanması, referans aralıkları hakkında karar verme, klinik karar vermeye katkı sayılabilir. Bu nedenle, BV analizlerini deneysel olarak doğru ve güvenilir bir şekilde elde etmek çok önemlidir. BV analizinde bir diğer önemli adım verilerin istatistiksel analizidir. Literatürde, güncel, kapsamlı ve ayrıntılı kılavuzlar bulunmasına rağmen, BV verilerinin istatistiksel analizini yapmak için güvenilir ve kapsamlı bir araç bulunmamaktadır.

Gereç ve Yöntem: Bu çalışmada, BV verilerinin literatürdeki mevcut ve güncel kılavuzlara dayanarak istatistiksel analizlerini gerçekleştirmek için çevrimiçi bir araç geliştirildi.

Bulgular: Bu çevrimiçi analiz aracı (i) aykırı değerleri saptamak, (ii) normalilik varsayımlını kontrol etmek, (iii) kararlı-durum koşulunu kontrol etmek, (iv) homojenlik

*Corresponding author: Selçuk Korkmaz, Trakya University, Faculty of Medicine, Department of Biostatistics and Medical Informatics Edirne, Turkey, E-mail: selcukkorkmaz@gmail.com. <https://orcid.org/0000-0003-4632-6850>

Gökmen Zarasız: Erciyes University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Kayseri, Turkey

Dinçer Göksülük: Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara, Turkey

Mehmet Senes: Ministry of Health, University of Health Sciences, Ankara Training and Research Hospital, Medical Biochemistry, Ankara, 06610, Turkey. <https://orcid.org/0000-0002-7924-5478>

Cem Sönmez: Turcosa Analytics Solutions Ltd Co, Erciyes Teknopark 5, Kayseri, Turkey

Dogan Yucel: Ankara Training and Research Hospital, Medical Biochemistry, Ulucanlar Caddesi, Ankara, 06340, Turkey

varsayımlarını kontrol etmek, (v) cinsiyet gruplarını karşılaştırmak, (vi) varyans bileşenlerini kestirmek ve (vii) analitik performans özelliklerini belirlemek için kullanılabılır. Ayrıca, her aykırı değer tespit admımda cinsiyet gruplarına göre aykırı değerleri görsel olarak denetlemek için grafikler oluşturulabilir. Bununla birlikte, tüm analiz sonuçlarını içeren otomatik bir rapor oluşturulabilir ve indirilebilir.

Sonuç: Bu çevrim içi araca <https://turcosa.shinyapps.io/biovar/> adresinden ücretsiz olarak erişilebilir. Tüm kaynak kodlara Github üzerinden erişilebilir: <https://github.com/selcukkorkmaz/BioVar>.

Anahtar kelimeler: Biyolojik varyasyon; Laboratuvar testleri; Varyans bileşenleri; Çevrimiçi araç.

Introduction

Laboratory tests have vital importance in diagnosing and monitoring patients. Therefore, it is essential to accurately analyze the biological variability of an analyte to interpret its result correctly [1]. Biological variation (BV) data is used to assess the analytical quality (i.e., bias, imprecision, and total error) of many laboratory tests [2]. There are three main components of variation of a laboratory test: analytical variation (CV_A), within-subject variation (CV_I), and between-subject variation (CV_G). Moreover, there is also a pre-analytical variation, which includes subject preparation and collection, transportation, storage, and handling of the biological analyte [3]. Analytical variation, which is associated with the analytical performance, represents the measurement errors. The BV, on the other hand, is related to physiological changes of laboratory test concentration, which occurs in a particular biological fluid in each subject because of certain biological factors [3, 4]. At each time point, the concentration of any constituent in a biological fluid is the result of a steady-state equilibrium, which is called homeostatic point [5]. Here, CV_I represents a random fluctuation of a measurand concentration around the homeostatic point [3, 5]. On the other hand, CV_G is the difference among homeostatic points for the same laboratory test in different subjects under the same physiological conditions [3, 4].

The BV data of a laboratory test allows the derivation of critical information regarding the correct application and interpretation of a measurand [6]. The BV data can be used to evaluate the utility of population-based reference intervals, to produce biological reasons for separating reference intervals by age and sex, to estimate significant

changes of results in an individual, to calculate the number of samples for a laboratory test to assure that mean value of the measurand is close to individual's homeostatic set point, and to derive analytical performance specifications to determine the availability of a laboratory test for its clinical use [3, 4].

There are many efforts in the literature to regulate the statistical analysis process of the BV data. One of the early effort made by Fraser and Harris (1989), where the authors presented experimental procedures and statistical methods for the evaluation of the BV data [6]. Fraser et al. (1997) proposed methods to calculate analytical performance specifications for quality levels (minimum, desirable, and optimal, for imprecision, bias, and total error) from the BV data [7]. In 1999, a conference held in Stockholm by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and International Union of Pure and Applied Chemistry (IUPAC) to derive the analytical goals of laboratory tests and to define quality specifications [8]. A follow-up conference held in Milan in 2014 to re-evaluate the Stockholm consensus [9].

Recently, several checklists and pipelines have been published to establish a reliable framework for the statistical analysis of the BV data. Bartlett et al. (2015) reported a critical appraisal checklist produced by the Working Group on Biological Variation (WG-BV, <https://www.eflm.eu/site/page/a/1148>) established by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM, <https://www.eflm.eu/>) to define standards for production and reporting and transmission of the BV data [10]. Braga and Panteghini (2016) used these checklist recommendations and presented a pipeline, where they represented aspects of experimental and statistical analysis procedures for the BV data [3]. Furthermore, recently, a critical appraisal checklist, called Biological Variation Data Critical Appraisal Checklist (BIVAC), created by the WG-BV and Task and Finish Group for the Biological Variation Database (TFG-BVD, <https://www.eflm.eu/site/page/a/1394>) on behalf of the EFLM to verify whether publications have included all vital elements of the BV data that may impact the veracity of the BV estimates [11].

According to these checklists and pipelines, the steps of analyzing the BV data can be simply divided into seven steps: (i) Detecting outliers, (ii) Controlling normality assumption, (iii) Checking steady-state condition, (iv) Testing homogeneity assumptions, (v) Performing subset analysis for genders, (vi) Performing analysis of variance, and (vii) Identifying analytical performance specifications. Although the statistical analysis of the BV data has critical importance to obtain correct results for laboratory tests,

there is no easily accessible, freely available, easy-to-use tool to perform such statistical analyses. Here, we presented an online tool for the statistical analysis of the BV data based on the guideline by Bartlett et al. (2015), the pipeline by Braga and Panteghini (2016), and the updated checklist by Aarsand et al. (2018).

Material and methods

In this section, we will explain steps of statistical analysis of the BV data and development of the online BioVar web-tool. First, we will begin with outlier detection tests. Next, we will move on to checking the normality assumption. Then, we will describe the steady-state condition, homogeneity assumptions, and subset analysis procedure to decide whether the estimates should be reported separately for each gender. Finally, we will introduce the analysis of variance (ANOVA) method to estimate variance components. Moreover, we will give information regarding the index of individuality (II), reference change values (RCV), and analytical performance specifications.

Outlier detection

The first step of statistical evaluation of the BV data is outlier detection. This step should be carried out rigorously and thoroughly since even a single outlying observation can significantly influence the variability of the BV data components. It is recommended that this step should be performed using Cochran's test and Reed's criterion in three steps [3, 4]:

Step 1 Use the Cochran's test to check outliers among replicates as follows:

- Calculate the variance of each replicate for each subject within each time sample:

$$\text{var}(r_{ij1}, r_{ij2}) = \frac{(r_{ij2} - r_{ij1})^2}{2}$$

where r_{ij1} is the value of replicate 1 for i^{th} subject and j^{th} sample, and r_{ij2} is the value of the replicate 2 for i^{th} subject and j^{th} sample.

- Calculate the ratio of maximum replicate variance ($\text{var}(m_r)$) to the sum of replicate variances ($\text{var}(s_r)$):

$$\text{var}(m_r) / \text{var}(s_r) = \frac{\max(\text{var}(r_{ij1}, r_{ij2}))}{\sum_i \sum_j \text{var}(r_{ij1}, r_{ij2})}$$

- Calculate the critical value (C) using Cochran's table [12] at a predefined significance level (i.e., 0.05).
- if $\text{var}(m_r) / \text{var}(s_r) > C$ then remove the replicate values with the highest variance [4].
- Repeat this procedure until $\text{var}(m_r) / \text{var}(s_r) < C$.

Step 2 Use the Cochran's test to determine whether a subject distribution was greater or smaller than those of the group as a whole. The procedure is as follows:

- Calculate individual variances of each subject, $\text{var}(i)$.
- Calculate the ratio of maximum individual variance ($\text{var}(m_i)$) to the sum of individual variances ($\text{var}(s_i)$):

$$\text{var}(m_i) / \text{var}(s_i) = \frac{\max(\text{var}(i))}{\sum_i \text{var}(i)}$$

- Calculate Cochran's critical value (C):
- if $\text{var}(m_i) / \text{var}(s_i) > C$ then remove the subject with the highest individual variance [3, 4].
- Repeat this procedure until $\text{var}(m_i) / \text{var}(s_i) < C$.

Step 3 Use Reed's criterion to determine whether the mean value of any subject is significantly greater or smaller compared to other subjects. Reed's criterion is a simple procedure and can be applied as follows [3, 6]:

- Calculate the mean value of each subject.
- Calculate the difference between the second-lowest mean value and the lowest mean value (lowest extreme, LE).
- Calculate the difference between the highest mean value and the second highest mean value (highest extreme, HE).
- Calculate the difference between the maximum and minimum mean values (i.e., range) and use one-third of the range as the critical value.
- If $LE > 1/3 Range$ then remove the subject with the lowest mean value from the analysis.
- If $HE > 1/3 Range$ then remove the subject with the highest mean value from the analysis.

After detecting and removing outliers appropriately, the analysis procedure can be carried out with the remaining samples and subjects.

Normality

After the application of outlier detection, the normality assumption of the BV data should be checked. As recommended by Braga and Panteghini (2016), the normality control should be performed in two steps: (i) the normality

of each measurand and (ii) the normality of mean values of all subjects [3]. The Shapiro–Wilk test is one of the most widely used and powerful tests for checking normality assumption [13]. The estimation of variance components using a nested ANOVA does not assume a normal distribution or any other distributions for the model effects [6, 14, 15]. However, calculation of confidence intervals and statistical significance tests usually require normality assumption [11].

Moreover, the direct calculation of RCV also requires normal distribution [16, 17]. An appropriate transformation, such as logarithmic or any other transformation method, can be applied to the BV data in the case of non-normal distribution. If a transformation applied to the BV data, the transformed-data should be re-tested for normality using the Shapiro–Wilk test.

Steady State

Since estimations of CV_I and RCV assume a steady-state condition [18], this assumption should be checked before performing statistical analysis for the BV data. This assumption requires that data from all subjects should be in a steady-state. In a steady-state condition, there should be no trend in the concentrations of measurand during the study period [16]. Linear regression analysis can be performed on mean [19] or median [20] group value over the whole study period to verify that all subjects are in the steady-state. Subjects are considered in the steady-state if the 95% confidence interval of the slope of the regression line contains 0 [19]. If this assumption does not hold, then a data transformation, such as multiples of the median (MoM) and its natural logarithm ($\ln MoM$), can be used to establish a kind of steady-state condition [18]. The MoM is a simple measure that shows how far a single test result from the median of the sample results. Let assume Y_{ijk} is the value of k^{th} replicate for j^{th} sampling time from i^{th} subject, then, the MoM value for each measurement of Y_{ijk} can be calculated as follows:

$$MoM(Y_{ijk}) = \frac{Y_{ijk}}{\text{Median}(Y_j)}$$

Finally, natural logarithm of MoM ($\ln(MoM(Y_{ijk}))$) is used to make the distribution of the pooled residuals Gaussian [16, 21].

Homogeneity

Another critical assumption is the homogeneity of within-subject and analytical variability (between replicates). The

Bartlett [22] test can be used to check the homogeneity assumptions. The analysis should not be continued in the case of heterogeneity since the heterogeneity will cause estimations not to be generalizable to the whole population [11]. One alternative, in this case, is to remove outliers until homogeneity is obtained [20].

Subset analysis

The significant differences in the CV_I and CV_G between gender groups (i.e., female and male) must be controlled to determine whether the estimates should be reported separately for genders. If the 95% confidence intervals of CV_I and CV_G overlap between genders, then it is concluded that there are no significant differences between gender groups [19]. Therefore, all-subjects results should be reported for CV_I and CV_G . In the case of non-overlapping confidence intervals, the results of CV_I and CV_G should be reported separately for each gender.

Moreover, the differences between the mean values of genders can be compared using a Student's t -test. If there is no significant difference between the mean values of males and females, all-subjects CV_I and all-subjects CV_G are used to calculate the analytical performance specifications. In the case of the significant difference between the mean values of genders, the lowest of the two CV_G estimate is used to calculate the analytical performance specifications [23].

Nested ANOVA

After detecting and removing outliers, controlling normality assumptions, checking the steady-state condition, testing homogeneity assumptions, and performing subset analysis, variance components (e.g., CV_G , CV_I , CV_A) can be estimated. A two-fold nested ANOVA is the most widely used type of method to estimate variance components for the BV data. Generally, a number of samples (j) for the BV data are collected from a certain number of subjects (i), and the samples are analyzed within replicates (k). A general structure of a nested ANOVA can be expressed as the following model:

$$Y_{ijk} = \mu + G_i + I_{ij} + A_{ijk}$$

where Y_{ijk} is the value of k^{th} replicate for j^{th} sample from i^{th} subject, μ is the population mean, G_i is the effect of CV_G , I_{ij} is the effect of CV_I , and A_{ijk} is the effect of CV_A . It is assumed that the variance components

(G_i, I_{ij}, A_{ijk}) are mutually independent and uncorrelated random variables with a mean of zero and variances σ_G^2 , σ_I^2 , σ_A^2 , respectively [16].

As previously mentioned in the *Normality* section, the calculation of confidence intervals and statistical significance tests as well as the direct calculation of RCV, require normality assumption. One option is to apply natural logarithmic (ln) transformation in the case of non-normal distribution. First, the raw data are transformed using ln transformation, and then the nested ANOVA method is applied to the transformed data. Finally, the transformed results are transformed back to the original scale.

Moreover, the estimation of CV_I is vital for diagnosing and monitoring patients and for setting analytical performance specifications [16]. A comprehensive simulation study performed by Roraas et al. (2016) showed that CV-ANOVA is the best-performing method, resulting in the least biased and narrowest range estimates based on the percentiles for the CVI compared to standard ANOVA and ln-ANOVA. In the CV-ANOVA method, each subject's data is divided by that subject's mean value, then a standard nested ANOVA procedure is applied [16] and this method can be applied independently of the data distribution [11, 16]. The CV-ANOVA, however, does not provide an estimate of CV_G , since each individual has a mean value of 1. Therefore, the CV_G can be estimated by the standard ANOVA [16].

Index of individuality and reference change values

The index of individuality (II) is a useful parameter that can be used to assess the utility of population-based reference intervals. The II compares the biological variation of a laboratory test within an individual to that between all individuals, and it can be calculated as follows [24]:

$$II = CV_I / CV_G$$

According to the study results by Harris (1981), if the II is <0.6 , the measurand shows low individuality meaning that the values for an individual cover only a small part of the reference interval. Low II indicates the limited utility of population-based reference intervals. On the other hand, if II is >1.4 , the measurand shows high individuality meaning that dispersion of values of each individual span much of the distribution of the reference interval [25]. Hence, population-based reference values are appropriate. When a measurand showing low II, it is recommended not to use conventional population-based reference intervals. Instead, it is suggested to use the reference change values

(RCV) for monitoring of longitudinal time changes of measurand concentrations in an individual [3]. The RCV is described as the difference needed between two serial results from an individual that a change to be statistically significant [26]. The RCV can be calculated as follow:

$$RCV(\alpha) = \pm z_\alpha \times 2^{1/2} \times (CV_A^2 + CV_I^2)^{1/2}$$

where z_α is 1.96 for 0.05 significance level and z_α is 2.58 for 0.01 significance level. This method assumes a normal distribution and results in symmetrical limits. On the other hand, Fokkema et al. (2006) suggested calculation of asymmetric limits when model effects are assumed to be log-normal distributed [27]. Asymmetric limits for the RCV can be calculated as follow:

$$RCV_{ln}(\alpha) = 100 \% \times \exp \left(\pm z_\alpha \times 2^{1/2} \times (CV_{lnA}^2 + CV_{lnI}^2)^{1/2} \right)$$

Here, CV_{lnI} is the CV on the log-transformed data, as $CV_{lnI} = (\ln(1+CV_I^2))^{1/2}$. Roraas et al. (2016) recommend the use of the asymmetrical limits regardless of the distribution of data [16].

Analytical performance specifications

There are two main sources of error for a measurand: random error and systematic error. Random error occurs in replicates and varies unpredictably, whereas systematic error remains constant or varies in a predictable manner [28]. The systematic error is evaluated by the bias, while the random error is assessed by the imprecision measured by the CV [29]. To measure the total error of a measurand, allowable bias and allowable imprecision are combined in a single metric as the total allowable error (TEa):

$$TEa = \text{Allowable Bias} + 1.65 \times \text{Allowable Imprecision}$$

Here, 1.65 implies 95% of the results will within the TEa limit under normal distribution [30]. Fraser et al. (1997) proposed to calculate minimum, desirable, and optimal quality levels for allowable imprecision, allowable bias, and TEa as following [7]:

- a. For imprecision:
 - i. Optimum allowable imprecision $<0.25 \times CV_I$
 - ii. Desirable allowable imprecision $<0.50 \times CV_I$
 - iii. Minimum allowable imprecision $<0.75 \times CV_I$
- b. For bias:
 - i. Optimum allowable bias $<0.125 \times (CV_I^2 + CV_G^2)^{1/2}$
 - ii. Desirable allowable bias $<0.250 \times (CV_I^2 + CV_G^2)^{1/2}$
 - iii. Minimum allowable bias $<0.375 \times (CV_I^2 + CV_G^2)^{1/2}$

The screenshot shows the 'Data upload' tab of a web-based tool. On the left, there's a sidebar for 'Input data' with options to 'Upload a file' or 'Load example data'. Below that, 'Select a data format:' has 'Long format' selected. Under 'Upload a delimited text file:', there's a 'Browse...' button pointing to 'example_data.txt', and an 'Upload complete' button. A 'Delimiter' checkbox is also present. At the top, tabs for 'Introduction', 'Data upload' (selected), 'Analysis', and 'Manual' are visible. Below the tabs is a search bar and a table titled 'Subject' with columns for Subject, Gender, Time, Replicate, and Measurement. The table contains 10 entries of female subjects with time points from 1 to 5 and measurements from 9.4633 to 9.084. Navigation buttons for 'Previous' and 'Next' are at the bottom.

Figure 1: Data upload tab of the tool. The dataset uploaded to the tool as in a long format.

c. For total error:

- i. Optimum TEa $< 1.65 \times (0.25 \times CV_I) + 0.125 \times (CV_I^2 + CV_G^2)^{1/2}$
- ii. Desirable
TEa $< 1.65 \times (0.50 \times CV_I) + 0.250 \times (CV_I^2 + CV_G^2)^{1/2}$
- iii. Minimum
TEa $< 1.65 \times (0.75 \times CV_I) + 0.375 \times (CV_I^2 + CV_G^2)^{1/2}$

The analytical performance specifications should be defined for each measurand to ensure that the measurement error does not affect the BV results. However, caution is needed when calculating the analytical performance

specifications. As pointed out by Oosterhuis (2011), the maximum allowable bias and imprecision can be derived separately from the BV data, and to combine these two values into a single expression has no theoretical foundation and may lead to gross over-estimation of the TEa [30].

Web-tool development

A web-based tool is developed for the BV data analysis using the R language environment. This tool can be used to perform outlier detection, to control normality assumption,

The screenshot shows the 'Analysis' tab of the tool. On the left, a sidebar titled 'Select Variables' includes dropdown menus for 'Analyte', 'Subject', 'Gender', 'Time', 'Replicate', and 'Choose calculation data' (with options 'Original' and 'Advanced Options'). At the bottom of this sidebar is a 'Run Analysis' button. At the top, tabs for 'Introduction', 'Data upload' (disabled), 'Analysis' (selected), and 'Manual' are visible. Below the tabs is a row of buttons labeled 'a)', 'b)', 'c)', 'd)', 'e)', 'f)', 'g)', and 'h)'. Each button corresponds to a different statistical analysis module: a) Outliers, b) Normality, c) Steady State, d) Homogeneity, e) Subset, f) ANOVA, g) Plots, and h) Report.

Figure 2: Analysis tab of the tool. a) Outlier results can be obtained. b) The normality assumption can be controlled. c) The steady-state condition can be checked. d) Homogeneity assumptions can be tested. e) Subset analysis can be performed. f) Analysis of variance (ANOVA) components can be estimated. g) Mean and absolute range plots can be created. h) A report can be generated.

Table 1: Results of Cochran's test to detect outliers in the sets of replicates.

Subject	Time	1st replication	2nd replication	Variance	var (m_i)/var (s_i) [*]	Critical value
3	1	11.9173	11.0311	0.393	0.071	0.031
30	5	11.815	11.0762	0.273	0.054	0.032
11	9	11.9003	12.4969	0.178	0.038	0.032

*var (m_i)/var (s_i): The ratio of maximum replicate variance to the sum of replicate variances.

to check the steady-state condition, to test homogeneity assumptions, to perform subset analysis for genders, to carry out nested ANOVA, and to generate plots. The tool is developed using shiny package of the Rlanguage and other languages, including HTML, CSS, and jQuery used to polish the front-end. Several R packages were used both in the back-end and front-end of the tool. The tool has an interactive interface. Users can upload their datasets in .txt format. There is an available example dataset in the tool to help users to test the applicability of the tool. In an appropriate dataset, rows must represent subjects, while columns must represent variables, and the first row must include a header that indicates variable names. Analysis procedures are easy to perform as well. Users can select any desired analysis method, choose appropriate variables, set options, and execute the analysis from the sidebar panel of the tool. Desired outputs will be displayed on the main panel of the tool. Table and plot results will be displayed on separate sub-tabs on the main panel to avoid confusion. An automatic report can be downloaded in HTML file format. A help page can be found on the web-page of the tool. The web-tool can be accessed freely at <https://turcosa.shinyapps.io/biovar/>.

Results

In this section, we will analyze a simulated BV dataset and present its results to illustrate the usage of the tool. First,

Table 2: Normality test on mean values of subjects.

Test	Test statistic	p-Value
Shapiro-Wilk	0.994	0.998

Table 3: Linear regression result for mean group values over the study period.

Variable	Estimate	Std. Error	t value	p-Value	Lower limit	Upper limit
(Intercept)	10.012	0.04	252.555	<0.001	9.921	10.103
Time	-0.0003	0.006	-0.039	0.97	-0.015	0.014

we will load our dataset using the *Data upload* tab. Datasets can be uploaded as long format or wide format. If a dataset uploaded as in wide format, it must be converted to the long format by selecting appropriate headers, including the name of the measurand, subject, gender, time, and replicate. After loading a proper dataset, one can move to the *Analysis* tab to perform the BV data analysis. Before running the analysis, suitable variable names, including the measurand, subject, gender, time, and replicate, must be selected.

Table 4: Homogeneity test results for analytical and within-subject variability.

Variability	Corrected chi-square	p-Value
Analytical	426.244	0.94
Within-subject	60.952	0.083

Table 5: Gender comparison for between-subject and within-subject variation.

Gender	Source	CV%	Lower limit	Upper limit
Female	Between	14.033	10.874	19.728
Male	Between	14.074	10.903	19.79
Female	Within	4.118	3.751	4.561
Male	Within	4.324	3.941	4.786

Table 6: Mean and associated confidence interval of the measurand, index of individuality (II) and the number of samples (n).

Mean	Lower Limit (95%)	Upper Limit (95%)	II	n
10.022	10.009	10.034	0.303	954

Table 7: Analysis of variance results for all subjects.

Source	Sigma	CV (%)	Lower (95%)	Upper (95%)
Between	1.397	13.937	11.577	17.487
Within	0.423	4.225	3.952	4.536
Analytical	0.099	0.987	0.928	1.054
Total	1.463	14.596	12.364	18.017

Table 8: Analytical performance specifications for all subjects.

Quality Levels	Imprecision (%)	Bias (%)	Total allowable error (%)
Optimal	<1.056	<7.36	<9.103
Desirable	<2.112	<14.721	<18.207
Minimal	<3.169	<22.081	<27.31

Table 9: Reference change value results.

RCV (%)	Lower limit (%)	Upper LIMIT (%)
12.026	11.292	12.866

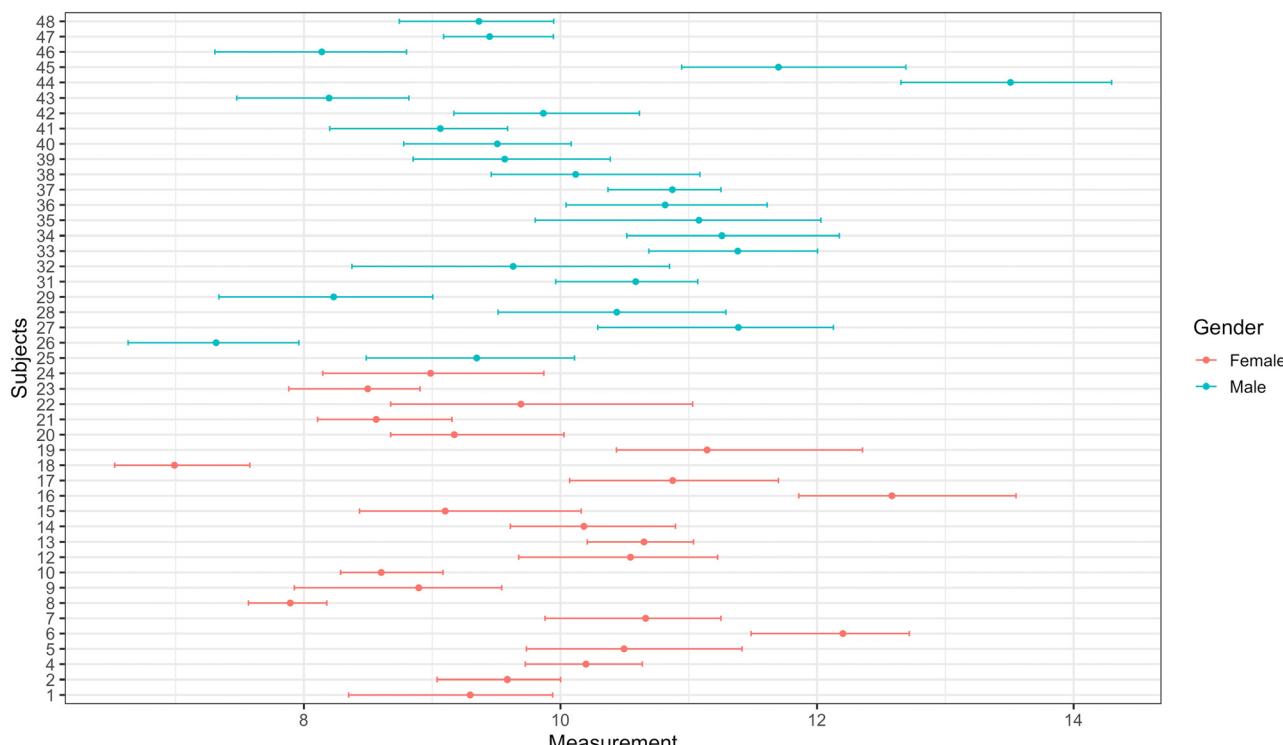
Moreover, there are several advanced options for selecting normality and homogeneity tests, removing outliers, changing the number of decimals, and type 1 error rate. After

choosing the desired calculation method among original, ln-transformed, transform back to original, CV-ANOVA, MoM, and lnMoM, the BV data analysis can be run. Here, we will examine outlier detection (*Outliers* tab), normality assumption (*Normality* tab), steady-state condition (*Steady State* tab), homogeneity assumptions (*Homogeneity* tab) subset analysis (*Subset* tab), ANOVA results (*ANOVA* tab), and mean and absolute range plots (*Plot* tab) for a simulated BV dataset. Furthermore, an automatic report can be generated and downloaded using the *Report* tab.

Dataset

We simulated a dataset based on R codes provided by Roraas (2017) [31] with a nested design following the scheme of 48 individuals (24 males and 24 females) with 20 samples (time points) measured in replicate. The CV_G , CV_I , and CV_A were set to 15, 5, and 1%, respectively. The simulated dataset uploaded to the tool as in a long format using the *Data upload* tab (Figure 1).

After uploading the data set, we can now move on to the *Analysis tab* to perform the BV data analysis (Figure 2). There are separate sub-tabs in this section. Outlier detection results can be obtained using Cochran's test and Reed's criterion (Figure 2a). The normality assumption can

**Figure 3:** Mean and absolute range plot after removing outliers among replicates using Cochran's test.

be checked using the Shapiro–Wilk test (Figure 2b). The steady-state condition can be controlled by performing linear regression (Figure 2c). Homogeneity of analytical and within-subject variabilities can be tested using the Bartlett test (Figure 2d). Subset analysis can be performed by comparing 95% confidence intervals of CV_1 and CV_G for males and females (Figure 2e). Nested ANOVA can be carried out to estimate variance components and analytical performance specifications for all-subjects and separately for males and females (Figure 2f). Mean and absolute range plots can be obtained after each outlier detection step (Figure 2g). An analysis report can be generated and downloaded in HTML file format (Figure 2h).

BV data analysis

First, outlier detection performed to detect outlying subjects using the *Outliers* tab. Three replicate measurements were detected as outliers using Cochran's test (Table 1), and no subjects found as outlying subjects using Cochran's test and Reed's criterion. Accordingly, the three outlying replicates were removed from the dataset.

Next, we move on to checking the normality assumption using the *Normality* tab. First, we performed normality tests on sets of results from each individual using the Shapiro–Wilk test and found that all subjects follow normal distributions (all p values > 0.05). Then, we move on to check the normality assumption on mean values of the subjects using the Shapiro–Wilk test, and it is concluded that the data is assumed to be normal on mean values of subjects ($p=0.998$, Table 2).

Next, the *Steady-State* tab is used to check whether all subjects were in the steady-state condition during the study period. Linear regression analysis performed on the mean group values over the study period (Table 3), and the steady-state assumption was concluded since 95% confidence interval of the regression coefficient included 0 (95% CI: $-0.015 - 0.014$).

The *Homogeneity* tab is used to test the homogeneity of within-subject and analytical variability (between replicates). The Bartlett test is used to test the homogeneity assumptions, and the test results showed that the homogeneity assumptions are valid (Table 4) for both analytical variability ($p=0.940$) and within-subject variability ($p=0.083$).

The *Subset* tab is used to determine whether CV_1 and CV_G results should be reported separately for gender groups. Since the 95% confidence intervals of both CV_1 and CV_G overlap between males and females, it is concluded

that all-subjects estimates should be reported for CV_1 and CV_G (Table 5).

Next, the *ANOVA* tab is used to display (i) mean of the measurand with its confidence interval, II and the number of samples (Table 6), (ii) ANOVA table (Tables 7), (iii) analytical performance specifications (Tables 8) and (iv) RCV result with lower and upper limits (Table 9). For all subjects; CV_A is founded as 0.987% (95% CI: 0.928% – 1.054%), CV_1 is 4.225% (95% CI: 3.952 – 4.536%), CV_G is 13.937% (95% CI: 11.577 – 17.487%), II is 0.303 and RCV is founded as 12.026% (11.292 – 12.866%).

The *Plot* tab can be used to create mean and absolute range plots (Figure 3). These plots are especially useful to inspect outlying subjects and compare subgroups (i.e., genders). Four plots will be displayed in this tab: (i) before outlier detection, (ii) after removing outliers among replicates using Cochran's test, (iii) after removing outlying subjects using Cochran's test, and (iv) after removing outlying subjects using Reed's criterion. Finally, one can download the whole report using the *Report* tab. The report can be downloaded as an HTML format. The report includes all the results and the tables generated by the tool.

Discussion

The BV data has crucial importance on the correct interpretation of laboratory tests, hence diagnosis and monitoring patients. Statistical analysis of the BV data is as essential as to derive them experimentally. Each step in the statistical analysis procedure should be carried out thoroughly and rigorously. Since even one influential observation might have a significant impact on the variance components of the BV data, this procedure should start with the outlier detection step. This step includes the detection of outliers between replicates, within-subjects, and between subjects, using Cochran's test and Reed's criterion. Another important step is checking the normality assumption. Since, the normality assumption is necessary for confidence interval and RCV to be calculated directly [11], this assumption must be controlled before proceeding the analysis. One option in the case of the non-normal distribution is to apply ln-transformation to the BV data. If data follow the normal distribution after the transformation, then the analysis procedure should be continued using the ln-transformed data. CV-ANOVA is another data transformation method, which can be applied independently of the data distribution [11, 16].

Moreover, the estimation of CV_1 and RCV requires a steady-state condition, where the concentration of

measurand does not change systematically during the study period [18]. A data transformation using MoM and lnMoM can be used to create a steady-state-like condition. Furthermore, homogeneity of analytical and within-subject variability should be checked prior to the nested ANOVA analysis to ensure that estimations can be generalized to the whole population [11]. After checking all the necessary assumptions and/or applying appropriate data transformation, the BV data can be analyzed using two-fold nested ANOVA. Here, variance components (CV_A , CV_I , and CV_G) can be estimated, and quality measures (imprecision, bias, and total error) can be obtained. Finally, mean and absolute range plots can be created to inspect the sample results before and after outlier detection visually.

Although the statistical analysis of the BV data has critical importance on obtaining correct results for laboratory tests, there is no easily accessible, freely available, easy-to-use tool to perform such statistical analyses. Here, we presented an online tool for statistical analysis of the BV data based on checklist recommendation by Bartlett et al. (2015) [10], the pipeline by Braga and Panteghini (2016) [3] and the critical appraisal checklist by Aarsand et al. (2018) [11]. The tool is freely available through <https://turcosa.shinyapps.io/biovar/>, and all source code is available on the Github repository at <https://github.com/selcukkorkmaz/BioVar>. BioVar has been previously used as a descriptor by biological variation: information site for laboratory medicine (<http://www.biologicalvariation.com/>). Please note that we do not have any connection with this organization.

Research funding: None declared.

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

Conflict of interest statement: The authors have no conflict of interest.

References

- Castilla JA, Alvarez C, Aguilar J, Gonzalez-Varea C, Gonzalvo MC, et al. Influence of analytical and biological variation on the clinical interpretation of seminal parameters. *Hum Reprod* 2006;21:847–51.
- Aarsand AK, Roraas T, Sandberg S. Biological variation – reliable data is essential. *Clin Chem Lab Med* 2015;53:153–4.
- Braga F, Panteghini M. Generation of data on within-subject biological variation in laboratory medicine: An update. *Crit Rev Clin Lab Sci* 2016;53:313–25.
- Fraser C. *Biological Variation: from Principles to Practice*. Washington (DC): AACC Press; 2001.
- Franzini C. Relevance of analytical and biological variations to quality and interpretation of test results: examples of application to haematology. *Ann Ist Super Sanita* 1995;31:9–13.
- Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989;27:409–37.
- Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8–12.
- Petersen PH, Fraser CG. Strategies to set global analytical quality specifications in laboratory medicine: 10 years on from the Stockholm consensus conference. *Accredit Qual Assur* 2010;15:323–30.
- Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, 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.
- Bartlett WA, Braga F, Carobene A, Coskun A, Prusa R, et al. A checklist for critical appraisal of studies of biological variation. *Clin Chem Lab Med* 2015;53:879–85.
- Aarsand AK, Roraas T, Fernandez-Calle P, Ricos C, Diaz-Garzon J, et al. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. *Clin Chem* 2018;64:501–14.
- Kokoska S, Christopher N. *Statistical tables and formulae*. New York, NY: Springer; 198.
- Razali NM, Wah YB. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J Stat Model Anal* 2011;2:21–33.
- Sahai H, Ojeda MM. *Analysis of variance for random models, volume 2: unbalanced data: theory, methods, applications, and data analysis*. Berlin: Springer Science & Business Media; 2004.
- Burdick RK, Borror CM, Montgomery DC. *Design and analysis of gauge R&R studies: Making decisions with confidence intervals in random and mixed ANOVA models*. New Delhi: SIAM; 2005.
- Roraas T, Stove B, Petersen PH, Sandberg S. Biological variation: the effect of different distributions on estimated within-person variation and reference change values. *Clin Chem* 2016;62:725–36.
- Braga F, Ferraro S, Ieva F, Paganoni A, Panteghini M. A new robust statistical model for interpretation of differences in serial test results from an individual. *Clin Chem Lab Med* 2015;53:815–22.
- Kristoffersen AH, Petersen PH, Sandberg S. A model for calculating the within-subject biological variation and likelihood ratios for analytes with a time-dependent change in concentrations; exemplified with the use of D-dimer in suspected venous thromboembolism in healthy pregnant women. *Ann Clin Biochem* 2012;49:561–9.
- Aarsand AK, Diaz-Garzon J, Fernandez-Calle P, Guerra E, Locatelli M, 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.
- Coskun A, Carobene A, Kilercik M, Serteser M, Sandberg S, et al. Within-subject and between-subject biological variation estimates of 21 hematological parameters in 30 healthy subjects. *Clin Chem Lab Med* 2018;56:1309–18.
- Palomaki GE, Neveux LM. Using multiples of the median to normalize serum protein measurements. *Clin Chem Lab Med* 2001;39:1137–45.
- Snedecor GW, Cochran WG. *Statistical methods*, 8th ed. Ames: Iowa State Univ. Press Iowa; 1989.

23. Carobene A, Roraas T, Solvik UO, Sylte MS, Sandberg S, et al. Biological variation estimates obtained from 91 healthy study participants for 9 enzymes in serum. *Clin Chem* 2017;63:1141–50.
24. Fraser CG. Inherent biological variation and reference values. *Clin Chem Lab Med* 2004;42:758–64.
25. Harris EK. Statistical aspects of reference values in clinical pathology. *Prog Clin Pathol* 1981;8:45–66.
26. Fraser CG. Reference change values. *Clin Chem Lab Med* 2011;50: 807–12.
27. Fokkema MR, Herrmann Z, Musket FA, Moecks J. Reference change values for brain natriuretic peptides revisited. *Clin Chem* 2006;52:1602–3.
28. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP, et al. The use of error and uncertainty methods in the medical laboratory. *Clin Chem Lab Med* 2018;56:209–19.
29. Biswas SS, Bindra M, Jain V, Gokhale P. Evaluation of imprecision, bias and total error of clinical chemistry analysers. *Indian J Clin Biochem* 2015;30:104–8.
30. Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. *Clin Chem* 2011;57: 1334–6.
31. Roraas T. Estimating biological variation: methodological and statistical aspects. Bergen: University of Bergen; 2017.