Hikmet Can Çubukçu*, Anna Carobene, Aasne K. Aarsand, Sverre Sandberg and Abdurrahman Coşkun, on behalf of the EFLM Committee on Biological Variation and Technical Committee on Biological Variation Database

The BV calculator: a free, open-source web app for estimation of within-subject and between-subject biological variation

https://doi.org/10.1515/cclm-2026-0863
Received May 31, 2026; accepted July 8, 2026;
published online July 27, 2026

Abstract

Objectives: Biological variation (BV) data are essential for many purposes in laboratory medicine, however, the complex statistical analysis remains a barrier for delivery of BV data. This study describes the development of a web-based BV Calculator based on the Biological Variation Data Critical Appraisal Checklist (BIVAC).

Methods: We developed an open-source BV Calculator (https://bvcalculator.streamlit.app/) using Python 3.12, built on the 14 quality items (QI) included in the BIVAC. It requires manual user input with uploading of raw data and manual selection of QIs related to the study and study population, followed by automated preprocessing of the data with outlier detection, variance homogeneity assessment, steady-state verification through linear regression, normality assessment with Shapiro-Wilk tests, and log transformation when appropriate. The tool uses nested two-level ANOVA for BV component estimation and supports both balanced and unbalanced study designs and delivers CV_I and CV_G data with 95 % confidence intervals, accompanied by BIVAC criteria and audit logs. Real world data for creatinine, total cholesterol, S100B and creatine kinase from the European Biological Variation Study (EuBIVAS) were assessed, and results compared to previously published EuBIVAS data derived by manual processing.

Results: Central BV components, including mean concentrations, CV_A, CV_I, and CV_G estimates were similar to the manual EuBIVAS estimates for all measurands. Substantial differences in the number of subjects retained following automated outlier and trend analyses were identified for creatinine and S100B.

Conclusions: The BV Calculator is a freely available tool that can be used to facilitate BIVAC compliant statistical analysis of BV data.

Keywords: biological variation; BIVAC; reference change value; analytical performance specification; open-source software; STARBIV

***Corresponding author: Hikmet Can Çubukçu**, MD, MSc, PhD, EuSpLM, Department of Medical Biochemistry, Sincan Training and Research Hospital, Sincan, Türkiye, E-mail: hikmetcancubukcu@gmail.com. https://orcid.org/0000-0001-5321-9354

Anna Carobene, Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Milan, Italy

Aasne K. Aarsand, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and The Norwegian Porphyria Centre, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway

Sverre Sandberg, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; The Norwegian Porphyria Centre, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; and Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway

Abdurrahman Coşkun, Department of Medical Biochemistry, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye; and Department of Medicine, Knappschaft Kliniken University Hospital Bochum, Ruhr-University Bochum, Bochum, Germany,
E-mail: coskun2002@gmail.com. https://orcid.org/0000-0002-1273-0604

Introduction

Biological variation (BV) refers to the natural fluctuation in the concentration or activity of a measurand involved in human metabolism over time [1]. In laboratory medicine, the use of BV data is one model for establishing analytical performance specifications (APS) for imprecision, bias, and measurement uncertainty [2]. BV data can also be used to aid the interpretation of patient results, including the calculation of the reference change value (RCV), which determines whether the difference between two consecutive measurements can be attributed to expected analytical and biological variation, with a given probability. RCVs may also be useful for delta-checking and auto-verification algorithms [3]. BV data further enable decisions to be made as to the comparative utility

of population-based (popRIs) and personalized reference intervals (prRIs) to assist in interpretation of patients' laboratory results [4–6]. In this context, the Index of Individuality (II=within-subject BV (CV_I)/between-subject BV (CV_G)) is an indicator of the usefulness of using popRIs in an individual patient: II<0.6 favors using RCVs or prRIs over popRIs, whereas II>1.4 suggests popRIs may be informative [7].

Following the 1st EFLM Strategic Conference, the EFLM Committee on Biological Variation and Technical Committee on Biological Variation Database (previously named the Working Group on Biological Variation and the Task Group for the Biological Variation Database) developed the Biological Variation Data Critical Appraisal Checklist (BIVAC) a structured appraisal tool to evaluate the methodological quality of BV studies [8] aiming to support quality and consistency in BV research. The BIVAC consists of 14 quality items (QIs) that assess various aspects of a study, including its design (QI 1–3), measurement procedure (QI 4), preanalytical (QI 5) and estimation of analytical variation (QI 6), and statistical handling and reporting of data (QI 7–14) [8]. Each is, depending on the type of QI, scored on a descending scale from A to D, with the study's overall grade being decided by the lowest score received for any single QI. Crucially, BV data deriving from a study receiving a "D" grade are considered outdated or not fit for purpose and should not be used [8]. The BIVAC has later been included in the recently published Standard for Reporting Biological Variation Data Studies (STARBIV) [9], which is recognized by the Equator network [10] and which is available as a mind map on the EFLM Biological Variation Database website [2].

While the STARBIV and the BIVAC guide researchers through the proper and complete steps for conducting and reporting BV studies, the statistical assessment of data is time-consuming and requires expert knowledge. Estimation of BV components depends on complex statistical analysis such as outlier analysis at three levels, assessment of trends, normality and variance homogeneity to be fully BIVAC compliant. To simplify these steps, we developed an open-source, web-based application, the "BV Calculator" [11], which is based on the BIVAC and enables automated estimation of BV components, streamlining the statistical analysis. In this study, we describe the development and validation of the BV Calculator tool and present results for European Biological Variation Study (EuBIVAS) data sets with different characteristics, as compared to manual calculations of BV components in accordance with the BIVAC and EFLM recommendations.

Methods

Application development

The BV Calculator was developed using Python 3.12 [11, 12]. Data manipulation and tabular analysis were performed with Pandas 2.2.2 [13]. Core numerical computations – including the nested two-level analysis of variance (ANOVA) algebra of Røraas et al., Cochran's C test, Reed's outlier rule, Bartlett's tests for homogeneity, Shapiro–Wilk normality tests, linear regression, and calculation of confidence limits – were conducted using NumPy 1.26.4 [14] and SciPy 1.13.1 [15]. Interactive graphical representations were generated with Plotly 5.24.1 [16], and the web-based user interface was built with Streamlit 1.37.1 [17].

To allow users to assess the data calculations and to ensure transparency, the app performs essential preprocessing (outlier detection/removal, log transformation, etc.) when appropriate [18] and delivers a chronological audit log in line with BIVAC recommendations [8]. The entire workflow of the BV Calculator analysis is shown in Figure 1.

User input and data entry routes

The application provides two data-entry routes. Users may upload study data as comma-separated (.csv) or Excel (.xlsx) files through the "Upload" tab. Alternatively, a Manual Entry tab offers an editable spreadsheet interface within the application. Submitted data must contain the four variables detailed in Table 1. A downloadable Excel template with the correct structure is provided in the sidebar to reduce mapping errors. Importantly, the calculator depends on a study design applying replicate (usually duplicate) analysis, as recommended by the BIVAC.

After uploading data, users are prompted to map their columns to the required variables (Figure 2). As a first step, users enter BIVAC QI 1–5 information (scale, subjects, samples, measurand/method, and pre-analytical phase) via a structured form (Figure 3) based on their own information and self-assessment of their study. These user-entered QIs are stored together with the analysis outputs and are automatically incorporated into the exported checklist. In addition, users indicate whether replicates were analyzed in the same analytical run ("same-run duplicates"), a selection used in scoring and documenting QI 6 (Estimates of analytical variation). The interface also supports optional gender/sex stratification for data analysis: if enabled, the user will map a sex/gender variable which leads to reporting also of subgroup results (Figure 3).

Figure 1: BV calculator analysis workflow.

If the user selects the "Estimate CV_I with CV-ANOVA" option, CV_I is calculated using the coefficient-of-variation ANOVA method [19], otherwise a standard ANOVA method is used for the estimation [20].

Table 1: Mandatory data columns/variables to be entered by the user.

Variable	Meaning	Example entries
Subject	Unique participant identifier	1, subject 23
Sample	Ordered time-point label	1
Replicate	Replicate number within a sample	2
Result	Quantitative measurand value	6.4

Data analysis and statistical assessment

All datasets undergo an automated, structured preprocessing workflow implemented in *_preprocess_bv_dataframe*, designed to operationalize BIVAC QIs related to outlier exclusion, statistical integrity and reporting (QI 6–14) [8]. The pipeline returns: (i) a cleaned dataframe, and (ii) a detailed chronological audit log of the process (*preprocess_log*).

Each step is optional and can be enabled/disabled via sidebar switches (Figure 4). In addition, the application supports two operational modes for outlier handling: (1) auto mode, where outliers detected by statistical test explained in the following sections are removed, and (2) manual mode, where detected outliers are presented to the user for action (for subject and sample, "remove" vs. "ignore"; and for replicate-level outliers, in addition, "keep one replicate" may also be selected). For assessment of outliers prior to CV-ANOVA (sidebar selection choice), data are normalized, i.e. each result is divided by the subject mean,

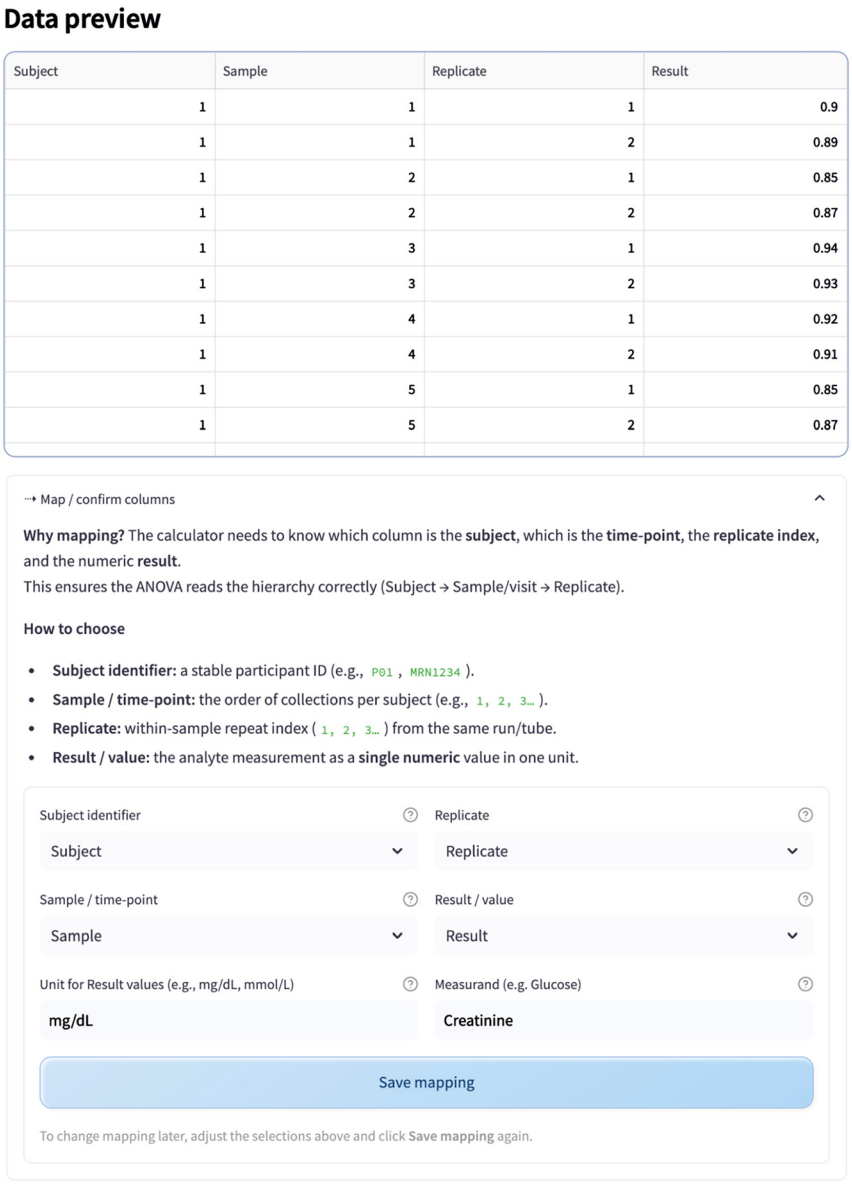

Figure 2: Data preview and mapping columns sections. This screenshot shows the upload workflow and the column-mapping dialog used to align user data to the four required variables (subject, sample, replicate, result). After mapping, users proceed to the preprocessing/analysis steps described in the Methods section.

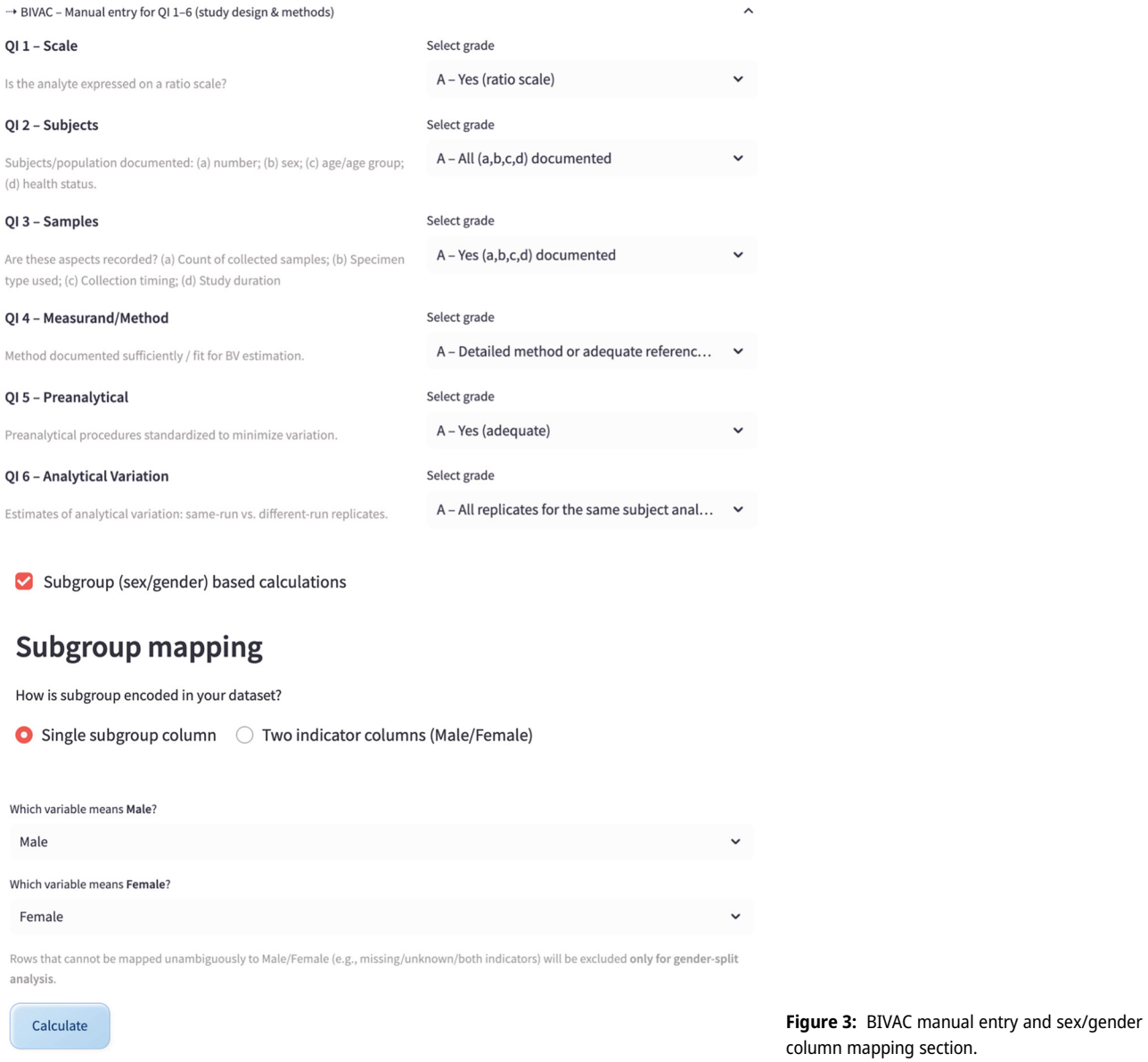

Figure 3: BIVAC manual entry and sex/gender column mapping section.

prior to analysis. User choices on outlier handling are stored and shown in the audit log. When all preprocessing checks are selected, the workflow proceeds in the following order:

Outliers in replicates (Cochran C test and Bartlett test), (QI 8)

For each sample from one individual, the variance of the replicate measurements is computed. Cochran's C test is then applied iteratively to identify outliers in all replicates. If an outlier is detected, the corresponding sample replicate pair is flagged.

In auto mode, the flagged sample pair is removed and the test repeats iteratively until no further Cochran outliers are flagged. In manual mode, users may choose to remove all results for the sample, ignore, or retain a single replicate (i.e. dropping the others). All actions and test statistics (G and G_{crit}) are recorded in the audit log.

Furthermore, to assess whether analytical imprecision is consistent across replicate sets, the application optionally applies a Bartlett test across replicate-level groups (each group is the replicate results belonging to a Subject × Sample pair). If heterogeneity is detected ($p<0.05$), the app iteratively flags and removes the highest-variance replicate pair, repeating the test until homogeneity is achieved. This step

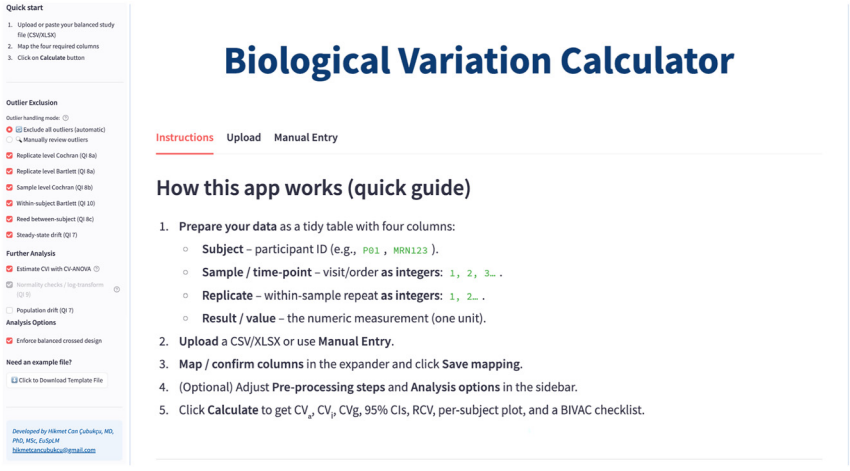

Figure 4: User interface of BV calculator. The interface presents (i) an instruction panel with a three-step quick start (upload/map/calculate); (ii) a checklist of BIVAC-aligned preprocessing options (e.g., replicate, within-subject and between-outlier checks QI 8, steady-state assessment and population drift QI 7, normality/log-transform QI 9, within-subject homogeneity QI 10); and (iii) analysis options (balanced vs. unbalanced design; optional CV-ANOVA for CV_I). These selections feed a chronological audit log, supporting transparent appraisal and re-runs.

can be user-guided in manual mode; and all removals/retentions are logged.

Outliers in samples per subject (Cochran's C test) (QI 8)

Within each subject, sample-wise replicate variances are evaluated, and Cochran's C applied to identify high-variance samples. In auto mode, detected outlier samples are removed (removing the full Subject × Sample pair) and detection continues. In manual mode, the user may remove or ignore each flagged sample, with details logged.

Variance homogeneity (Bartlett test) (QI 10)

Variance homogeneity across subjects is assessed using Bartlett's test on within-subject variances. Subjects exhibiting significantly heterogeneous variances compared to others are identified and excluded in auto mode. In manual mode subjects can be removed or ignored based on user choice. All exclusions and statistics are logged.

Outliers between-subjects (Reed test) (QI 8)

Between-subject outliers are screened using Reed's criterion, and any excluded subject is recorded in the log. The subject means are sorted and the overall range – the difference between the largest and smallest mean – is calculated. The most extreme subject mean, at either the high or low end, is flagged as a potential outlier when the gap between it and its nearest neighbouring mean exceeds one third of this range. Any flagged subject is excluded, and the test is repeated on the remaining subjects, with the range recalculated each time, until no extreme mean exceeds the threshold (a minimum of three subjects is required). In manual review mode, each detected outlier is presented with a radio button offering "Remove" or "Ignore" per subject. Upon user's request the app conducts outlier exclusion.

Steady-state (QI 7)

To identify subjects with significant temporal trends that should be investigated further, the app performs subject-level drift testing for subjects with a minimum of three samples. For each subject, replicate measurements are averaged per sample, and linear regression is fitted for result vs. sample order. The slope, its p-value, and a 95 % confidence interval (CI) for the slope are determined. If $p<0.05$, the subject is excluded in auto mode as well as presented to the user for selection in manual mode, with full details recorded in the audit log. Users must always assess that subjects have been excluded and why, to ensure representability of the reported BV components.

Population-level trend analysis (descriptive)

To assess cohort-level drift, the BV Calculator computes a pooled within-subject slope using a subject fixed-effects "within" estimator. Replicates are first averaged per subject and sample to minimize analytical noise. Data are then demeaned by subtracting each subject's average from their individual values, removing baseline differences. Ordinary least squares regression on the demeaned data yields the common slope, CIs and p-values. This check is descriptive only and does not automatically exclude subjects from analysis.

Normality assessment and transformation (QI 9)

By default, the application assesses normality in two tiers and, where indicated, applies a reversible natural-log

transformation. First, for every subject with ≥3 observations, the Shapiro–Wilk test is applied to the raw results, and the proportion of subjects consistent with a Gaussian distribution ($p>0.05$) is recorded in the audit log. As a pragmatic default, if this proportion is ≤50 %, all results are natural log transformed. Log-transformation of skewed BV data prior to variance estimation is a recognised approach [18]; the specific 50 % proportion used here is a screening threshold adopted for the application rather than a value prescribed by a formal statistical rule. Second, when≥3 subjects are available, the distribution of subject means is tested with Shapiro–Wilk; a significant result triggers a one-time log transformation (if not already applied) and immediate retesting. If normality is still not achieved after transformation, the application does not stop; it proceeds on the log scale and records a caution in the audit log indicating that the estimates should be interpreted carefully. When a transformation is used, values are exponentiated back to their original units before analysis, so all subsequent values and plots remain on the "native" scale.

Because normality testing at the individual-subject level is inherently uncertain when only a few samples are available per subject, this subject-level screen is deliberately paired with the assessment of the distribution of subject means, and the automated decision is intended to be reviewed by the user rather than applied blindly. The automated step can be switched off entirely using the "Normality checks/log-transform (QI 9)" option in the sidebar; when it is disabled, no transformation is applied, the omission is recorded in the audit log, and the user is free to inspect the data distributions visually and to decide on, and justify, any transformation on the basis of expert judgment. Conversely, if the user selects CV_I estimation based on CV-ANOVA, the application performs normalization by default, followed by the outlier handling process described previously.

Achieving a dataset with balanced design (analysis option)

By default, "enforce balanced crossed design" is enabled. The app keeps only subjects with the modal number of samples and removes any Subject–Sample pairs whose replicate count differs from the modal replicate number, logging all actions. Final retained counts of Subjects x Samples x Replicates are provided in the logs.

If the option is disabled, the app proceeds with an unbalanced analysis, retaining all the data. The number of subjects and the mean samples per subject and replicates per sample are logged. In the full pre-processing log, all details related to the data analysis and processing, including what data have been included and excluded, are displayed.

Statistical calculations

Following uploading of data and preprocessing steps, BV parameters are calculated using the *calculate_bv function*.

Balanced workflow

The function first computes the overall (grand) mean of all remaining results. It then partitions total variability into three parts: analytical (from replicates), within-subject (sample-to-sample for the same person), and between-subject. From these, it reports the corresponding coefficients of variation (CV_A, CV_I, CV_G), estimated by standard ANOVA. 95 % CIs are provided for CV_A and CV_I using chi-square limits and for CV_G using a log-scale ANOVA with exact limits back-transformed to percent [21]. A 95 % CI for the grand mean is also given. Subsequently, the app reports the one-sided 95 % reference-change value ($RCV_{(95\,\%)}$) using a lognormal, asymmetric formulation that yields distinct limits for decreases ($RCV_{(95\,\%)}\downarrow$) and increases ($RCV_{(95\,\%)}\uparrow$), reflecting the combined impact of analytical and within-subject variation, at a given probability [22]. The RCVs reported by the BV calculator represent the study population from which the uploaded data set has been derived. However, in clinical practice, an RCV can be applied to an individual patient only if the inherent within-subject biological variation of the target population is the same as that of the reference population used to derive it, and with a CV_A estimate reflecting the analytical imprecision of the method applied to the target population for a relevant time period, and derived with a relevant probability level. For these reasons, the suggested values should not be transferred directly to clinical practice (for RCV calculations intended for clinical use and with different probabilities, see the EFLM Biological Variation Database [2]). If the user selects the "Estimate CV_I with CV-ANOVA" option from the sidebar panel, data are additionally analyzed using the coefficient-of-variation ANOVA method for the estimation of CV_I, as described by Røraas et al. [19].

Unbalanced workflow

When balance enforcement is turned off, the app uses method-of-moments approach that accommodate unequal numbers of samples and replicates across subjects. The same outputs are produced – CV_A, CV_I, CV_G with 95 % intervals; a 95 % CI for the mean; and the 95 % RCV – and CV_G is again

obtained on the log scale using an unbalanced variant. CV_I by CV-ANOVA is also available in this mode.

All results are reported on the original measurement scale. The audit log records which analysis branch (balanced vs. unbalanced) was used and summarizes all derived quantities.

APS derivation and reporting

APS for imprecision, bias, and expanded measurement uncertainty (MAU) are derived from BV estimates and reported at three quality levels – minimum, desirable, and optimal – as previously described [23] for both full population and sex/gender subgroups if subgroup analysis is selected.

Graphical output

Exploratory visualization is performed by *plot_subject_ranges* function, which produces an interactive Plotly figure in which each included subject appears as a blue marker representing the mean concentration and with a thin grey whisker spanning the individual's minimum and maximum values.

BIVAC generation

The *_build_qi_checklist* function in the code parses the audit log together with analysis outputs to populate QI 7–14 automatically. QI 1–6 are taken from the form completed by the user/user input. Each item receives a score with a varying set of options from A–D as described by the BIVAC [8]; the overall grade represents as per BIVAC recommendation the lowest score achieved for any QI. The checklist can be exported as an .xlsx file.

Subgroup calculations

If the user selects subgroup analysis (sex/gender), the entire analysis is stratified, and results are presented for both the overall population and subgroups. Furthermore, the system provides specific reporting recommendations based on the comparison of 95 % CIs. Specifically, if the 95 % CIs for CV_I overlap between subgroups, it may be considered appropriate to report a single combined value. Similarly, if the 95 % CIs for mean concentrations overlap indicating similar population subgroup set-points, subgroup-specific reporting for CV_G may not be relevant.

Code and application availability

The entire application code – encompassing data preprocessing, statistical analysis, plot visualization, and checklist generation – is consolidated into a single open-source script (*bv_streamlit_v3. py*) shared at https://github.com/hikmetc/bv_calculator. The BV Calculator is also freely accessible at https://bvcalculator.streamlit.app/ [11]. An instruction describing the application's functionality and step-by-step usage is provided on the application website and will be updated as the tool evolves.

Data used for demonstration

To illustrate the full workflow, we prepared and analyzed a synthetic creatinine BV study dataset containing 48 subjects, each with 10 serial samples measured in duplicate (48 × 10 × 2 raw results). The code used to generate synthetic data is available at https://github.com/hikmetc/bv_calculator/blob/main/sythetic_data_generation/synth_bv_pipeline.py. The simulated creatinine data were used solely for demonstration of the BV Calculator's workflow and reporting; no validation inferences were drawn from this dataset. The data were uploaded without manual editing, the four required columns were mapped, and calculation was initiated.

Evaluation of BV calculator using real-world data

To evaluate the concordance between BV estimates derived from the newly developed BV Calculator and those previously published from the large scale, highly powered EuBIVAS [24, 25], a selection of measurands was chosen to represent most commonly used measurands and also representing different challenges in calculating BV data. Creatinine was selected as a representative measurand with data from two different analytical principles (enzymatic and Jaffe methods), enabling assessment across methods with distinct analytical characteristics [26]. Total cholesterol was included as a representative lipid measurand [27], creatine kinase (CK) as a representative enzyme [28], and S100B as an example of a protein analyte [29]. CK was intentionally selected because it presented a relatively high number of outliers in the original EuBIVAS dataset, allowing evaluation of the BV Calculator under conditions of increased analytical and biological variability.

BV estimates reported in the original EuBIVAS publications were compared with those obtained using the BV

Calculator under two conditions: with and without application of steady state assessment (subject-specific trend analysis). Comparisons were performed for CV_A, CV_I, and CV_G coefficients of variation, reported for the overall cohort and stratified by sex when applicable.

Results

Demonstration of the BV calculator using simulated data

During automated outlier detection, at replicate-level (BIVAC QI 8) the application reported no extreme replicate variances as assessed by Cochran's test, and Bartlett's test confirmed homogeneous replicate variances. Subgroup analysis showed higher mean creatinine concentrations in males (1.03 mg/dL, 95 % CI 1.01–1.04) than in females (0.90 mg/dL, 95 % CI 0.89–0.92). The 95 % CIs for CV_I and CV_G overlapped between subgroups (male CV_I CV-ANOVA 3.89 % [3.48–4.41] vs. female 4.61 % [4.14–5.18]; male CV_G 14.04 % [10.55–20.93] vs. female 15.67 % [11.92–22.84]) (Figure 5). Assessment for outliers in samples within subjects (BIVAC QI 8) identified one sample – Subject 16, Sample 10 (replicates 0.68 and 0.60) – which was automatically removed, as shown in the audit log(Supplemental Data). Because this exclusion left subject 16 with 9/10 required samples, the subject was excluded to preserve a fully crossed balanced design. Assessment of between-subject outliers (BIVAC QI 8) by Reed's criterion detected no outlying subject means. Assessment of steady-state by time-trend analysis (BIVAC QI 7) indicated significant drift in six participants (subjects 1, 8, 28, 34, 40, and 41), who were excluded from subsequent automated analyses. A pooled population-level estimate of drift (fixed-effects within estimator, pre-balance) was not significant (slope 0.00038, 95 % CI −0.00102 to 0.00179; p=0.594). Assessment of normality (BIVAC QI 9) showed that 37 of 42 retained subjects (88 %) met the Shapiro–Wilk criterion on the raw scale, and the distribution of subject means did not deviate from normality (Shapiro–Wilk p=0.638). To respect distributional assumptions at the group level, CV_G was estimated on the natural-log scale and back-transformed. Assessment of variance homogeneity across subjects (BIVAC QI 10) indicated homogeneous within-subject variances (Bartlett p=0.182). Thus, no additional exclusions were applied at this step. The resulting balanced crossed dataset comprised 41 subjects × 10 samples × 2 replicates and was used for ANOVA-based estimation (Figure 5).

With the final dataset, key overall metrics – including the overall mean, CV_A, CV_I (CV-ANOVA and standard-ANOVA), CV_G, and the lognormal RCV (95 %) – are summarized in Figure 6. Based on the CV_I and CV_G estimates, the resulting APS for imprecision (CV_A), bias, and MAU (k=2) at the minimum, desirable, and optimal levels are presented in Figure 6. All preprocessing actions are shown in the on-screen preprocessing log provided in the Supplementary Material. Final results are presented as a "Key metrics" panel, a "Summary of variation metrics", an APS table, and per-subject distribution plots (Figure 6). BIVAC scores are documented in an automatically generated list (which for the creatinine test set included overall grade A) (Figure 7), downloadable as a spreadsheet file.

Concordance between the EuBIVAS publications and the BV calculator output

In Table 2, BV estimates together with the number of subjects and samples effectively retained for each scenario are shown, reflecting the datasets remaining after exclusion of analytical outliers and, where applicable, exclusion of subjects with lack of steady state/significant temporal trends as applied in the BV calculator. Across the selected measurands, CV_I and CV_G estimates demonstrated a high degree of concordance between the EuBIVAS publications and the BV Calculator outputs (Table 2). For the overall cohort, all CV_I and CV_G estimates showed overlapping 95 % CIs across all three pairwise comparisons (EuBIVAS, BV Calculator without trend analysis, and BV Calculator with trend analysis). Similarly, complete overlap of 95 % CI for both CV_I and CV_G was observed for all evaluated measurands in males and in females, where sex-specific comparisons were applicable.

CV_A estimates also showed agreement across the three different approaches. For most measurands and population strata, CV_A values and their 95 % CI overlapped between the EuBIVAS publications and both BV Calculator outputs, with only minor numerical differences observed following application of trend analysis.

Application of subject-specific trend analysis resulted in exclusion of a variable proportion of subjects, depending on the measurand. Specifically, the BV calculator trend analysis led to exclusion of 33 out of 91 subjects (36 %) for enzymatic creatinine, 23 out of 91 subjects (25 %) for Jaffe creatinine, 24 out of 91 subjects (26 %) for S100B, 38 out of 90 subjects (42 %) for total cholesterol, and 45 out of 84 subjects (54 %) for CK. These exclusions were accompanied by proportional reductions in the number of samples contributing to BV estimation.

Subgroup	Mean concentration (95% CI)	CV_A % (95% CI)	CV_I (based on standard ANOVA) (95% CI)	CV_I (based on CV-ANOVA) (95% CI)	CV_G % (95% CI)	RCV (95%)
Male	1.03(1.01–1.04)	1.82 (1.65–2.02)	3.76 (3.36–4.26)	—	14.04 (10.55–20.93)	−9.23% / +10.17%
Female	0.90(0.89–0.92)	2.03 (1.85–2.24)	4.48 (4.03–5.03)	—	15.67 (11.92–22.84)	−10.77% / +12.07%

Subgroup-based reporting recommendations

CVI 95% CIs **overlap** between subgroups → separate CVI reporting for the subgroups may not be required.

✅ **Subgroup-based recommendation (CVG):** Mean concentration 95% CIs **do NOT overlap** between subgroups. **This may indicate that subgroup (Male and Female) CVG estimates may be more appropriate for use.**

Subgroup-stratified Per-subject Distribution

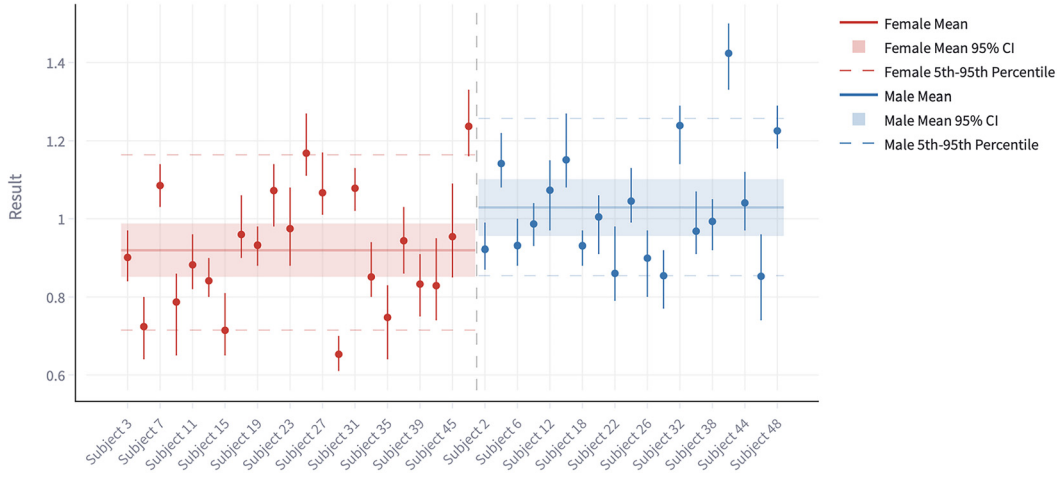

Figure 5: Sex/gender-based reporting recommendations.

For CK, a marked reduction in the number of included subjects was already evident in the absence of trend analysis. Compared with the original EuBIVAS cohort (84 subjects), BV estimation without trend analysis included 47 subjects, corresponding to exclusion of 37 subjects (44 %), primarily due to analytical outlier exclusion. Subsequent application of trend analysis further reduced the CK dataset to 39 subjects, corresponding to an additional exclusion of 8 subjects (17 %) relative to the no-trend scenario.

Discussion

In the present study, we describe the features of the BV calculator, which has been designed to facilitate the delivery of high-quality BV data by automating the statistical approach in line with the BIVAC and STARBIV. To assess the performance of the BV calculator, we have compared the results delivered by the calculator with previously published EuBIVAS results derived from manual analysis for

Study design

Parameter	Value
Number of subjects (I)	36
Samples per subject (S)	10
Replicates per sample (R)	2

Core estimates

Parameter	Value	CI (95%)
Mean concentration mg/dL	0.98 mg/dL	0.97–1.00 mg/dL
Analytical CV (%)	1.87 %	1.74–2.02 %
Within-subject CV (based on standard ANOVA) (%)	3.75 %	3.45–4.10 %
Within-subject CV (based on CV-ANOVA) (%)	3.89 %	3.58–4.25 %
Between-subject CV (%)	16.99 %	13.73–22.30 %
Reference change value (95%) (lognormal)	−9.26 % / +10.20 %	

Analytical Performance Specifications (APS)

Overall Male Female

Specification	CV_A	Bias	MAU (k=2)
Minimum	2.9	6.5	5.8
Desirable	1.9	4.4	3.9
Optimal	1.0	2.2	1.9

CVI used for APS: **CV-ANOVA** → 3.89% (95% CI 3.58–4.25%); MAU: Maximum Allowable Expanded Uncertainty.

Per-subject distribution (overall)

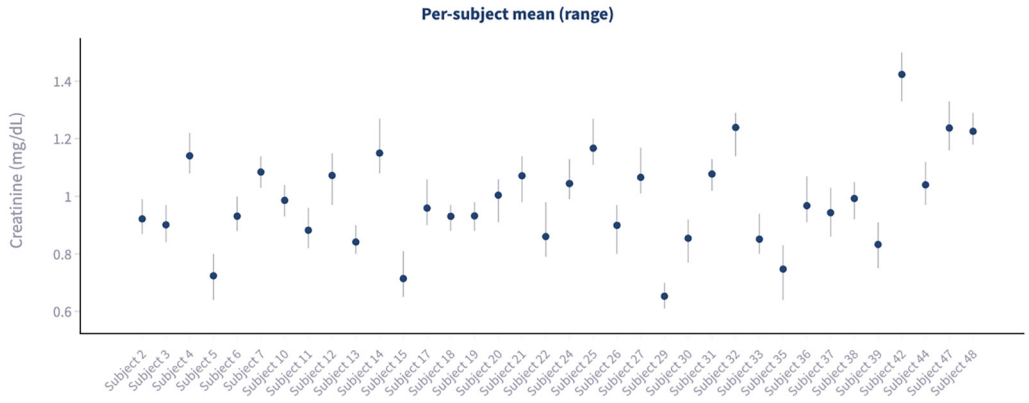

Figure 6: Key metrics tables, analytical performance specifications (APS) table, and per-subject distribution plot. The key metrics table reports study design (included subjects, samples per subjects, replicates per sample) and core estimates (the mean (with 95 % CI), CV_A, CV_I, CV_G with 95 % CIs, and the asymmetric log-normal RCV). The APS table summarizes BV-based APS at minimum/desirable/optimal levels. The per-subject plot shows each subject's mean (marker) and range (whisker = min–max), allowing quick visual assessment of within-subject dispersion.

QI	Score	Comment	Details
QI 1	A	User-selected A.	Manual: Scale → A.
QI 2	A	User-selected A.	Manual: Subjects → A.
QI 3	A	User-selected A.	Manual: Samples → A.
QI 4	A	User-selected A.	Manual: Measurand/Method → A.
QI 5	A	User-selected A.	Manual: Preanalytical → A.
QI 6	A	Estimates presented (2 replicates/sample); same run analysis.	User selected: Grade A. ▸ CV_a 1.95% (95% CI 1.82–2.09%).
QI 7	A	Individual trend analysis performed; drifting subjects removed.	QI 7 – Drift outlier removed (user selection) → Subject 1 slope=-0.00603 (95% CI -0.01195 to -0.0001112, p=0.04627) ▸
QI 8	A	Outlier testing performed: (a) Replicates, (b) Samples, (c) Subjects.	QI 8a – Cochran (replicate sets): G = 0.029, Gcrit = 0.031 → no outlier detected ▸ QI 8b – sample Cochran outlier remo
QI 9	A	Distribution assessed for normality.	Normality check (QI 9) – 37/42 subjects (88%) passed Shapiro-Wilk p>0.05 on raw results. ▸ Normality check (QI 9) –
QI 10	A	Variance homogeneity examined; variances are homogeneous.	QI 8a – Bartlett test (replicate sets): χ^2 = 214.70, χ^2crit = 468.78, p = 1.0000 → variances homogeneous. ▸ QI 10 – Bartle
QI 11	A	Nested ANOVA with variance decomposition (balanced design).	Balanced crossed design; closed-form variance decomposition for CV_a, CV_i, CVg.
QI 12	A	95% confidence limits for CVI presented.	CIs: CV_a 1.82–2.09%, CV_i 3.78–4.44%, CVg 13.49–21.23%.
QI 13	A	Results: 820 used / 960 raw (140 excluded).	Kept 820 of 960 measurements; 140 excluded by QC.
QI 14	A	Mean concentration: 0.97 (95% CI 0.96–0.98).	Mean = 0.97; 95% CI = 0.96–0.98.

Figure 7: BIVAC checklist output table. The auto-generated table lists item-level scores (A–D) for QI 1–14 and the overall grade (defined as the lowest item score). QI 1–5 derive from the user's study-design entries (including the "same-run duplicates" option relevant to QI 6), while QIs 6–14 are populated from the analysis.

creatinine, S100B, total cholesterol, and CK (Table 2), which represent different scenarios which must be considered when analysing BV data. Overall, the BV calculator delivers similar results as the manual EuBIVAS analysis (Table 2). Furthermore, despite substantial differences in the number of subjects retained following automated outlier and trend analyses, central BV estimates (mean, CV_A, CV_I, and CV_G) remained overlapping, indicating that e.g. individual trends are of less importance for the overall results for large-scale studies such as the EuBIVAS. It is, however, essential that automated BV tools are not used without assessing the data and with the BV calculator, the flagged results in the audit-log must be considered in detail. This must include visual inspection of data and data distributions and consideration of known biological factors that may influence results. In principle, automated BV estimation tools should be regarded as valuable complements to, rather than replacements for, informed methodological judgment, especially when dealing with heterogeneous populations, smaller data sets or measurands influenced by physiological factors.

Total cholesterol illustrates a scenario in which careful contextualisation of BV estimates is particularly important, as the BV calculator does not currently support stratification according to age or other subject characteristics. Age-related biological differences, especially in females around the menopausal transition, are well known and were explicitly addressed in the EuBIVAS study through age-stratified BV estimates. In the present comparison, BV estimates obtained using the BV Calculator remained overlapping with those derived from EuBIVAS at the aggregate level. Application of subject-specific trend analysis also resulted in exclusion of a substantial number of subjects. For creatinine, in the EuBIVAS, temporal changes were evaluated and, when necessary, corrected at the population level. This decision was based on extensive simulations showing that a large proportion of subjects (approximately 30–40 %) exhibited statistically significant individual trends, with positive and negative slopes occurring with similar frequency. Under such conditions, exclusion of subjects based on individual trend testing would result in substantial data loss without clear biological justification, which can be explained by the mathematical properties of regression-based trend testing. When analytical and within-subject biological variation are very low, as is the case for creatinine, statistically significant slope values may be observed even in the absence of biologically meaningful change. This effect was less pronounced for the Jaffe method, where higher analytical imprecision partially masked small temporal changes. Individual trend assessment may be more appropriate for measurands such as trace elements or tumor markers, where higher biological variability makes statistically significant trends more likely to reflect meaningful biological changes. This supports re-addressing individual trend analysis in the next version of BIVAC. Furthermore, this highlights the importance of assessing data that are flagged, and for the case of trends automatically excluded in standard mode, and to consider if they should be manually selected for inclusion to ensure representativeness of the data. For S100B, BV estimates were

Table 2: Biological variation estimates for selected measurands derived from EuBIVAS and from the BV calculator with and without application of trend analysis, reported for the full cohort and stratified by sex.

Measurand (unit)	Data source	Group	Number of subjects	Number of results	Mean value (95 % CI)	CV_A % (95 % CI)	CV_I % (95 % CI)	CV_G % (95 % CI)
Enzymatic creatinine, µmol/L	EuBIVAS paper	All	91 (0)	1,716 (0)	70.7 (70.0–71.4)	1.1 (1.05–1.2)	4.4 (4.2–4.7)	17.1 (14.9–20.1)
		M	38 (0)	708 (0)	79.5 (78.5–80.5)		4.2 (4.0–4.7)	13.8 (11.1–17.8)
		F	53 (0)	1,008 (0)	64.5 (63.9–65.1)		4.6 (4.3–4.9)	13.4 (11.1–16.5)
	BV calc. no trend analysis	All	91 (0)	1,690 (26)	70.7 (70.0–71.4)	0.96 (0.91–1.02)	4.4 (4.2–4.7)	17.2 (15.0–20.3)
		M	38 (0)	694 (14)	79.8 (78.8–80.9)		4.4 (3.99–4.86)	14.0 (11.4–18.2)
		F	49 (4)	996 (12)	64.6 (63.9–65.3)		4.4 (4.06–4.82)	13.5 (11.2–16.9)
	BV calc. trend analysis applied	All	58 (33)	1,174 (542)	70.5 (69.6–71.5)	0.95 (0.89–1.04)	4.3 (4.0–4.7)	17.7 (14.9–21.8)
		M	27 (11)	516 (192)	79.5 (78.3–80.7)		4.4 (4.0–5.0)	13.8 (10.8–19.0)
		F	28 (25)	658 (350)	63.7 (62.8–64.6)		4.3 (4.0–4.7)	13.3 (10.5–18.3)
Jaffe creatinine, µmol/L	EuBIVAS paper	All	91 (0)	1,704 (12)	65.6 (64.9–66.4)	4.4 (4.2–4.6)	4.7 (4.4–5.1)	19.0 (16.6–22.4)
		M	38 (0)	704 (4)	75.0 (73.8–76.1)		4.1 (3.5–4.8)	16.5 (13.2–21.2)
		F	53 (0)	1,002 (6)	59.0 (58.4–59.6)		5.0 (4.6–5.5)	13.9 (11.6–17.3)
	BV calc. no trend analysis	All	90 (1)	1,648 (68)	65.8 (65.2–66.5)	4.3 (4.1–4.5)	4.7 (4.5–5.1)	19.3 (16.7–22.7)
		M	37 (1)	666 (42)	74.0 (73.2–74.9)		4.4 (4.0–4.9)	14.9 (12.0–19.4)
		F	52 (1)	976 (32)	59.2 (58.6–59.8)		5.0 (4.6–5.4)	14.1 (11.8–17.6)
	BV calc. trend analysis applied	All	68 (23)	1,259 (457)	66.2 (65.4–67.0)	3.4 (3.1–3.7)	4.5 (4.2–4.9)	20.3 (17.3–24.5)
		M	29 (9)	504 (204)	74.2 (73.1–75.3)		4.4 (3.9–4.9)	16.6 (13.1–22.7)
		F	38 (15)	731 (277)	59.1 (58.5–59.8)		4.6 (4.2–5.1)	14.4 (11.7–18.7)
S100-b, mg/dL	EuBIVAS paper	All	91 (0)	1,728 (23)	44.5 (43.7–45.4)	4.3 (4.1–4.6)	10.2 (9.6–10.7)	32.7 (28.1–38.6)
	BV calc. no trend analysis	All	91 (0)	1,734 (17)	44.8 (44.0–45.6)	4.4 (4.2–4.7)	10.6 (10.1–11.3)	33.5 (30.7–42.1)
	BV calc. trend analysis applied	All	67 (24)	1,276 (475)	44.5 (43.7–45.4)	4.5 (4.3–4.8)	10.5 (9.8–11.2)	34.4 (29.1–42.1)
Total cholesterol, mg/dL	EuBIVAS paper	All	90 (1)	1,640 (54)	191.0 (189.0–192.9)	0.98 (0.9–1.0)	5.2 (4.9–5.5)	17.4 (15.1–20.4)
		M	38 (0)	686 (12)	191.8 (188.7–192.6)		5.2 (4.8–5.6)	17.7 (14.6–23.2)
		F<50	43 (0)	804 (4)	176.7 (174.7–179.0)		6.1 (5.7–6.6)	13.4 (10.9–17.1)
		F>50	9 (1)	172 (22)	232.0 (225.0–235.8)		4.1 (3.5–4.9)	14.6 (9.8–28.5)
	BV calc. no trend analysis	All	90 (1)	1,663 (31)	187.9 (185.3–188.5)	0.96 (0.91–1.00)	5.3 (5.0–5.6)	17.5 (15.2–20.5)
		M	38 (0)	688 (10)	189.4 (186.9–191.0)		6.5 (5.1–7.1)	13.4 (10.9–17.1)
		F	52 (1)	975 (27)	186.6 (184.5–188.7)		6.1 (5.8–6.6)	16.9 (14.1–21.1)
	BV calc. trend analysis applied	All	52 (39)	993 (701)	179.0 (177.2 181.5)	1.03 (0.97–1.10)	5.8 (5.4–6.2)	16.8 (14.0–21.0)
		M	20 (18)	373 (325)	178.6 (175.2 182.1)		5.4 (4.9–6.1)	19.1 (14.5–28.3)
		F	32 (21)	618 (384)	179.6 (177.2 182.0)		6.0 (5.6–6.6)	15.6 (12.4–20.9)
CK, U/L	EuBIVAS paper	All	84 (7)	1,485 (236)	105.9 (103.4–108.3)	0.9 (0.9–1.0)	14.5 (13.8–15.4)	37.9 (32.8–45.8)
		M	35 (3)	624 (94)	134.5 (130.4–138.6)		16.0 (14.8–17.5)	31.5 (25.2–42.5)
		F	49 (4)	883 (134)	87.5 (85.2–89.8)		15.7 (14.6–16.8)	30.5 (24.8–38.4)
	BV calc. no trend analysis	All	47 (44)	904 (817)	110.7 (107.7–113.7)	0.9 (0.9–1.0)	15.1 (14.1–16.2)	37.7 (30.8–48.4)
		M	25 (13)	473 (245)	137.9 (133.1–142.6)		15.9 (14.5–17.6)	34.5 (26.5–49.6)
		F	24 (29)	419 (598)	88.8 (86.4–91.3)		15.6 (14.3–17.3)	28.4 (21.7–40.9)
	BV calc. trend analysis applied	All	39 (52)	701 (1,020)	105.3 (102.5–108.1)	1.0 (0.9–1.1)	14.0 (13.0–15.2)	34.5 (28.2–46.1)
		M	19 (19)	309 (409)	128.0 (124.0–132.0)		14.2 (12.8–15.9)	26.4 (19.6–40.2)
		F	21 (32)	371 (646)	86.6 (83.4–89.2)		14.3 (12.9–15.9)	29.3 (22.0–43.5)

evaluated only for the overall cohort, as no sex-related differences were observed in the EuBIVAS dataset. Application of subject-specific trend analysis led to exclusion of a proportion of subjects comparable to that observed for Jaffe creatinine. Despite the reduction in the number of included subjects following trend analysis, mean concentrations as well as CV_I and CV_G estimates for S100B remained fully overlapping.

The calculator applies chi-square limits (Burdick and Graybill method) for deriving 95 % CI for CV_I and for CV_G a log-scale ANOVA approach with chi-square limits and back-transformation to the original scale [21], whereas in the EuBIVAS publications and in the EFLM Biological Variation Database [2], 95 % CIs are approximated for the two-fold nested ANOVA model using the Burdick and Graybill method for unbalanced designs [30]. The two approaches deliver similar results for the large scale EuBIVAS data sets investigated in our study (Table 2), but for smaller or more heterogeneous data sets there may be differences, which must be considered when assessing the results.

CK is a measurand characterised by higher biological variability and a substantial number of outliers in the EuBIVAS database. Mean concentrations as well as CV_I and CV_G estimates for CK remained overlapping across all approaches, despite marked differences in the number of subjects retained for analysis. A notable finding for CK was the high number of subjects excluded as outliers, already evident in the absence of trend analysis. This exclusion primarily reflects the application of a Bartlett test at the individual subject level, which identifies subjects whose within-subject variance is statistically different from that of the population and leads to exclusion of the entire subject. As a consequence, a considerable proportion of subjects was removed before any evaluation of temporal trends. In the original EuBIVAS analysis, when the Bartlett test indicated excessive within-subject variability, the corresponding data were visually inspected at the individual profile level. Where appropriate, single aberrant samples were removed rather than excluding the entire subject, and the variance was subsequently re-evaluated in an iterative manner. Although this approach was highly demanding and required substantial manual effort, it allowed preservation of the dataset while maintaining control over analytical and biological consistency. The results observed for CK once again emphasizes the importance of data visualisation as a complement to automated statistical procedures. It is important to be aware that if too many data are excluded due heterogeneity of within-subject variances, the end results are not representative of the group.

Despite numerous BV studies having been published, grade A studies, indicating full compliance with BIVAC, are scarce; as shown by systematic reviews published by the EFLM Technical Committee Biological Variation Database for hematological parameters [31], glycemic markers [32], measurands related to kidney function [33], thyroid hormones [34], and tumor markers [35]. Typically, the BIVAC items related to statistical analysis are often the cause of the lower BIVAC grades. The BV Calculator tool lowers methodological barriers and supports the correct implementation of the statistical components of BIVAC-compliant BV studies and generates an auditable, itemized BIVAC with scores, thereby offering a comprehensive solution for high-quality BV studies. Earlier attempts to simplify BV data analysis – most notably the BioVar web tool published in 2020 [36] and the openly available R script reported in 2023 [37] – leaves important gaps that the new BV Calculator [11] seeks to fill. BioVar provides a point-and-click Shiny interface that conducts outlier detection, normality testing, steady-state verification, homogeneity assessment, nested ANOVA, and report generation [36]. Yet it leaves critical exclusion decisions – removing flagged samples, eliminating subjects with significant trends – to the user. Although the authors describe BioVar as “based on” BIVAC, the tool neither creates a structured checklist nor assigns quality scores [36]. The R script by van Winden et al. offers no graphical user interface, so users must be comfortable with R and its packages [37] and it also does not include assessment of BIVAC items.

However, even if the BV Calculator provides support for statistical quality, to achieve full BIVAC A classification, researchers must also adhere to methodological and pre-analytical requirements as defined in the EFLM recommendations [8]. Recently, the EFLM groups published the STARBIV, which provides an interactive mind-map tool to guide researchers through proper conducting and reporting of BV studies [9]. While STARBIV standardizes documentation [9], the BV Calculator [11] complements it by delivering the BV estimates. The tool performs BV component estimation in accordance with current BIVAC criteria. Alternative frameworks for estimating BV – such as Bayesian methods [38] and data mining approaches [39, 40] – exist and can be integrated in the future versions of the BV Calculator [11]. There is a clear need to validate and standardize the statistical tools used to estimate BV, ensuring they align with BIVAC recommendations. As a step in this direction, the BV Calculator [11] extends prior attempts [36, 37] by embedding BIVAC compliance, audit-ready reports, and user-friendly interface into one freely available web application. Furthermore, the BV calculator can be further improved and new recommendations for delivery of BV components implemented as they become available.

In conclusion, the BV Calculator is a freely available, user-friendly tool developed to facilitate BIVAC compliant statistical analysis of BV data. The Calculator should be used in combination with expert methodological judgment, especially when dealing with heterogeneous populations, smaller data sets or measurands influenced by physiological factors.

Research ethics: To validate the BV Calculator, we utilized published data from studies approved by the Institutional Ethical Review Board of San Raffaele Hospital (Milan, Italy; Protocol WG-BV #001, 50/INT 2014) and complying with the 2013 Helsinki Declaration.
Informed consent: Not applicable.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The authors state no conflict of interest.
Research funding: None declared.
Data availability: None declared.

References

1. Sandberg S, Carobene A, Bartlett B, Coskun A, Fernandez-Calle P, Jonker N, et al. Biological variation: recent development and future challenges. Clin Chem Lab Med 2023;61:741–50.
2. The EFLM biological variation database. Available from: https://biologicalvariation.eu/.
3. Fraser CG. Reference change values. Clin Chem Lab Med 2011;50:807–12.
4. Coskun A, Sandberg S, Unsal I, Serteser M, Aarsand AK. Personalized reference intervals: from theory to practice. Crit Rev Clin Lab Sci 2022;59:501–16.
5. Coşkun A, Sandberg S, Unsal I, Topcu DI, Aarsand AK. Reference intervals revisited: a novel model for population-based reference intervals, using a small sample size and biological variation data. Clin Chem 2024;70:1279–90.
6. Coşkun A, Sandberg S, Unsal I, Cavusoglu C, Serteser M, Kilercik M, et al. Personalized reference intervals in laboratory medicine: a new model based on within-subject biological variation. Clin Chem 2021;67:374–84.
7. Fraser CG. Inherent biological variation and reference values. Clin Chem Lab Med 2004;42:758–64.
8. 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.
9. Bartlett WA, Sandberg S, Carobene A, Fernandez-Calle P, Diaz-Garzon J, Coskun A, et al. A standard to report biological variation data studies - based on an expert opinion. Clin Chem Lab Med 2025;63:52–9.
10. EQUATOR Network. Enhancing the QUAlity and transparency of health research. Available from: https://resources.equator-network.org.
11. BV Calculator; 2025. Available from: https://bvcalculator.streamlit.app/.
12. Van Rossum G, Drake FL. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
13. Team TPD. pandas-dev/pandas: Pandas, 2020. Zenodo. Available from: https://www.doi.org/10.5281/zenodo.10957263.
14. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585:357–62.
15. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat Methods 2020;17:261–72.
16. Collaborative data science publisher: Plotly Technologies Inc; 2015. Available from: https://plot.ly [Accesed 21 June 2023].
17. Streamlit • A faster way to build and share data apps. Available from: https://streamlit.io/ [Accesed 12 April 2023].
18. 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.
19. Røraas T, Støve 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.
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. Burdick RK, Graybill FA. Confidence intervals on linear combinations of variance components in the unbalanced one-way classification. Technometrics 1984;26:131–6.
22. Fokkema MR, Herrmann Z, Muskiet FA, Moecks J. Reference change values for brain natriuretic peptides revisited. Clin Chem 2006;52:1602–3.
23. Sandberg S, Coskun A, Carobene A, Fernandez-Calle P, Diaz-Garzon J, Bartlett WA, et al. Analytical performance specifications based on biological variation data - considerations, strengths and limitations. Clin Chem Lab Med 2024;62:1483–9.
24. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, et al. Sample collections from healthy volunteers for biological variation estimates' update: a new project undertaken by the Working Group on Biological Variation established by the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2016;54:1599–608.
25. Carobene A, Aarsand AK, Bartlett WA, Coskun A, Diaz-Garzon J, Fernandez-Calle P, et al. The European Biological Variation Study (EuBIVAS): a summary report. Clin Chem Lab Med 2022;60:505–17.
26. 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.
27. 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.
28. Carobene A, Røraas T, Sølvik U, Sylte MS, Sandberg S, Guerra E, et al. Biological variation estimates obtained from 91 healthy study participants for 9 enzymes in serum. Clin Chem 2017;63:1141–50.
29. Carobene A, Guerra E, Locatelli M, Ceriotti F, Sandberg S, Fernandez-Calle P, et al. Providing correct estimates of biological variation-not an

easy task. The example of S100-β protein and neuron-specific enolase. Clin Chem 2018;64:1537–9.
30. Burdick RK, Graybill FA. Confidence intervals on variance components. CRC Press; 1992.
31. Coskun A, Braga F, Carobene A, Tejedor Ganduxe X, Aarsand AK, Fernández-Calle P, et al. Systematic review and meta-analysis of within-subject and between-subject biological variation estimates of 20 haematological parameters. Clin Chem Lab Med 2019;58:25–32.
32. Ricós C, Fernández-Calle P, Gonzalez-Lao E, Simón M, Díaz-Garzón J, Boned B, et al. Critical appraisal and meta-analysis of biological variation studies on glycosylated albumin, glucose and HbA(1c). Adv Lab Med 2020;1:20200029.
33. Jonker N, Aslan B, Boned B, Marqués-García F, Ricós C, Alvarez V, et al. Critical appraisal and meta-analysis of biological variation estimates for kidney related analytes. Clin Chem Lab Med 2022;60:469–78.
34. Fernández-Calle P, Díaz-Garzón J, Bartlett W, Sandberg S, Braga F, Beatriz B, et al. Biological variation estimates of thyroid related measurands - meta-analysis of BIVAC compliant studies. Clin Chem Lab Med 2022;60:483–93.
35. Marques-Garcia F, Boned B, González-Lao E, Braga F, Carobene A, Coskun A, et al. Critical review and meta-analysis of biological variation estimates for tumor markers. Clin Chem Lab Med 2022;60:494–504.
36. Korkmaz S, Zarasız G, Göksülük D, Senes M, Sönmez C, Yucel D. BioVar: an online biological variation analysis tool. Turk J Biochem 2020;45: 479–89.
37. van Winden LJ, van Rossum HH. An openly available R script for the estimation of biological variation based on EFLM guidelines. J Appl Lab Med 2023;8:218–20.
38. Røraas T, Sandberg S, Aarsand AK, Støve B. A bayesian approach to biological variation analysis. Clin Chem 2019;65:995–1005.
39. 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.
40. Marqués-García F, Nieto-Librero A, González-García N, Galindo-Villardón P, Martínez-Sánchez LM, Tejedor-Ganduxé X, et al. Within-subject biological variation estimates using an indirect data mining strategy. Spanish multicenter pilot study (BiVaBiDa). Clin Chem Lab Med 2022;60:1804–12.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2026-0863).