

Mini Review

Jorge Díaz-Garzón Marco*, Pilar Fernández-Calle and Carmen Ricós

Models to estimate biological variation components and interpretation of serial results: strengths and limitations

<https://doi.org/10.1515/almed-2020-0063>

Received February 9, 2020; accepted May 22, 2020; published online August 10, 2020

Abstract: Biological variation (BV) has multiple applications in a variety of fields of clinical laboratory. The use of BV in statistical modeling is twofold. On the one hand, some models are used for the generation of BV estimates (within- and between-subject variability). Other models are built based on BV in combination with other factors to establish ranges of normality that will help the clinician interpret serial results for the same subject. There are two types of statistical models for the calculation of BV estimates: A. Direct methods, prospective studies designed to calculate BV estimates; i. Classic model: developed by Harris and Fraser, revised by the Working Group on Biological Variation of the European Federation of Laboratory Medicine. ii. Mixed-effect models. iii. Bayesian model. B. Indirect methods, retrospective studies to derive BV estimates from large databases of results. Big data. Understanding the characteristics of these models is crucial as they determine their applicability in different settings and populations. Models for defining ranges that help in the interpretation of individual serial results include: A. Reference change value and B. Bayesian data network. In summary, this review provides an overview of the models used to define BV components and others for the follow-up of patients. These models should be exploited in the future to personalize and improve the information provided by

the clinical laboratory and get the best of the resources available.

Keywords: biological variation; methods; statistical design.

Introduction

The concept of biological variation (BV) was first formulated by Harris and Fraser in the mid-twentieth century [1]. Since then, multiple prospective experimental studies of different levels of complexity have been conducted to produce accurate BV estimates that could serve as reference values and be generalized to a diversity of populations, settings and conditions.

BV has two components: within-subject and between-subject variation. Within-subject variation is defined as the fluctuation of a measurand around its homeostatic setting point within the same subject, whereas between-subject BV is described as the variation between the homeostatic points of different subjects [1–3]. Both estimates are expressed as coefficients of variation and are referred to as CV_I and CV_G , respectively, in accordance with the standard terms recommended by Simundic et al. [4].

BV has a multiplicity of applications in the clinical laboratory, including internal control of the analytical process, external quality assurance programs, estimation of the reference change value (RCV) [1], verification and validation of analytical methods and as the criterion for establishing the maximum allowable deviation caused by biological sample instability [5]. BV is also used in interference studies and for establishing the limit of quantification of a measurand [6].

There are two ways to link the concept of BV to statistical modeling. On the one hand, some models are used for generation of BV estimates (within- and between-subject variation). On the other hand, other models are built based on BV in combination with other factors in order to establish intervals that help clinicians to interpret serial results in an individual.

*Corresponding author: Jorge Díaz-Garzón Marco, Comisión de Calidad Analítica, SEQC^{ML}, Barcelona, Spain; and Servicio de Análisis Clínicos, Hospital Universitario La Paz, Paseo de la Castellana, 261, Madrid, 28046, Spain, E-mail: jdgmarco@gmail.com

Pilar Fernández-Calle: Comisión de Calidad Analítica, SEQC^{ML}, Barcelona, Spain; Servicio Análisis Clínicos, Hospital Universitario La Paz, Madrid, Spain

Carmen Ricós: Comisión de Calidad Analítica, SEQC^{ML}, Barcelona, Spain

Models for the generation of BV estimates

There are two types of statistical models for the calculation of BV estimates:

- A. Direct methods: Prospective studies specifically designed to calculate BV estimates
 - i. classic model: developed by Harris and Fraser, revised by the Working Group on Biological Variation of the *European Federation of Laboratory Medicine* (EFLM) (EFLM-BVWG)
 - ii. mixed-effect models
 - iii. Bayesian model
- B. Indirect methods: Retrospective studies to derive BV estimates from large databases—*Big data*.

Understanding their characteristics of these models is crucial as they determine their applicability in different settings and populations.

A. Direct methods

i. Classic method

The classic method, developed by Fraser and Harris in the 60s, is still used in most studies aimed at the generation of BV estimates, fitting the model more or less strictly. The Analytical Quality Commission of Spanish Society of Laboratory Medicine (SEQC^{ML}) published in 1997 a database that compiled all studies published to date on BV, that was periodically upgraded until 2014 [7, 8]. This database has been a valuable tool for laboratory professionals.

The classic model described by Fraser and Harris, which requires that samples are assayed in duplicate, assumes three requirements: BV data must be normally distributed; the distribution of variables must be homoscedastic (homogeneity of variance across the whole interval of concentrations, which implies that the distribution of results is homogeneous between the replicates of the samples and between the samples of different subjects) and finally, subjects cannot show any trend throughout the study period (stable status). The statistical method recommended for the calculation of BV estimates consists of a nested analysis of variance (ANOVA) after a thorough search for outliers at three levels (between duplicates, within subjects and between subjects)

This approach is based on the concept that the total variation of all measurements, expressed as coefficient of

variation (CV_T) is the sum of preanalytical (CV_{PRE}), analytical (CV_A) and within-subject BV (CV_I). Once analytical variation is known and controlled, and in standard preanalytical conditions, this last component can be ruled out, thereby having only CV_I left in the equation. CV_I must be equal to the difference between total variation and analytical variation.

$$VT = CV_{PRE} + CV_A + CV_I \rightarrow CV_I = CV_T - CV_A$$

However, upon this model, for BV estimates to be robust, a set of requirements must be met. The number of subjects, samples per subjects and replicates per sample must be adequate. To such purpose, the ratio of analytical imprecision of the measurement method to the expected CV_I must be considered [9]. It is recommended that such a ratio should be <1 , and it would be desirable to be <0.5 [1].

The EFLM-BVWG, in cooperation with the *Task Group on Biological Variation Database*, recently published a review of studies using a BV data critical appraisal checklist designed by the EFLM-BVWG itself. Thus, studies were classified into four categories (A, B, C and D) in ascending order of quality. A meta-analysis was carried out to estimate the components of BV, CV_I , and CV_G and their 95% confidence intervals [10]. The new BV database was published on the EFLM website [11] in May 2019. Most of the studies published and included in the new database are adapted to the classic model. However, in most of these studies, CV_I estimates were obtained by simple subtraction of variances (subtracting the analytical variation from the total observed variation) instead of using a most appropriate statistical method such as ANOVA. As a result, these studies have a poorer methodological quality. Other limitations of these studies are related to the assumption of individual stability, the method used to calculate CV_A , the homoscedasticity of variances and the method for testing outliers [10, 12].

The most relevant advantage of this approach is that it has been extensively tested and validated, and it is widely used and known. It is a prospective method wherein pre-analytical variables are well controlled, which guarantees the robustness of estimates.

However, this method has some limitations:

- Some measurands may not meet homoscedasticity requirements.
- If the process of testing outliers is carried out with excessive zeal, so that even moderate outliers are eliminated, the transferability and statistical power of the study are limited. This is due to the manual processing of outliers even though some standard criteria

are fulfilled (Reed, Cochrane, etc.). Valuable results that are representative of the subjects may be eliminated.

- Most studies use a small sample of subjects to reduce costs and facilitate the study. In addition, the number of samples obtained per subject (visits) is not generally high as it compromises the participant's adherence to the study.
- The presence of significant trends that affect the stability of the subject (homeostasis) over the study period may compromise the generation of reliable estimates. If trends are observed in a certain subject, this subject could be considered as an outlier. If instead trends affect all results such as in the case of seasonal variations, corrections can be performed to minimize the effects of this factor. Unfortunately, no recommendations or clear criteria have been established in the literature to solve this problem.
- The possible presence of noncontrolled variables, as observed by the researcher, may cause bias in the obtained estimates.

ii. Mixed-effects models

The formula used in this method is the same as the one of the classic models as it is based on a nested ANOVA. However, the estimation is performed using a mixed model that includes further variables apart from homeostatic regulation. In this model, it is also necessary that results are normally distributed, and homogeneity of variance should be checked for BV estimates to be reliable [13]. The main difference with respect to the classic method is that it may include other variables that may affect the BV of the studied measurands. These variables might be included in the model, and the magnitude of their effect on the BV is measured (influence).

In general, this model includes two types of effects: fixed effects, which are variables with potential influence (age, sex, medication, health status, to name a few), and random effects, which are BV estimates and imprecision (CV_I , CV_G and CV_A).

The main advantage of this model relies on the ability to determine if BV is affected by any variable and to measure such an effect. On the other hand, if the variables potentially affecting BV are previously identified, outlier handling can be less rigorous [13].

This model has the same limitations as the classic model but adds the difficulty of requiring previous awareness of potentially influencing variables. In addition, this statistical method is more complex and requires advanced mathematical skills. For several variables to be included

simultaneously in this model, a very large sample size is required. Finally, an important limitation of this model is that it has not been fully validated in BV studies [13].

iii. Bayesian model

This model was recently developed and published by Røraas et al. [14] that is based on a Bayesian method that does not require the variances to be homogeneous. Thus, this model loosens strict outlier handling, thereby preventing that potentially valuable data for BV estimation are lost [9].

The fact that this model does not require a three-level outlier analysis and homoscedasticity simplifies and automates BV component estimation, which makes these estimates more reliable and, even more important, standardizable. In addition, this method can be used by researchers to estimate individualized CV_I , allowing to have the estimates expressed as percentiles (median and interquartile range for CV_I and CV_G), instead of yielding a central value as other models do (mean).

This model has been validated against traditional methods and has demonstrated to yield similar BV estimates. To perform these comparisons, Røraas et al. [14] used the data obtained in the EuBIVAS study for chloride and triglycerides [15]. In other words, this model has already been validated with real BV data from robust studies, which guarantees the validity of this approach.

However, the main limitation of this model is the level of mathematic and programming skills required to be able to operate the statistical software needed for its implementation. Another disadvantage is that this model requires the use of BV estimates from previous studies as *a priori* information (hyperparameters); therefore, if the *a priori* information entered in the model is not correct, the conclusions obtained may be incoherent or inconsistent with the hypotheses defined in the design of the study.

In summary, this type of model should only be used when solid statistical support is available. This model is extremely useful to study BV in nonhomoscedastic measurands that yield a high percentage of outliers that cannot be explained by the information recorded during the study.

B. Indirect methods – Big data

Although this method has gained popularity in recent years, few studies have been conducted, and consensus has not been reached on its optimal design (statistical method, inclusion criteria, among other factors).

This strategy is based on the assumption that most estimates are not affected by health status because abnormal results are excluded after an outlier analysis [16]. Some examples include the studies conducted by Loh et al. [17] and Loh and Metz [18] in the pediatric population assisted in primary care, with similar CV_I results to those obtained in healthy adults.

The most relevant advantages of this method are that it enables assessing differences by age, gender, study duration [16–18] and even between different diseases [19]. As the number of subjects included in this type of studies is larger than that in the prospective studies (up to several thousand because the method is based on the laboratory database), they have a higher degree of generalizability and statistical power. This method is economical as it requires limited human and material resources and does not require an experimental phase.

Some of its limitations are that preanalytical variables are not standardized; the homeostatic status of the subjects is unknown; CV_A is derived from the quality assurance results (stabilized control samples with a different matrix from patients' samples and different concentrations from those in patients' samples) instead of duplicate analysis measured in subjects' samples. These studies often have a low number of samples per subject and variable sampling intervals, which, in the case of some measurands, may limit the robustness of results. Some of these multicentric studies include results from different measurement methods and reagent lots.

Models for defining ranges that help in the interpretation of individual serial results

The following are the models for defining ranges that help in the interpretation of individual serial results:

- A. Reference change value
- B. Bayesian data network

A. Reference change value

The RCV defines an interval from which the difference between serial results from an individual could be considered biologically significant. The RCV is derived from the measurement of an analytical error and the fluctuation of a measurand within an individual. For such

purpose, the laboratory's CV_A and the CV_I are included in the calculation [20]. This concept is based on the assumption that all individuals have the same CV_I , and it requires robust and representative CV_I from the population to which this concept is applied for monitoring. The RCV is calculated using the following formula, with a Z probability of 1.96 and 1.65 for bidirectional and unidirectional changes, respectively [1]:

$$RCV = 2^{1/2} Z (CV_A^2 + CV_I^2)^{1/2}$$

In contrast with the classic model, which is based on the assumption that all measurands are normally distributed and the interval is symmetrical, the method currently recommended for the calculation of the RCV is based on a logarithmic method that defines an asymmetrical interval [21]. The formula is as follows:

$$RCV_{pos} = (\exp(1.96 \cdot 2^{1/2} \cdot \alpha) - 1) \cdot 100$$

$$RCV_{neg} = -(\exp(1.96 \cdot 2^{1/2} \cdot \alpha) - 1) \cdot 100$$

$$\alpha = (\ln(CV_T + 1))^{1/2}; \quad CV_T = CV_A + CV_I$$

The measurands that can be interpreted by this method are those that are subject to a greater homeostatic regulation, i.e., measurands with a low index of individuality ($II < 0.6$) as in the case of creatinine, electrolytes and some hematology parameters. This index relates within- and between-subject BV estimates ($II = CV_I/CV_G$) [1]. The serial results of measurands with a high individuality (low II) should be interpreted by the RCV as an alternative to biological reference intervals.

The main limitation of this concept is that it is unknown to most clinicians, and laboratory information management systems are not fully adapted to this model.

Another aspect to be considered is that preanalytical conditions have been standardized in most of the BV studies to avoid the influence of preanalytical factors and bias when calculating BV estimates. However, in a real setting, these variables are not always standardized by laboratories, and this source of variation is not considered when the range is defined. In this context, RCV could be too strict, and changes between serial results could be interpreted as biologically significant when they are not.

In addition, CV_A included in the equation is derived from the internal quality control. This implies that in measurands whose CV_A depends on the measurand concentration, a different CV_A should be applied for each concentration, based on the results obtained. This limitation hinders the application of the model.

B. Bayesian data network

Bayesian models make it possible to update and adjust the probability of obtaining a result within a range based on previous information. These models are flexible and improve progressively as further information is integrated. At baseline, previous information can be obtained from the literature or reliable sources [11].

These models have been developed by Sottas et al. [22] in the field of laboratory medicine and are applied by the World Anti-Doping Agency in the detection of illicit drug consumption. The method used in the follow-up of athletes is named biological passport. Instead of detecting the illicit substance in the body, indirect markers of doping are monitored (hemoglobin and reticulocytes, among others).

To predict the specific interval of variation of a measurand in a subject, the model incorporates previous information based on both result distribution in the specific population group to which the subject belongs (age, gender, ethnicity, disease) and individual results [23].

This model relies on a hierarchical Bayesian network, consisting of several levels, that is built from patient results and heterogeneity variables such as gender, age and other factors. In the first level, it separates the data into different distributions (groups) as a function of these heterogeneity variables. In the second level, this model uses BV estimates derived from these groups as *a priori* information (hyperparameters) [24].

Then, after including the subject results in the model, distributions that define a range predicting the individual result with a higher degree of robustness than in former levels are obtained.

In sum, this model defines specific reference intervals for a subject based on previous information. In other words, it is based on preliminary distributions and incorporates data that provide information to the model, which generates more robust distributions. In addition, if the volume of previous results included in the database for the population of study is high enough, previous specific BV studies would be not necessary (hyperparameters).

According to the literature, the number of patients or samples needed to meet the requirements of a clinical trial may decrease significantly, or the trial may be completed earlier if this model is used. For instance, Sottas et al [24] indicate that when population-based reference intervals are used, 600 subjects are needed to detect a change of 0.06 mg/dL in serum creatinine concentrations. In contrast, if an analysis of covariance with respect to basal values is performed, 210 subjects are needed, whereas 20

subjects are required if the Bayesian model is used. The use of this model in clinical practice may help clinicians in the interpretation of laboratory results [24].

The Bayesian network model assumes that there is no analytical error. To achieve that, all results must be obtained using the same analytical procedure, which guarantees the absence of bias and ensures low and fully controlled imprecision. However, many of the measurands routinely assayed in clinical laboratories for disease monitoring are not measured under these conditions. Indicators of analytical error obtained by internal quality assurance strategies could be incorporated into this model as hyperparameters to obtain a more robust prediction of ranges derived.

Another limitation is that these databases are not public, and they cannot be freely accessed by clinical laboratories.

In summary, this review aimed to give an overview of the models used to define BV components as well as others for patient monitoring. It should be exploited in the future to personalize and improve the information provided by the clinical laboratory, taking advantage from the available resources.

Research funding: None declared.

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

Competing interests: Authors state no conflict of interest.

References

1. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989; 27:409–37.
2. Harris EK, Kanofsky P, Shakarji G, Cotlove E. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. II. Estimating biological components of variation. *Clin Chem* 1970;16:1022–7.
3. Simón M, Corte Z, Variación BE, Ricós C, Álvarez V, Perich C, et al. Laboratorio clínico y calidad. Nuevas perspectivas 1ª ED. Barcelona: Fundació pel Control de la Qualitat del Laboratoris Clínics; 2017:589–614 p.
4. Simundic AM, Kackov S, Miler M, Fraser CG, Petersen PH. Terms and symbols used in studies on biological variation: the need for harmonization. *Clin Chem* 2015;61:438–9.
5. Gómez-Rioja R, Segovia Amaro M, Díaz-Garzón J, Bauçà JM, Martínez Espartosa D, Fernández-Calle P. A protocol for testing the stability of biochemical analytes. Technical document. *Clin Chem Lab Med* 2019;57:1829–36.
6. CLSI. Evaluation of detection capability for clinical laboratory measurement procedures. Approved guideline-second edition. CLSI document EP17–A2. Wayne, PA: Clinical and Laboratory Standard Institute; 2012.

7. Minchinela J, Ricós C, Perich C, Fernández-Calle P, Alvarez V, Doménech MV, et al. Base de datos de los componentes de Variación Biológica, con las especificaciones de la calidad analítica (deseable, mínima y óptima). Actualización del año 2014. Available from: https://www.seqc.es/docs/Comisiones/Calidad_Analitica/Anexo_I-Especificaciones_de_calidad_DESEABLES_2014.pdf [Accessed 31 Jan 2020].
8. Minchinela J, Ricós C, Perich C, Fernández-Calle P, Álvarez V, Doménech MV, et al. Biological variation database and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update. Available from: [Accessed https://www.westgard.com/biodatabase-2014-update.htm](https://www.westgard.com/biodatabase-2014-update.htm) [16 Jan 2020].
9. 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.
10. Aarsand AK, Røraas T, Fernandez-Calle P, Ricós C, Díaz-Garzón J, Jonker N, et al. European federation of clinical chemistry and laboratory medicine working group on biological variation and Task and finish group for the biological variation database. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. *Clin Chem* 2018;64: 501–14.
11. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM biological variation database. Available from: <https://biologicalvariation.eu/> [Fecha de consulta 31 Ene 2020].
12. Ricós C, Fernández-Calle P, González-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 HbA1c. Available from: <https://doi.org/10.1515/almed-2020-0029> [Accessed 4 May 2020].
13. Díaz-Garzón J. Variación Biológica de magnitudes bioquímicas, hematológicas y gasometría en atletas. Tesis doctoral. Universidad Autónoma de Madrid. Fac Med 2019. Available from: <https://www.educacion.gob.es/teseo/irGestionarConsulta.do> [Accessed 3 May 2020].
14. Røraas T, Sandberg S, Aarsand AK, Støve B. A Bayesian approach to biological variation analysis. *Clin Chem* 2019;65: 995–1005.
15. 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.
16. 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.
17. Loh TP, Ranieri E, Metz MP. Derivation of pediatric within-individual biological variation by indirect sampling method: an LMS approach. *Am J Clin Pathol* 2014;142:657–63.
18. Loh TP, Metz MP. Indirect estimation of pediatric between-individual biological variation data for 22 common serum biochemistries. *Am J Clin Pathol* 2015;143:683–93.
19. Sottas PE, Kapke GF, Vesterqvist O, Leroux JM. Patient-specific measures of a biomarker for the generation of individual reference intervals: hemoglobin as example. *Transl Res* 2011;158: 360–8.
20. Lund F, Hyltoft Petersen P, Fraser CG. Calculation of limits for significant bidirectional changes in two or more serial results of a biomarker based on a computer simulation model. *Ann Clin Biochem* 2015;52:434–40.
21. Klersy C, d'Eril GV, Barassi A, Palladini G, Comelli M, Moratti R, et al. Advantages of the lognormal approach to determining reference change values for N-terminal propeptide B-type natriuretic peptide. *Clin Chim Acta* 2012;413:544–7.
22. Sottas PE, Robinson N, Rabin O, Saugy M. The athlete biological passport. *Clin Chem* 2011;57:969–76.
23. Lobigs LM, Sottas PE, Bourdon PC, Nikolovski Z, El-Gingo M, Varamenti E, et al. The use of biomarkers to describe plasma-, red cell-, and blood volume from a simple blood test. *Am J Hematol* 2017;92:62–7.
24. Sottas PE, Kapke GF, Leroux JM. Adaptive Bayesian analysis of serum creatinine as a marker for drug-induced renal impairment in an early-phase clinical trial. *Clin Chem* 2012;58: 1592–6.

Article Note: The original submission can be found here: <https://doi.org/10.1515/almed-2020-0017>