



A Bayesian Approach to Biological Variation Analysis

Thomas Røraas,^{1*} Sverre Sandberg,^{1,2,3} Aasne K. Aarsand,³ and Bård Støve⁴

BACKGROUND: Biological variation (BV) data have many applications for diagnosing and monitoring disease. The standard statistical approaches for estimating BV are sensitive to “noisy data” and assume homogeneity of within-participant CV. Prior knowledge about BV is mostly ignored. The aims of this study were to develop Bayesian models to calculate BV that (*a*) are robust to “noisy data,” (*b*) allow heterogeneity in the within-participant CVs, and (*c*) take advantage of prior knowledge.

METHOD: We explored Bayesian models with different degrees of robustness using adaptive Student *t* distributions instead of the normal distributions and when the possibility of heterogeneity of the within-participant CV was allowed. Results were compared to more standard approaches using chloride and triglyceride data from the European Biological Variation Study.

RESULTS: Using the most robust Bayesian approach on a raw data set gave results comparable to a standard approach with outlier assessments and removal. The posterior distribution of the fitted model gives access to credible intervals for all parameters that can be used to assess reliability. Reliable and relevant priors proved valuable for prediction.

CONCLUSIONS: The recommended Bayesian approach gives a clear picture of the degree of heterogeneity, and the ability to crudely estimate personal within-participant CVs can be used to explore relevant subgroups. Because BV experiments are expensive and time-consuming, prior knowledge and estimates should be considered of high value and applied accordingly. By including reliable prior knowledge, precise estimates are possible even with small data sets.

© 2019 American Association for Clinical Chemistry

Biological variation (BV)⁵ data, describing the natural variation observed for most constituents in the human body, have many applications for diagnosing and monitoring disease (1). The magnitude of within-participant BV (CV_I) describes the expected variation of an analyte around a homeostatic set point within an individual and the between-participant BV (CV_G) variation between these set points (2). For the BV estimates to be clinically applicable on a population basis, they need to be representative for the population from which they have been derived and to which they will be applied. This fit can be established by testing for variance homogeneity to verify the applicability of a global CV_I (2). Nonetheless, variance homogeneity testing is often lacking, as shown by review of 128 BV publications by the Biological Variation Data Critical Appraisal Checklist (3). In the setting in which variance homogeneity cannot be verified, data must be checked for outliers or assessed for further subgrouping e.g., sexes or age groups. If this still does not deliver variance homogeneity, it may be necessary to accept heterogeneity because generalizable global BV estimates cannot be derived (4).

Estimating the components of BV is traditionally performed with frequentist methods, such as an ANOVA or a form of generalized linear model (2–4). Simply put, these methods return an average of the observed variances depending on the estimator chosen (5) and typically require a number of laborious statistical operations: assessment for outliers, variance homogeneity for both analytical CV (CV_A) and CV_I, and normality for the construction of the CI (6–8). This is the typical recommended approach by Fraser and Harris, based on a prior assumption of homogeneity (2). However, removing data points categorized as outliers or trimming data to achieve variance homogeneity might possibly remove valuable information. If more than a few observations are excluded, results may no longer be generalizable and applicable to the population from which they were derived.

¹ Norwegian Organisation for Quality Improvement of Laboratory Examinations (NOKLUS), Haraldsplass Deaconess Hospital, Bergen, Norway; ² Department of Global Health and Primary Care, Faculty of Medicine and Dentistry, University of Bergen, Norway; ³ Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; ⁴ Department of Mathematics, University of Bergen, Norway.

* Address correspondence to this author at: NOKLUS, PO 6165, N-5892 Bergen, Norway. Fax +47-55-97-95-10; e-mail thomas.roraas@icloud.com.

Received November 26, 2018; accepted May 13, 2019.

Previously published online at DOI: 10.1373/clinchem.2018.300145

© 2019 American Association for Clinical Chemistry

⁵ Nonstandard abbreviations: BV, biological variation; CV_G, between-participant BV; CV_I, within-participant BV; CV_A, analytical CV; CV_{P(i)}, personal within-participant BV; d(CV_{P(i)}), distribution of CV_{P(i)}; CrI, credible interval(s); EuBIVAS, European Biological Variation Study; LMM, linear mixed model; B_{NNN}, Bayesian normal-normal-normal model; B_{NNT}, Bayesian normal-normal-Student-*t* model; B_{NTN}, Bayesian normal-Student-*t*-normal model; B_{NNI}, Bayesian normal-Student-*t*-Student-*t* model; RCV, reference change value; ν_A, degrees of freedom for Student *t* distribution for samples; ν_i, degrees of freedom for Student *t* distribution for replicates; μ(CV_{P(i)}), mean of d(CV_{P(i)}); σ(CV_{P(i)}), SD of d(CV_{P(i)}); IQC, internal quality-control.

This is trying to shape the data to fit a model and not the other way around: find a model that fits and best describes the data at hand.

When estimating components of BV, we often have prior knowledge about their likely magnitude from previously published studies or from a different population group. In standard approaches, this information is only used to compare the results of a study or to discuss reasons for discrepancies between results from different studies (3, 9–11). **We propose that a better use of this knowledge is to include it in the prior distributions of a Bayesian model to achieve more precise estimates with Bayesian updating when analyzing new data.**

The aims of this study were to develop models to calculate BV by use of the Bayesian Markov Chain Monte Carlo approach as follows: (i) can be robust to extreme values ("noisy data," heavy tailed); (ii): **allow heterogeneity in the within-participant CV, in that each individual is allowed their own personal within-participant BV ($CV_{P(i)}$)**; and (iii) include prior knowledge, taking into account our belief in its trustworthiness and applicability. Further, the Bayesian models can be used to predict the distribution of the $CV_{P(i)}$, abbreviated as $d(CV_{P(i)})$. This result provides information about heterogeneity within a group and allows for better comparison between groups. The posterior distributions from the fitted model deliver credible intervals (Crl) for all parameters in the model.

To assess the applicability of these models, they were applied to the raw chloride and triglyceride data from the European Biological Variation Study (EuBIVAS) (9, 10). These 2 measurands were selected owing to their apparent characteristics, i.e., a high degree of homogeneity and heterogeneity, respectively, based on a preliminary test of the Bayesian models. **The results were compared to those achieved when applying a linear mixed model (LMM) for random effects with a maximum log-likelihood estimator and to previously published estimates (9).**

Methods

MATERIALS, METHODS, AND DATA

The EuBIVAS is a Biological Variation Data Critical Appraisal Checklist compliant, large-scale study designed to deliver updated BV estimates for a large number of measurands (10). The study population from which EuBIVAS data have been derived included 91 healthy volunteers: 38 men and 53 women, age 21–69 years. Fasting blood samples were drawn weekly for 10 consecutive weeks, and serum samples were kept frozen at -80°C until analysis, as previously described in detail (10). All samples from the same individual were analyzed in 1 run. The average number of samples per participant was 9.5 for both triglycerides and chloride, all measured

in duplicate (9). Data were as detailed in (9), normalized before outlier and variance homogeneity assessment (between replicates, Bartlett test; between samples, Cochran test) and thereafter analyzed by a CV-ANOVA method (12). CV_G was estimated by ANOVA on natural log-transformed data after applying the Dixon q test. The following results for duplicates (data points), samples (both duplicates), and participants (all data points for the participant) were excluded: (a) chloride: (i) overall study population, 3 duplicates, 1 sample; (ii) female subgroup, 3 duplicates, 1 sample; (iii) male subgroup, 1 sample; and (b) triglycerides: (i) overall study population, 1 duplicate, 7 samples, 2 participants; (ii) male subgroup, 4 duplicates, 1 participant.

MODEL DESCRIPTION

In the present work the following notation will be used: between-participant SD and CV are denoted as SD_G and CV_G , within-participant SD and CV as SD_I and CV_I , and analytical SD or CV as SD_A and CV_A (13). We will adopt the denotation of Harris, using $SD_{P(i)}$ and $CV_{P(i)}$ as the personal within-participant BV of the i -th individual (4).

Let $m_{sr(i)}$ be the measured value of the r -th replicate of the s -th sample from the i -th individual. Allowing for an unbalanced design we let $r = 1, \dots, R_{s(i)}$ denote the replicates for sample s for participant i ; $s = 1, \dots, S_{(i)}$ denote samples from participant i ; and $i = 1, \dots, I$ denote the participants. Thus, we allow for variation in both the number of replicates for each sample and the number of samples from each individual. **The basic 3-level hierarchical random model used for estimating BV is as follows (14):**

$$m_{sr(i)} = \mu + G_{(i)} + I_{s(i)} + A_{sr(i)} \quad (\text{model 1}),$$

Here it is assumed that we have an independent and additive effect, with $G_{(i)}$ being the random effect of person i on the group mean μ , specifically the difference between the group mean μ and the homeostatic set point μ_i for individual i . $I_{s(i)}$ is the random effect of the within-participant change on the value $\mu_{s(i)}$ of sample $s(i)$ for individual i . The analytical bias and variability is defined as $A_{sr(i)}$. If all samples from an individual are analyzed in 1 run, as in the data used in this study, it is assumed that the analytical bias is constant at group level or participant level. This assumption entails that bias will not influence variability of samples and will be included in the group mean μ , or in the individual effect $G_{(i)}$ if it is participant dependent. Thus, it is assumed that $A_{sr(i)}$ has an expected value of 0.

Because it is common to work with CVs, it is typically assumed that the model above is achieved through a log-transformation because CVs are assumed to be the most homogeneous measure of variability (12). Thus, the model can be written as (12):

$$\log(m_{sr(i)}) = \mu + G_{(i)} + I_{s(i)} + A_{sr(i)} \text{ (model 2).}$$

In a standard ANOVA or LMM model, all the effects are assumed random normally distributed, i.e., $G_{(i)} \sim N(0, SD_G)$, $I_{s(i)} \sim N(0, SD_I)$, and $A_{sr(i)} \sim N(0, SD_A)$ (8–10). Allowing for heterogeneity of the within-participant SD gives us instead that $I_{s(i)} \sim N(0, SD_{P(i)})$.

By back-transforming the SDs estimated on the log-transformed data, the corresponding CV on the original scale can be recovered by (15):

$$CV = 100 \times \sqrt{(\exp(SD^2) - 1)} \text{ (formula 1).}$$

Reference change value (RCV) in percentage can be calculated from the SD as:

$$RCV_{CV} = 100 \times (\exp(RCV_{SD}) - 1),$$

in which $RCV_{SD} = \sqrt{2} \times Z \times \sqrt{(SD_A^2 + SD_{P(i)}^2)}$ and Z depends on the direction and probability of change we want to detect (15–17).

The above model 2 is the ideal, but it assumes that the data are collected without errors and under unrealistically perfect conditions. What we typically observe is model 2 but with an added component of noise:

$$\log(m_{sr(i)}) = \mu + G_{(i)} + I_{s(i)} + A_{sr(i)}$$

+ Noise (model 3).

The term noise can be related to participants, samples, or (measurement results for individual) replicates. An adaptive Student *t* distribution can be applied to the random effects in the model, in which the degrees of freedom parameter in the Student *t* distribution is estimated from data (18). This estimation allows heavier tails, i.e., the model can be made more robust to extreme observations, depending on the estimated degrees of freedom parameters for the Student *t* distributions (18). A low degrees of freedom parameter gives a more heavy-tailed Student *t* distribution, whereas a high degrees of freedom parameter reflects a Student *t* distribution that is more gaussian like with lighter tails.

BAYESIAN ESTIMATION OF BV

A central idea of Bayesian inference is to use data to update prior knowledge or beliefs. This revision is done by fitting a probability model on data (likelihood) and summarizing the result in a revised posterior probability distribution (19). The prior knowledge for an uncertain parameter is typically summarized in a prior probability distribution that expresses one's beliefs about this parameter before observations are taken into account (19). From the resulting posterior distributions, CRI can be estimated on the percentiles for each of the estimated parameters in the model (19).

Bayes' theorem gives us the posterior for the parameter(s) θ conditional on data:

$$p(\theta|\text{data}) = p(\theta)p(\text{data}|\theta)/p(\text{data}),$$

in which $p(\theta)$ is the prior and $p(\text{data}|\theta)$ is the likelihood for data conditional on θ , which can be a vector of parameters of length ≥ 1 .

We aim to estimate the components of variation, CV_G and CV_A , and predict the distribution of personal $CV_{P(i)}$, $d(CV_{P(i)})$, for the group the data are made to represent, including the median, mean, $\mu(CV_{P(i)})$, and SD of $\sigma(CV_{P(i)})$.

The following prior distributions and hyperparameters—meaning simply the parameters of prior distribution(s), as opposed to parameters of the model—were chosen according to available estimates from the online 2014 Ricos' BV database and for CV_A data from Haukeland University Hospital (20, 21), as indicated by the “(database)” subscript:

- $SD_G \sim N(SD_{G(\text{database})}, \frac{1}{4}SD_{G(\text{database})})$;
- $SD_{P(i)} \sim N(\mu(SD_{P(i)}), \sigma(SD_{P(i)}))$;
- $\mu(SD_{P(i)}) \sim N(SD_{I(\text{DATABASE})}, \frac{1}{4}SD_{I(\text{DATABASE})})$;
- $\sigma(SD_{P(i)}) \sim N(0, 100)$;
- $SD_A \sim N(SD_{A(\text{database})}, \frac{1}{4}SD_{A(\text{database})})$.

Variances (hence also SDs) are inherently nonnegative, but negative estimates can occur in the Bayesian Markov Chain Monte Carlo process for divergent models. Because SDs are defined to be positive, the prior, likelihood, and posterior distributions of the SDs will not contain negative values because these estimates are rounded up to zero. This rounding to zero is to be kept to a minimum for a suitable model and should be evaluated from the posterior distributions.

Four different Bayesian models were explored, with different degrees of robustness relating to the use of Student *t* likelihoods: (i) the Bayesian normal-normal-normal (B_{NNN}) model, with normal likelihoods for homeostatic set points, samples, and replicates (corresponding with the assumptions for the classic ANOVA model (6–8, 19)), (ii) the Bayesian normal-normal-Student *t* (B_{NNT}) model with an adaptive Student *t* likelihood for the replicates; (iii) the Bayesian normal-Student-*t*-normal (B_{NTN}) model with adaptive Student *t* likelihoods for samples; and (iv) the Bayesian normal-Student-*t*-Student-*t* (B_{NTT}) model with adaptive Student *t* likelihoods for samples and replicates. The 4 models are defined by (on the log scale):

- Homeostatic set point for individual i : $\mu_i \sim N(\mu, SD_G)$,
- Sample s from individual i :
 1. B_{NNN} and B_{NNT} model: $\mu_{s(i)} \sim N(\mu_i, SD_{P(i)})$;
 2. B_{NTN} and B_{NTT} model: $\mu_{s(i)} \sim \text{Student } t(\nu_i, \mu_i, SD_{P(i)})$;
- Observed replicate r from sample s from individual i :
 1. B_{NNN} and B_{NTN} model: $m_{s(i)r} \sim N(\mu_{s(i)}, SD_A)$;
 2. B_{NNT} and B_{NTT} model: $m_{s(i)r} \sim \text{Student } t(\nu_A, \mu_{s(i)}, SD_A)$.

The hyperparameters ν_1 and ν_A are the degrees of freedom describing how heavily tailed the adaptive Student t distributions are. To calculate the corresponding normal SD_{normal} from the estimated SD_{Student t} with ν degrees of freedom, the following formula was applied:

$$\text{SD}_{\text{normal}} = \sqrt{(\text{SD}_{\text{Student } t}^2 \nu / (\nu - 2))} \quad (\text{formula 2}).$$

ν_1 and ν_A are defined the model to be >2.1 so as not to get “division by 0”-type problems when used (formula 2). The priors were set to $\nu_1 \sim N(\mu = 5, \sigma = 10)$ and $\nu_A \sim N(\mu = 3, \sigma = 5)$. These priors are based on prior knowledge from other data sets and analysis but also on the common approach of using a Student t distribution with $\nu = 4$ (18).

As an example, the CV₁ from the database for triglycerides is 19.9 (20). Thus, the corresponding SD₁ is $\sqrt{(\log((19.9/100)^2 + 1))} = 0.197$ when the inverse of formula 1 is used. The corresponding SD_{Student t} ² with $\nu = 5$ is then $\text{SD}_{\text{Student } t}^2 = \sqrt{[(\nu - 2)\text{SD}_{\text{normal}}^2/\nu]} = \sqrt{[5 - 3] \times 0.197^2/5} = 0.153$, when the inverse of formula 2 is used.

DATA ANALYSIS

Without first removing any aberrant results, the raw EuBIVAS data for triglycerides and chloride were assessed by the 4 Bayesian models B_{NNN}, B_{NNNT}, B_{NTN}, B_{NTT} and an LMM approach. The $\mu(\text{CV}_{\text{P(i)}})$, $\sigma(\text{CV}_{\text{P(i)}})$, CV_G, and CV_A with corresponding SDs were estimated with the resulting posterior distributions. Randomly generated CV_{P(i)} based on the estimated parameters were used to generate predicted distributions, d(CV_{P(i)}), and to calculate the 20%, 50%, and 80% percentiles and the SD for the predicted d(CV_{P(i)}). These results were compared to the CV_A, $\mu(\text{CV}_{\text{P(i)}})$, and CV_G estimated with the LMM approach and to previously published EuBIVAS estimates (9). The posterior distribution of the parameters in the model were also used to deliver the personal CV_{P(i)} and their 95% credible intervals (Crls) for the 91 individuals.

As a measure of heterogeneity, the Harris–Brown heterogeneity ratio, i.e., $100 \times (\sigma(\text{CV}_{\text{P(i)}})/\mu(\text{CV}_{\text{P(i)}}))$ (16) was calculated for the overall group and the subgroups defined by sex. We also calculated the predicted Harris–Brown ratio by using the median and SD from the predicted d(CV_{P(i)}).

SOFTWARE

We used R (version 3.5.1, <https://www.r-project.org/>) with *tidyverse* (version 1.2.1, <https://www.tidyverse.org/>) and *rstan* (version 2.17.0, <https://mc-stan.org/users/interfaces/rstan>) running on a Manjaro Gnu/Linux box. Stan is a probabilistic programming language for implementation of Bayesian inference (22). For LMM, the *lme4* package (version 1.1–18–1, <https://cran.r-project.org/web/packages/lme4/index.html>) was applied. Codes

with a simulated data set are given as supplemental text files and are available at https://gitlab.com/thoror/bayesian_bv_paper.

Results

The 4 Bayesian approaches gave different predicted distributions d(CV_{P(i)}) for chloride, whereas the distributions clearly exhibited greater similarity for triglycerides (Fig. 1). The estimates of CV_A, $\mu(\text{CV}_{\text{P(i)}})$, and $\sigma(\text{CV}_{\text{P(i)}})$ were strongly dependent on the method used (Table 1). The estimated degrees of freedom for the adaptive Student t distributions, ν_1 and ν_A , indicated extreme observations in the data, especially between replicates. There was also some apparent heterogeneity in triglycerides. When grouped based on sex, the predicted d(CV_{P(i)}) for women and men were similar for chloride but differently distributed for triglycerides (B_{NNT} model) (Fig. 2). Individual CV_{P(i)} with their corresponding 95% CrI for each of the 91 participants are presented in Fig. 3. No correlations between the personal CV_{P(i)} and the corresponding homeostatic set points, μ_i , were observed (see Fig. 1 in the Data Supplement that accompanies the online version of this article at <http://www.clinchem.org/content/vol65/issue9>). The estimated $\mu(\text{CV}_{\text{P(i)}})$ with corresponding 95% CI or CrI from Table 1 are visually summarized in Fig. 4. The SDs in the priors for BV derived from the Ricos database and the CV_A estimates have little effect on the estimated d(CV_{P(i)}) for triglycerides on the basis of data from the 38 men. Different priors for heterogeneity have as expected more effect on d(CV_{P(i)}), as illustrated for triglycerides in Fig. 5.

Discussion

The main advantages of using the presented Bayesian approach is the flexibility it allows when estimating the components of BV. With the Bayesian approach we can use adaptive Student t distributions for the different levels in the hierarchical model and estimate the ν , which describes the weight of the tails, from data. This procedure allows us to disregard the assumption of normality and makes the model robust to extreme observations. The ability to include prior knowledge as hyperparameters, and give these weights on the basis of either statistical or subjective reasoning, can give more precise estimates with narrower posteriors. The ability to generate estimates of the individual CV_{P(i)} is primarily useful to evaluate the degree of heterogeneity, to explore trends for evaluating subgroups, to identify participants that appear not to be belonging to the group, and to assess if the individual CV_{P(i)} are related to the participants' homeostatic set points.

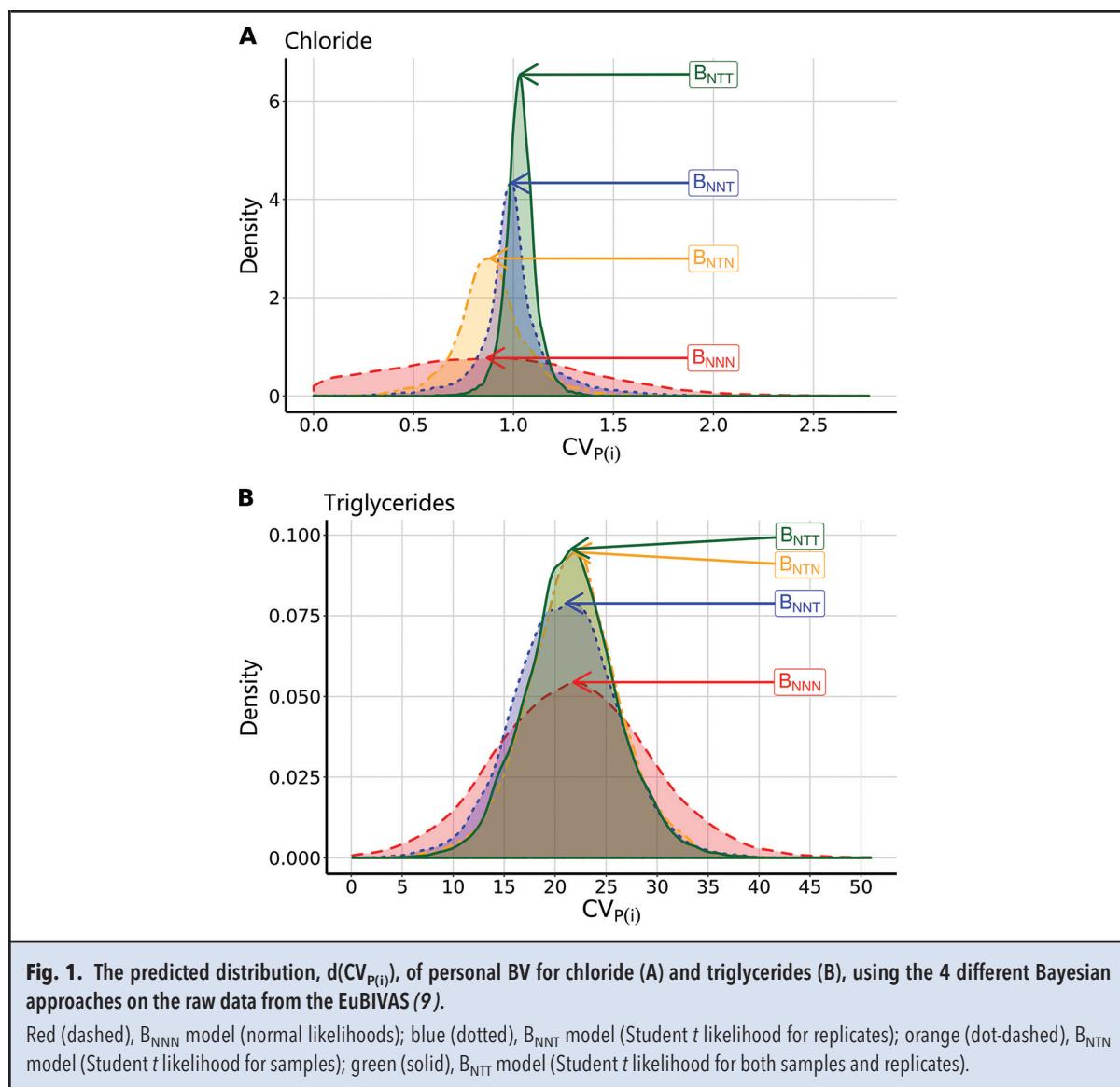


Fig. 1. The predicted distribution, $d(\text{CV}_{\text{P}(i)})$, of personal BV for chloride (A) and triglycerides (B), using the 4 different Bayesian approaches on the raw data from the EUBIVAS (9).

Red (dashed), B_{NNN} model (normal likelihoods); blue (dotted), B_{NNT} model (Student t likelihood for replicates); orange (dot-dashed), B_{NTN} model (Student t likelihood for samples); green (solid), B_{NTT} model (Student t likelihood for both samples and replicates).

In our study, we have tested 4 different Bayesian models to deliver BV estimates, applied to 2 large-scale data sets for chloride and triglycerides. The B_{NNN} model, which assumes normal likelihoods, is evidently the least adaptive of the 4 Bayesian models for both analytes (men and women combined)—more strikingly so for chloride than for triglycerides—as shown by the low height (density) and broad spread of the B_{NNN} distribution in each panel of Fig. 1. This result can be an indicator of noise or extreme observations in the data set, and therefore the normality assumptions for the B_{NNN} model do not fit the data. In contrast, the 4 Bayesian models were more in agreement for triglycerides (Fig. 1B, Table 1). This result shows that for data with no highly deviant values, the models

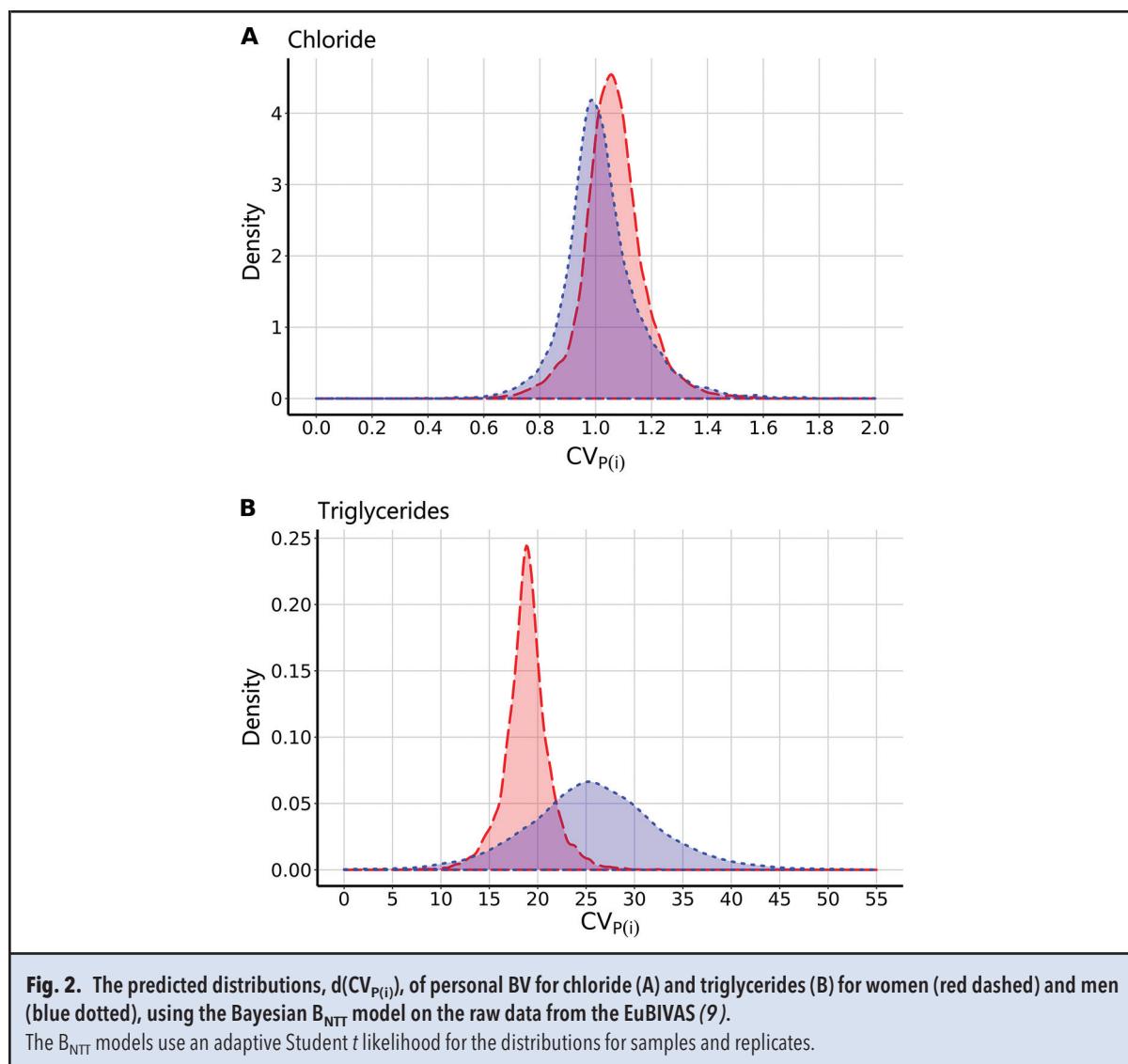
using Student t distributions really do adapt. This finding can also be observed from Table 1, in which ν_1 is larger for triglycerides than chloride, signifying a less heavy-tailed distribution for the latter.

For the combined data sets, we also observe from Table 1 that mean $\mu(\text{CV}_{\text{P}(i)})$ was nearly the same for all models other than the LMM, whereas mean $\sigma(\text{CV}_{\text{P}(i)})$ and mean CV_A were smaller for the more robust B_{NNT} and B_{NTT} models. This result also shows the LMM's sensitivity to outliers in that, by removing 1 participant (ID = 51), the LMM approach applied to the combined triglycerides data set would yield a mean $\mu(\text{CV}_{\text{P}(i)})$ of 21.1% instead of 25.1%. Moreover, for both chloride and triglycerides, out of the 4 Bayesian models, B_{NTT} delivered mean CV_A values for the combined data closest

Table 1. Estimated parameters derived from the 4 Bayesian models with corresponding hyperparameters (parameters of prior distribution), the LMM (linear mixed model) approach on raw data, and the previously published results using ANOVA combined with outlier removal (9).

Analyte (data set)	Statistical method	Sex	ν_I		ν_A		Predicted d($CV_{P(i)}$)				Estimated $\mu(CV_{P(i)})$		Estimated $\sigma(CV_{P(i)})$		CV_G		CV_A		Harris-Brown heterogeneity ratio ^a	
			Mean	SD	Mean	SD	20%	Median	80%	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Estimated	Predicted
Chloride (combined)	Hyperparameters	All	5.0	10.0	3.0	5.0	NA ^b	NA	NA	NA	1.20	0.30	0.00	100.00	1.50	0.38	2.00	0.50	NA	NA
	LMM		NA	NA	NA	NA	NA	NA	NA	NA	1.20	NA	NA	NA	1.33	NA	0.97	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	0.98	NA	NA	NA	1.34	NA	0.40	NA	NA	NA
	B_{NNN}		NA	NA	NA	NA	0.44	0.87	1.29	0.48	0.86	0.07	0.51	0.06	1.39	0.11	0.97	0.02	59.3	55.4
	B_{NNT}		NA	NA	2.3	0.2	0.90	0.99	1.09	0.19	1.01	0.04	0.15	0.11	1.41	0.09	0.65	0.13	14.5	19.1
	B_{NTN}		3.4	0.6	NA	NA	0.76	0.88	1.01	0.19	0.89	0.10	0.08	0.05	1.38	0.11	0.98	0.02	9.4	21.3
	B_{NTT}		7.5	1.5	2.4	0.2	0.98	1.03	1.09	0.08	1.03	0.04	0.05	0.03	1.38	0.11	0.58	0.11	4.7	7.6
Triglycerides (combined)	Hyperparameters	All	5.0	10.0	3.0	10.0	NA	NA	NA	NA	19.9	5.0	0.0	100.0	32.7	8.2	3.0	0.8	NA	NA
	LMM		NA	NA	NA	NA	NA	NA	NA	NA	25.1	NA	NA	NA	42.0	NA	3.3	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	19.8	NA	NA	NA	40.3	NA	1.8	NA	NA	NA
	B_{NNN}		NA	NA	NA	NA	15.6	21.8	27.9	7.4	21.7	1.0	7.1	0.8	41.4	3.1	3.3	0.1	32.7	33.9
	B_{NNT}		NA	NA	2.4	0.2	16.9	21.0	25.1	5.0	21.0	0.8	4.7	0.7	38.1	2.2	2.8	0.6	22.4	24.0
	B_{NTN}		8.0	2.1	NA	NA	18.1	21.8	25.4	4.6	21.8	0.9	3.6	1.0	41.4	3.2	3.3	0.1	16.7	21.3
	B_{NTT}		11.7	5.8	2.5	0.3	18.0	21.6	25.2	4.5	21.5	0.9	3.7	0.9	41.3	3.1	2.5	0.5	17.1	20.9
Chloride (by sex)	LMM	F	NA	NA	NA	NA	NA	NA	NA	NA	1.34	NA	NA	NA	1.31	NA	0.37	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	0.99	NA	NA	NA	1.35	NA	0.40	NA	NA	NA
	B_{NTT}		6.4	1.5	3.9	0.8	0.99	1.06	1.14	0.11	1.06	0.06	0.06	0.04	1.41	0.14	0.38	0.03	6.0	10.3
	LMM	M	NA	NA	NA	NA	NA	NA	NA	NA	0.95	NA	NA	NA	1.35	NA	1.44	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	0.96	NA	NA	NA	1.34	NA	0.40	NA	NA	NA
	B_{NTT}		13.1	6.8	2.2	0.1	0.93	1.01	1.11	0.14	1.02	0.09	0.08	0.05	1.40	0.16	0.72	0.13	7.5	13.7
Triglycerides (by sex)	LMM	F	NA	NA	NA	NA	NA	NA	NA	NA	18.8	NA	NA	NA	36.1	NA	1.9	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	19.1	NA	NA	NA	38.5	NA	1.8	NA	NA	NA
	B_{NTT}		16.9	5.9	4.9	1.4	17.3	18.9	20.5	2.4	18.9	0.8	1.7	1.0	36.7	3.7	1.9	0.1	9.1	12.5
	LMM	M	NA	NA	NA	NA	NA	NA	NA	NA	32.1	NA	NA	NA	44.4	NA	4.6	NA	NA	NA
	CV-ANOVA (outliers removed)		NA	NA	NA	NA	NA	NA	NA	NA	22.7	NA	NA	NA	42.9	NA	1.8	NA	NA	NA
	B_{NTT}		6.6	2.3	2.3	0.2	20.4	25.6	31.0	6.9	25.7	2.0	4.9	1.7	42.4	4.5	2.6	0.5	19.1	26.7

^a The Harris-Brown heterogeneity ratio = $100 \times (\sigma(CV_{P(i)})/\mu(CV_{P(i)}))$ (16).^b NA, not applicable; F, females; M, males.

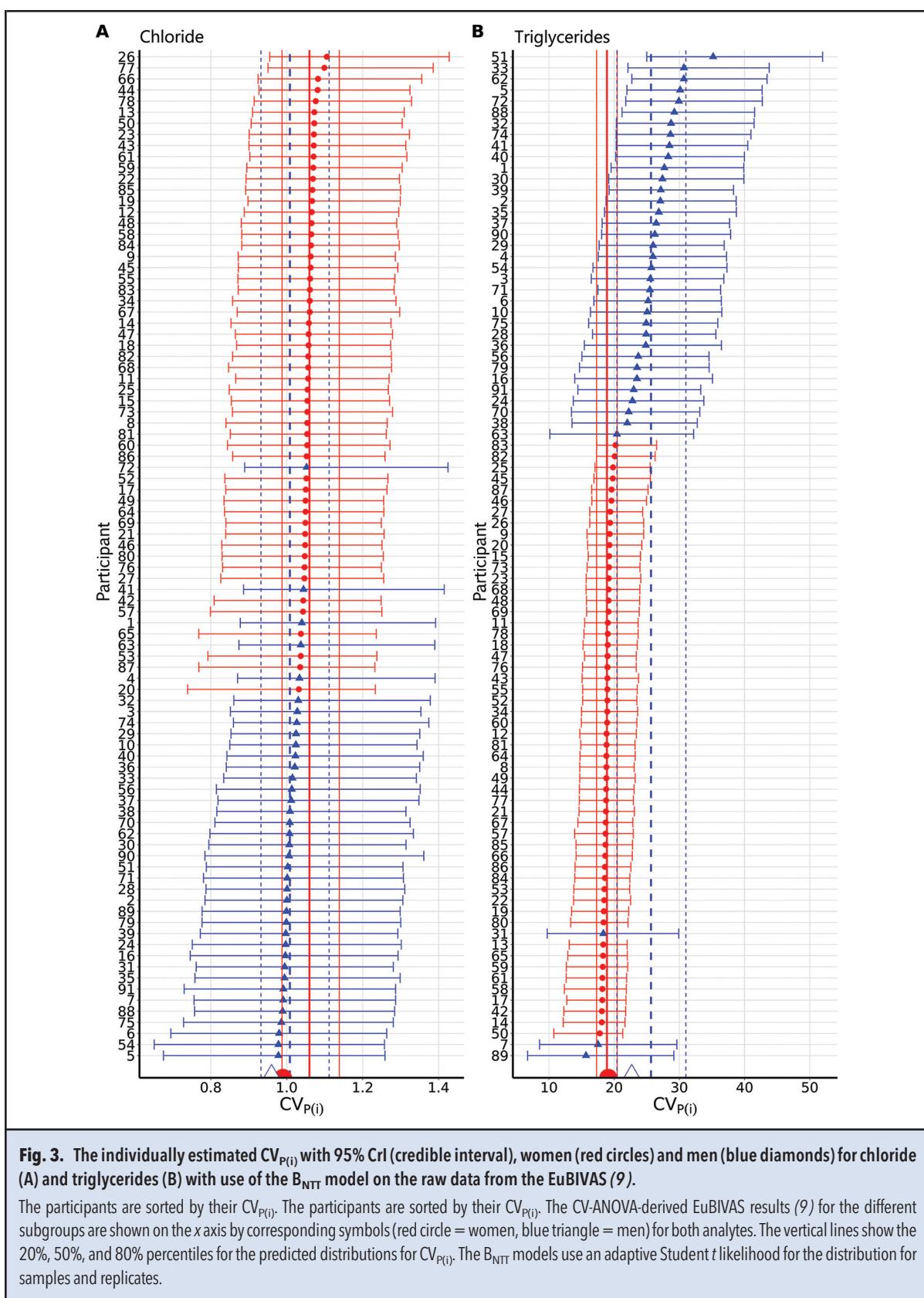


to those of the approach consisting of CV-ANOVA with outlier removal (Table 1).

Because, for both analytes, the B_{NTT} model appeared to be the best choice for estimating BV components from the combined data, we selected it for analyzing the data sets subgrouping by sex. A similar $d(CV_{P(i)})$ was observed for both groups for chloride (Fig. 2A). This finding is expected considering the narrow $d(CV_{P(i)})$ observed for the combined group (Fig. 1A). Both 60% prediction intervals are narrow and severely overlapping, so assuming homogeneity and using a global estimate of CV_1 seems adequate for chloride (Table 1). For triglycerides, however, we observe a much narrower $d(CV_{P(i)})$ for women than men when using the B_{NTT} model (Table 1). This result indicates that women make up a fairly homogeneous

group, whereas men are more heterogeneous, as is also illustrated in Figs. 2B and 3B. When ordering the participants by $CV_{P(i)}$ we get a near-perfect grouping by sex for chloride, except in the middle (Fig. 3A). For triglycerides, there are a couple of men mixed in with the women at the lower end of the scale (Fig. 3B). Therefore, although using the individual $CV_{P(i)}$ to evaluate subgroups can be useful, with such a narrow $d(CV_{P(i)})$ as observed for the chloride example, further subgrouping seems unnecessary, especially when based on only 10 samples from each individual.

Use of the predicted Harris-Brown heterogeneity ratio is recommended, because it shows the extent of variability in relation to the mean in the population. For a homogeneous population with an average number of S samples from each individual, this ratio is expected to be below $100/\sqrt{2S}$ for



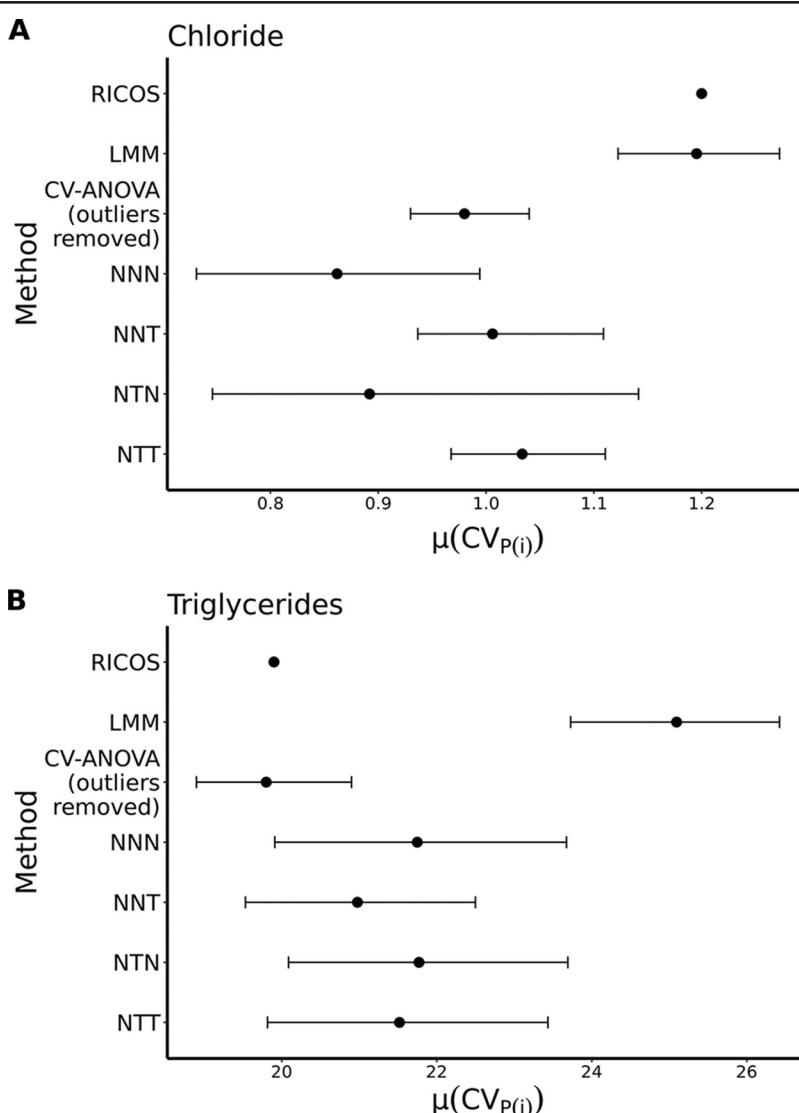
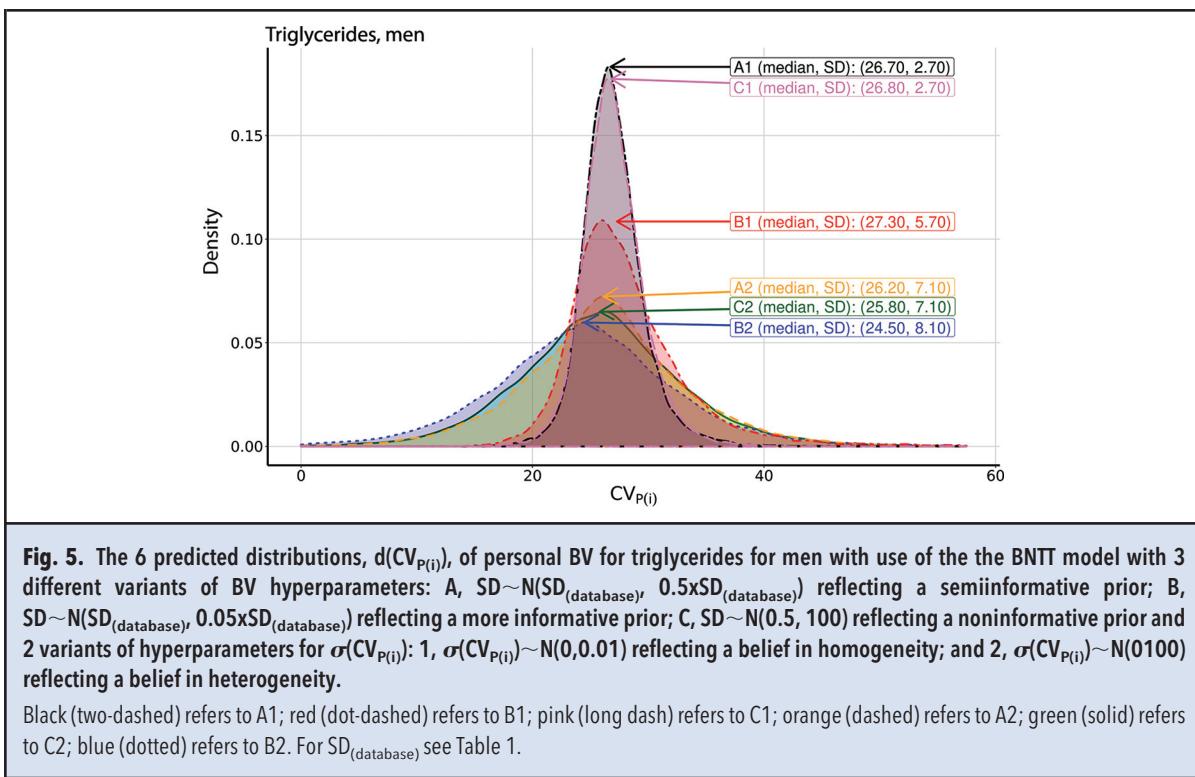


Fig. 4. Estimated mean $CV_{P(i)}$, $\mu(CV_{P(i)})$, for chloride (A) and triglycerides (B) applying the 4 Bayesian approaches (B_{NNN} , B_{NNT} , B_{NTN} , B_{NTT}) and the LMM on the raw data from the EuBIVAS, the published CV-ANOVA with outlier removal (9) and the estimate in the Ricos' database (20), provided with corresponding 95% CI or CrI.

a normal approximation (16). In a homogeneous population from which 9.5 samples have been included, this ratio is expected to be <22.9%. From Table 1 we observe that the Harris–Brown heterogeneity ratio is <22.9% for both the chloride and triglycerides examples when using the estimated ratio based on the results from the recommended B_{NTT} model, failing to indicate heterogeneity of $CV_{P(i)}$ for either of the combined data sets. When using the predicted ratio based on the median and SD of $d(CV_{P(i)})$, the conclusion is the same, but for triglycerides it is 20.9, which is quite large compared to the corresponding ratio of 7.6 for the combined chloride data. When reviewing the subgroups,

the Harris–Brown heterogeneity ratio indicates that the female group is homogeneous for triglyceride but the male group appears to be heterogeneous, with a predicted ratio of 26.7. These conclusions are also in agreement with Fig. 1 and Fig. 2.

In the Bayesian models we have applied, the hyperparameters for SD_G , $\mu(SD_{P(i)})$, and SD_A can and should be adjusted according to the reliability and relevance of prior knowledge. Examples could be to let the prior for SD_A reflect results from internal quality-control (IQC) using replicates, $SD_A \sim N(SD_{IQC}, SD(SD_{IQC}))$, or the priors and their hyperparameters could be based on pub-



lished estimates and their CI or CrI. The flexibility of the Bayesian approach would also allow us to model the analytical imprecision based on precision profiles and not simply as a CV_A estimate.

We applied hyperparameters for BV derived from the Ricos' database (Table 1). The reliability and relevance of these can be adjusted in the hyperparameters for the SDs of these BV estimates, in which we used $\frac{1}{4}SD$: $SD \sim N(SD_{(database)}, \frac{1}{4}SD_{(database)})$. For a more informative prior, a hyperparameter of 5% of the BV component can be applied: $SD \sim N(SD_{(database)}, 0.05 \times SD_{(database)})$. For a less informative prior, a hyperparameter of 50% of the BV component can be applied: $SD \sim N(SD_{(database)}, \frac{1}{2}SD_{(database)})$. If we have no prior information, we could use a noninformative prior for a BV component, for example $SD \sim N(0.5100)$. For the heterogeneity parameter we used a noninformative prior, $\sigma(SD_{P(i)}) \sim N(0, 100)$. For an informative prior reflecting a belief of a nonheterogeneous group, $\sigma(SD_{P(i)}) \sim N(0, 0.01)$ can be applied. The effect of applying these different priors on the $d(CV_{P(i)})$ for triglycerides in men can be seen in (Fig. 5). The 6 curves reflect how different levels of reliability given to the Ricos BV estimates (C, A, B = not reliable, somewhat reliable, very reliable) and whether assumption of homogeneity for $CV_{P(i)}$ (1, 2 = yes, no) affect the predicted density $d(CV_{P(i)})$ using the B_{NTT} model. The example shows that the prior for homogeneity has most influence on $d(CV_{P(i)})$ because the SD for the distribu-

tions using $\sigma(SD_{P(i)}) \sim N(0, 0.01)$ (A1, B1, and C1) are less than the SD for the distributions in which noninformative priors for $\sigma(SD_{P(i)})$ were applied (A2, B2, and C2). The distribution with 5% SD (B2, blue, dotted) is influenced by the hyperparameter of 19.9% for $\mu(CV_{P(i)})$ the most; thus, the median is the lowest for this model (Fig. 5).

The effect of the applied BV hyperparameters will depend on how much trust we have in them and the size of the data set being analyzed. If we have a small data set and trustworthy and relevant hyperparameters we can gain more from prior information.

Yet another advantage of the Bayesian methods included in our study is that the crude personal $CV_{P(i)}$ for the 91 participants are estimated simultaneously with the CV_A . We do not get the problem of negative estimates of $CV_{P(i)}$ for certain individuals, as observed by Harris when calculating each $SD_{P(i)}$ independently and then subtracting the global SD_A estimated from control material (4, 16). In addition, the Bayesian approaches also return a CrI for all parameters in the model using the posterior distributions, making it superfluous to do normality checks or bootstrap approaches otherwise necessary when establishing CI (23).

Conclusion

In our study, we have demonstrated that with a Bayesian approach it is possible to make models that give direct infer-

ence regarding the degree of heterogeneity of $CV_{P(i)}$ and that adapt to data and are therefore relatively robust to extreme observations. This approach is an advantage over classical frequentist methods such as ANOVA and also reduces the need for laborious statistical operations and possible subjectivity in data trimming to achieve homogeneity and exclusion of outliers. Furthermore, it solves the issue that if more than a few observations are excluded, results may no longer be generalizable and applicable to the population from which they were derived. The ability to crudely estimate personal $CV_{P(i)}$ can be used to explore relevant subgroups or to identify individuals not belonging to the group. The Bayesian approach also provides the ability to use prior knowledge, which makes precise inference from smaller data sets possible if the priors are reliable and relevant. We recommend that, in future BV studies, an adaptive Bayesian model is used to deliver BV estimates.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising

the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

T. Røraas, statistical analysis.

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: None declared.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: T. Røraas, postdoctoral funding from the Western Norway Regional Health Authority.

Expert Testimony: None declared.

Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors would like to thank the EFLM Working Group on BV for the use of data from the EuBIVAS and the Western Norway Regional Health Authority for providing postdoctoral funding for the first author. The authors would also like to thank the reviewers whose input has greatly improved this paper.

References

1. Aarsand AK, Røraas T, Sandberg S. Biological variation—reliable data is essential. *Clin Chem Lab Med* 2015;53:153–4.
2. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989;27:409–37.
3. 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.
4. Harris EK. Distinguishing physiologic variation from analytic variation. *J Chronic Dis* 1970;23:469–80.
5. Faraway JJ. Extending the Linear Model with R. 2nd edition. Boca Raton (FL): CRC Press; p. 196–201.
6. Burdick RK, Graybill F. Confidence intervals on variance components. New York (NY): Marcel Dekker; 1992.
7. Sahai H, Ojeda MM. Analysis of variance for random models: volume 1 balanced data. Boston (MA): Birkhäuser; 2004.
8. Sahai H, Ojeda MM. Analysis of variance for random models: volume 2 unbalanced data. Boston (MA): Birkhäuser; 2005.
9. 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.
10. 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.
11. Carobene A, Marino I, Coşkun A, Sertereser 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.
12. Røraas 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.
13. Simundic A-M, 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.
14. 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.
15. Fokkema MR, Herrmann Z, Muskiet FAJ, Moecks J. Reference change values for brain natriuretic peptides revisited. *Clin Chem* 2006;52:1602–3.
16. Harris EK, Brown SS. Temporal changes in the concentrations of serum constituents in healthy men. *Ann Clin Biochem* 1979;16:169–76.
17. Harris EK, Yasaka T. On the calculation of a "reference change" for comparing two consecutive measurements. *Clin Chem* 1983;29:25–30.
18. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd edition. Boca Raton (FL): CRC Press; 2013. pp. 435–445.
19. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd edition. Boca Raton (FL): CRC Press; 2013.
20. Minchinela J, Ricos C, Perich C, Fernández-Calle P, Alvarez V, Doménech MV, et al. Biological variation database and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update. <http://www.westgard.com/biodatabase-2014-update.htm> (Accessed May 2018).
21. Helse Bergen, Haukeland University Hospital. Analyseoversiktene. www.analyseoversiktene.no (Accessed December 2018).
22. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. *J Stat Softw* 2017;76:1–32.
23. Røraas T, Stove B, Petersen PH, Sandberg S. Biological variation: evaluation of methods for constructing confidence intervals for estimates of within-person biological variation for different distributions of the within-person effect. (Report). *Clin Chim Acta* 2017;468:166–73.