Thesis

Peón Pena, Gonzalo

2025-01-08

Table of contents

Abstract

The Concordance Correlation Coefficient has been one of the key analysis measures of concordance in particular for the repeated measures outcomes in the continuous scale. The coefficient relies on a key distributional assumption of normality of the response variable. In this thesis the main estimation approaches for the longitudinal CCC have been reviewed through simulation sets under distributional misspecifications, where both bayesian and bootstrap Bca approaches obtained the better coverage, but nonetheless all reviewed methods failed to reach its nominal coverage under strong right-skewness on the response variable. The application of popular transformations for the asymptotic estimate did not provide visible coverage improvements, while in the presence of a drop-out pattern robust methods that require complete case analysis were significantly limited for the longitudinal CCC estimation.

1 Motivation

In many scenarios of statistical practice, as the comparison between alternative measuring techniques, which might be relevant for medical practice scenarios if the novel method had been shown to be less invasive than the current gold standard[1]; or in machine learning trying to gauge the effectiveness of a knowledge distillation strategy[2]; or when trying to assess the degree of agreement of different raters[3]; it is highly relevant to provide a measure of agreement consistent with the typology of the observed data, and assesses the variability and biases present accordingly. Within the nominal and ordinal scales, the use of the κ index, and its weighted version, are widely used. They were first proposed in Cohen (1960) and in Cohen (1968) for the situation for binary and multiple classification. Both indices reflect the difference from observed agreement between the evaluation methods and the agreement by chance, with values close to 1 reflecting perfect agreement and values close to zero a poor degree of agreement. For the numerical scale, and based on a analysis of variance (ANOVA) model, Fisher (1925) proposed the Intraclass Correlation Coefficient (ICC), as a measure of agreement between observers, which specifically can be interpreted and computed as a ratio between the variability in the responses due to the subjects and the total variability observed in the responses. Values close to one would indicate that the within-subjects variability has negligible impact on the response values, what would indicate high levels of agreement, while high method variability or high levels of unexplained variance would indicate instead little reliability between the methods.

The underlying ANOVA model implicitly assumes when fitting a linear relationship between the observed values and the components, that the subject effects are independent and identically distributed and follow a normal distribution with zero mean and σ^2 variance, which is the same for the random errors, and both are mutually exclusive. The observer effects can be either considered as another normal random effect, if the expectation is that there are no systematic differences between any pair of methods, or as a fixed effect, with the added restriction of centrality.

Still, this coefficient intrinsically depends on the ANOVA conditions to hold, in order to be correctly specified and retain interpretability[4]. Given its limitations, and the interest to extend the reliability study to replications, a mathematical comparable analogue was then proposed that was not withholden to the ANOVA assumptions.

[1] Gatti et al. (2024)
[2] Mitra et al. (2022) .
[3] Armour et al. (2024) .
[4] Chen and Barnhart (2008) .

The concordance correlation coefficient (CCC) is one of the most well established concordance indices in the literature for the continuous scale, and since its initial proposal in L. I.-K. Lin (1989), it has been extended to multiple raters[5], longitudinal[6] and count data[7], and its connection with the Intraclass Correlation Coefficient, and with the Cohen's weighted kappa has been explored.

Given the distributional assumption of multivariate normality on which the estimate was conceived, this opens up to a potential and limiting issue of the CCC under this misspecification. Neither the coefficient's formula will be correct under a different distribution, nor the model strategy (if fitted with a Linear Mixed Model) will be valid. As such, it has been one of the most covered issues of the coefficient[8] [9] [10], even if other misspecifications may also thread the model's validity (e.g., heteroscedasticity and non-linearity of residuals or the effect of a missingness mechanism).

[5] Barnhart, Haber, and Song (2002) .

[6] Josep L. Carrasco et al. (2013) .

[7] Josep L. Carrasco (2009) .

[8] King, Chinchilli, and Carrasco (2007) .

[9] Feng, Baumgartner, and Svetnik (2015) .

[10] Feng, Baumgartner, and Svetnik (2018) .

2 Aims

The misspecification of the distributional assumption of the Concordance Correlation Coefficient has led to the exploration of two avenues in order to address it: Fit an appropriate model and update the coefficient's formula accordingly (Josep L. Carrasco (2009)), or to find robust methods of estimation of the coefficient (King and Chinchilli (2001b),Feng, Baumgartner, and Svetnik (2018)). In this thesis the focus will be placed on the latter, which has been the main line of study in the literature (already in the coefficient's proposal, Lin tested the robustness the coefficient under misspecification of the distribution), comparing the robustness (through coverage and bias) of different estimation methods that have been expanded or proposed since Josep L. Carrasco et al. (2007). These methods will be compared in a simulation study under five different levels of longitudinal concordance, and three levels of skewness.

The impact of transformations of the data, in particular the log-transformation, in order to obtain normality and its connection to the correlation coefficient, will be assessed in the first case study, a small dataset with moderate skewness of bacterial colony formations in paired molar samples. The second case study will review a large longitudinal dataset with time-varying concordance, potential covariates, and a significant dropout process, where possible under or overspecification has been reviewed, accompanied with an assessment of the impact of the observed missing pattern on the estimation methods.

3 The Concordance correlation coefficient

The Concordance Correlation Coefficient was initially proposed in L. I.-K. Lin (1989) for the measure of the degree of concordance between two random variables, Y_1 and Y_2, that follow a normal bivariate distribution[1]. The index was defined as a scaled (in order to be bounded between 1 and -1) and inverted (1 minus the value) version of the expected value of the squared difference ($E[(Y_1-Y_2)^2]$), which then could be expressed as the product of the Pearson correlation coefficient ρ (as a measure of precision) and the bias correction factor C_b that captures both the scale and location shifts from the 45 degree line (complete concordance).

$$CCC = \rho C_b = 1 - \frac{E[(Y_1 - Y_2)^2]}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

Lin also develops an asymptotic estimate of the variance of the sample Concordance Correlation Coefficient given the consistency of the estimator, which was tested under the uniform and Poisson distributions, and the asymptotical normality of its distribution (it is shown that this estimate can be improved using Fisher's Z transformation). Finally, it also delivers the first proposal for the generalized version of the coefficient for more than two readers.[2]

The extension of the coefficient for multiple observers was later fully developed in King and Chinchilli (2001b) where the Generalized Concordance Coefficient was proposed and an U-statistic estimate of the coefficient was introduced and reviewed; also as the Overall Concordance Correlation Coefficient in Barnhart, Haber, and Song (2002) with a Generalized Estimating Equations given, and also a third approach (which was also an expression of the GCCC) was proposed in Josep L. Carrasco and Jover (2003) through the variance components methods, which was later extended for longitudinal data and for repeated measures in King, Chinchilli, and Carrasco (2007) and in Josep L. Carrasco, King, and Chinchilli (2009).

$$OCCC = \frac{2\sum_{j=1}^{J-1}\sum_{k=j+1}^{J}\sigma_{jk}}{(J-1)\sum_{j=1}^{J}\sigma_j^2 + \sum_{j=1}^{J-1}\sum_{k=j+1}^{J}(\mu_j - \mu_k)^2}$$

With μ_j and σ_j^2 referring to the mean and variance of the jth individual, and σ_{jk} the covariance between individuals j and k.

[1] $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim MN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$

[2] Both the generalized coefficient and the estimate for the variance of the coefficient would be further corrected in "Correction: A Note on the Concordance Correlation Coefficient" (2000).

3.1 The robust U-statistics estimate

In contrast to the Overall CCC, which was developed and defined as a weighted average of the pairwise CCC, focusing on extending the coefficient for multiple observers, the focus of the Generalized CCC when it was first proposed in King and Chinchilli (2001a), was to extend the coefficient for the cases where the squared distance function ($(Y_1 - Y_2)^2$) might not have been appropriate, which is the case under non-normality, to a whole family of convex continuous distance functions ($g(Y_1 - Y_2)$), with some Winsored versions and Huber's function showing greater robustness. For the comparison of two methods, the GCCC formula is the following:

$$CCC = \frac{[E_{F_{Y_1}F_{Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1}F_{Y_2}} g(Y_1 + Y_2)] - [E_{F_{Y_1 Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1 Y_2}} g(Y_1 + Y_2)]}{E_{F_{Y_1}F_{Y_2}} g(Y_1 - Y_2) - E_{F_{Y_1}F_{Y_2}} g(Y_1 + Y_2) + \frac{1}{2}(E_{F_{Y_1}} g(2Y_1) + E_{F_{Y_2}} g(2Y_2))}$$

Which could be reduced to the following form for estimation by replacing the terms with the U-statistics which they showed to be unbiased estimators for the previous terms for a sample size of size n:

$$C\hat{C}C = \frac{(n-1)(U_3 - U_1)}{U_1 + nU_2 + (n-1)U_3}$$

Then, in King and Chinchilli (2001b), the authors extended the coefficient to the multiple raters case by replacing the U_1, U_2 and U_3 by their sum across all the unique pairs U_{1S}, U_{2S} and U_{3S}, given the U-statistics property that the sum of the U-statistics is in itself a U-statistic, and proved that the multi-rater version of the coefficient still was asymptotically normal.

$$C\hat{C}C = \frac{(n-1)(U_{3S} - U_{1S})}{U_{1S} + nU_{2S} + (n-1)U_{3S}}$$

It was finally extended to the repeated measures case, with particular attention for longitudinal data, in King, Chinchilli, and Carrasco (2007), where nU was the unbiased estimator of the expected value of the weighted squared differences $E[(X-Y)'D(X-Y)]$ and $U + (n-1)V$ the unbiased estimator of the previous expected value under independence.

$$C\hat{C}C = \frac{(n-1)(V - U)}{U + (n-1)V}$$

3.2 The Variance Components Approach

In this thesis the Concordance Correlation Coefficient formula which be mainly discussed is the one proposed in Carrasco (2003), estimated through the variance components. It suggests a linear mixed measurement model where the continuous variable is measured a total of m times (the product between the n number of subjects, k number of raters, and l number of replicates), which requires that every possible combination of raters and times to be accounted. Also, the explanation has been limited to a specific type of block diagonal, with the condition of identity, for the variance-covariance matrix of the random effects of the underlying LMM. The proposed model was the following: $Y_{ijl} = \mu + \alpha_i + \beta_j + \epsilon_{ijl}$, with $i = 1, ..., n$ subjects, $j = 1, ..., k$ raters and $l = 1, ..., r$ replicates; where μ, the overall mean, and β_j, the rater j effect, are the fixed effects; and $\alpha_i \sim N(0, \sigma^2_\alpha)$, the individual effect, and $\epsilon_{ijl} \sim N(0, \sigma^2_\epsilon)$, the random error, which are the random effects of the model.

$CCC = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\epsilon}$.

Working with the restriction on the sum of the rater effects have to be equal to zero ($\sum_{j=1}^{k} \beta_j = 0$), the variability among all raters can be derived:

$$\hat{\sigma}^2_\beta = \frac{\sum_{j=1}^{k-1} \sum_{j'=j+1}^{k} (\beta_j - \beta_{j'})^2}{k(k-1)}$$

It has also been derived in terms of the raters means, given that the rater effect (under the expressed conditions) is just the difference between the mean of the particular rater and the global mean, which was a step taken to proof the concordance with the other CCC proposals. The issue of this estimate is that in Searle, Casella, and McCulloch (1992) it was proven to be biased, which then led to the expression of the unbiased estimate in the matrix form $\hat{\sigma}^2_\beta = \frac{1}{k(k-1)}\{\hat{\beta}' LL' \hat{\beta} - trace(LL'\hat{\Sigma}_\beta)\}$, where L is the fixed effects contrast matrix and $\hat{\Sigma}_\beta$ is the variance-covariance matrix of the fixed effects[3].

The formula was first expanded in L. Lin, Hedayat, and Wu (2007) with a subject-rater interaction term (random effect) $\alpha\beta_{ij} \sim N(0, \sigma^2_{\alpha\beta})$, and then further expanded for repeated measurements, with particular attention to the longitudinal case (Carrasco, 2009), which led to the inclusion of a further subindex t to allow multiple replicates per individual per time. The Variance Components approach to estimate the longitudinal CCC with repeated measures rests on the following Normal-Normal Linear Mixed effects model:

$$Y_{ijtl} = \mu + \alpha_i + \beta_j + \gamma_t + \alpha\beta_{ij} + \alpha\gamma_{it} + \beta\gamma_{jt} + \epsilon_{ijtl}$$

where β_j and γ_t are the fixed effects related with method and time, μ is the overall mean, and the rest of the components are the random effects of the model (which are assumed to

[3]This correction for the CCC was later introduced (Carrasco, 2009), specifically for the longitudinal case.

be normal). For the estimation of the variance that comes from the systematic differences between observers across time (the fixed effects) a similar correction was applied: $\hat{\sigma}^2_{\beta\gamma} = \frac{1}{pk(k-1)}\{\hat{\beta}'LL'\hat{\beta} - trace(LL'\hat{\Sigma}_\beta)\}$, with p reflecting the number of times. The final formula is the following:

$$CCC = \frac{\sigma^2_\alpha + \sigma^2_{\alpha\gamma}}{\sigma^2_\alpha + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma} + \sigma^2_\epsilon}$$

3.3 A Bayesian Estimate

A proposal for the CCC was later developed under the bayesian approach in Bhattacharjee and Bhattacharyya (2015), which was then expanded in Feng, Baumgartner, and Svetnik (2015) and in Feng, Baumgartner, and Svetnik (2018) by defining robust priors to handle skewed data, and in Vanbelle and Lesaffre (2017) the distributions for bounded data. While this approach can be computationally expensive, and require an extensive number of iterations for the CCC to reach a reasonable effective sample size in the MCMC[4], and be burdened to misspecification due to the explicit definition of the model assumptions that are required, through a proper selection of selection of the priors, a bayesian CCC estimate might show to be more robust than the estimates were the assumptions are taken implicitly. For the purpose of this thesis, the estimate has been extended to the longitudinal case through the Variance Components approach previously explained, and the fitting of the Linear Mixed Model has fully implemented in R, only requiring the installation of R packages and without the requirement of external software (previous implementations required the installation of jags, but now the bayesian inference is carried through stan (2024))[5]. The hierarchical priors for the standard deviations and for the error term of the LMM are truncated at zero, in order to avoid negative values, t-student distributions with 3 degrees of freedom and a standard deviation of 2.5, while the fixed intercept μ has also a t-student prior, which it is not truncated, and has a default value of 5.8. The selection of the priors has purposely been confined to weakly informative priors, which were modified appropriately for the lognormal-normal model extension[6].

[4]Feng, Baumgartner, and Svetnik (2015) .

[5]There are alternative native implementations of the MCMC algorithm in R, as the *MCMCglmm* package, that would allow a comparable implementation.

[6]See following section.

4 Handling non-normality in Concordance

While the main line of research in the literature, and in this thesis, was to focus in finding robust estimators of the CCC to misspecification, there has been an important line of research into properly characterizing the underlying distribution of the data, mostly through a bayesian framework, as in Vanbelle and Lesaffre (2017) or in Feng, Baumgartner, and Svetnik (2018); but with a key frequentist contribution in Josep L. Carrasco (2009), where a version of the generalized concordance correlation coefficient through the variance components of a generalized linear mixed model was provided. The rest of this section discusses first the findings from this paper, to then expand it, as it underscores the simulation approach, which is detailed in its section.

4.1 Adapting the CCC distributional assumption

Another definition of the Concordance Correlation Coefficient[1], developed for the distributions of the exponential family, is the intraclass correlation between any measurement from different raters on the same subject:

$CCC = \frac{cov(Y_{ijl}, Y_{ij'l'})}{Var(Y_{ijl})}$, with the marginal variance can be further develop (through the law of total variance) into: $Var(Y_{ijl}) = var_u(\mu_{ij}) + E_u[\phi h(\mu_{ij})]$, with ϕ being the dispersion parameter. The marginal covariance can also be expanded, given the assumption that the effects that condition the observations are independent, as $cov(Y_{ijl}, Y_{ij'l'}) = cov_u(\mu_{ij}, \mu_{ij'})$.

This led to the definition of a general formula for the Generalized Concordance Correlation Coefficient: $GCCC = \frac{cov_u(\mu_{ij}, \mu_{ij'})}{var_u(\mu_{ij}) + E_u[\phi h(\mu_{ij})]}$.

Given that under a Normal distributed response variable the terms can be expressed in the form (as in Josep L. Carrasco, King, and Chinchilli (2009)): $cov_u(\mu_{ij}, \mu_{ij'}) = \sigma^2_\alpha + \sigma^2_{\alpha\gamma}$, $var_u(\mu_{ij}) = \sigma^2_\alpha + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma}$, and $E_u[\phi h(\mu_{ij})] = \sigma^2_\epsilon$ given $\phi = \sigma^2_\epsilon$ and $h(\mu_{ij}) = 1$ under the normal distribution for the longitudinal case. This leads to the general expression of $CCC = \frac{\sigma^2_\alpha + \sigma^2_{\alpha\gamma}}{\sigma^2_\alpha + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma} + \sigma^2_\epsilon}$.

[1] There are other alternate definitions of the CCC, as the inter-CCC, for inter-rater agreement, and intra-CCC, for intra-rater agreement, as defined in L. Lin, Hedayat, and Wu (2007), which are not discussed in this thesis. Using the same nomenclature, the CCC index discussed is the one of *Total Agreement*, which has also been referenced as *Overall CCC* (OCCC) in the literature.

For the Poisson and overdispersed Poisson case it can be developed into: $CCC = \frac{\mu(e^{\sigma^2_{\alpha}+\sigma^2_{\alpha\gamma}}-1)}{\mu(e^{\sigma^2_{\alpha}+\sigma^2_{\alpha\beta}+\sigma^2_{\alpha\gamma}+\sigma^2_{\beta\gamma}}-1)+\phi}$, with ϕ being the overdispersion parameter (when fixed equal to one gives the coefficient's estimate for the Poisson distribution). An extended review of this parametrization for Variance Components, U-statistics and GEE can be found in Tsai and Lin (2018).

Similarly for the Negative-Binomial distribution: $CCC = \frac{\mu(e^{\sigma^2_{\alpha}+\sigma^2_{\alpha\gamma}}-1)}{\mu((r+1)e^{\sigma^2_{\alpha}+\sigma^2_{\alpha\beta}+\sigma^2_{\alpha\gamma}+\sigma^2_{\beta\gamma}}-1)+\phi}$. An adaptation for the presence of zero-inflated cases is also provided (Carrasco, 2009).

4.2 The Lognormal-Normal Linear Mixed Model

While the previous extensions provided reasonable adaptations for count and overdispersed data, when handling skewed data, it has been proposed (Carrasco, 2007) to keep the normality of the response and of the residuals and assume that the subject effect follows a log-normal. Given that the variance formula of a log-normal random variable is the following $\sigma^2_{\alpha} = \exp(2\mu_{log\alpha} + \sigma^2_{log\alpha})(\exp(\sigma^2_{log\alpha}) - 1)$, with $\mu_{log\alpha}$ and $\sigma^2_{log\alpha}$ representing the mean and variance of the logarithm of α, then the non-longitudinal version of the coefficient is the following:

$$CCC = \frac{\exp(2\mu_{log\alpha} + \sigma^2_{log\alpha})(\exp(\sigma^2_{log\alpha}) - 1)}{\exp(2\mu_{log\alpha} + \sigma^2_{log\alpha})(\exp(\sigma^2_{log\alpha}) - 1) + \sigma^2_{\beta} + \sigma^2_{e}}$$

The methodological discussion behind its extension to the longitudinal case is whether the subject-method and subject-time interactions should be considered log-normal, or the skewness can it itself be captured by the subject effect (this has been the underlying assumption for generating skewed data for the simulations, as it offers a less complex analytical form, given that a sum of log-normals does not have a closed form[2]). The CCC formula for the latter would be the following:

$$CCC = \frac{\exp(2\mu_{log\alpha} + \sigma^2_{log\alpha})(\exp(\sigma^2_{log\alpha}) - 1) + \sigma^2_{\alpha\gamma}}{\exp(2\mu_{log\alpha} + \sigma^2_{log\alpha})(\exp(\sigma^2_{log\alpha}) - 1) + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma} + \sigma^2_{e}}$$

[2]While for the purposes of the simulations under skewness and misspecification the approach followed can be considered as sufficient, for the appropriate implementation of the the lognormal-normal estimation method, the proper distributional form of the interactions should be discussed.

4.3 The Skew-Normal Model

The last proposal to be reviewed can be seen as a hybrid approach between the correct specification and the robust estimators. In Feng, Baumgartner, and Svetnik (2018), the authors propose the skew-normal distribution (a generalization of the normal distribution that allows skewness), while maintain the standard CCC formula, after applying a correction on the estimated parameters, with $\mu_j^* = \mu_j + \sqrt{\frac{2}{\pi}}\delta_j$ and $\Sigma^* = \Sigma + (1 - \frac{2}{\pi})\delta\delta'$. The development for a skew-t-student distribution was also provided. As of 2024, while there has been no proposed extension to the longitudinal case, since june the underlying model required can now be fitted through the *glmmtmb* formula of the glmmtmb[3] R package (from a frequentist perspective) or through the function provided developed here for the bayesian method (which relies on the *brms*[4]implementation).

[3]Brooks et al. (2017)
[4]Bürkner (2017a)

5 Simulation & Methods

5.1 Methods

Across the literature, four main approaches have been identified in order to obtain robust estimates of the Concordance Correlation Coefficient and its standard error: The U-statistics approach, which was initially proposed in King and Chinchilli (2001b) and then implemented in R with the Josep Lluis Carrasco and Pena (2024) package; the Generalized Estimated Equations approach proposed in Barnhart, Haber, and Song (2002); the Variance Components approach proposed in Josep L. Carrasco and Jover (2003) (same implementation as the U-statistics[1]); and the Bayesian framework discussed in Feng, Baumgartner, and Svetnik (2015) and implemented in the **agRee**[2] R package. Given the interest of working with repeated measures and longitudinal data, the GEE approach has not been evaluated (given that we currently lack an implementation that fulfills our requirements), while the bayesian framework has been adapted in order to handle a longitudinal response, and to not require external applications.

The Variance Components approach has been split into two categories with regards to the estimation of the standard deviation of coefficient. The first one of those is the one initially presented in L. I.-K. Lin (1989) for the bivariate case ($m = 2$, which I'll be refering as the F2 transformation), taking advantage of the properties of the delta method, which then it could also be transformed, using Fisher's[3] z-transformation $z = \frac{1}{2}\ln(\frac{1+(m-1)\rho_{ccc}}{1-\rho_{ccc}})$, or to the modified Z-transformation by Konishi and Gupta (1987), $z = \sqrt{\frac{m-1}{2m}}\ln(\frac{1+(m-1)\rho_{ccc}}{1-\rho_{ccc}})$ which has a slight bias in the mean of $\frac{7-5m}{N\sqrt{18m(m-1)}}$ and a variance of $\frac{1}{N}$. This later transformation was recommended in Donner and Zou (2002) for situations with more than two replicates (m) per individual. The second category refers to the usage of resampling techniques in order to obtain the estimates of the standard error and construct confidence intervals for the parameter. Within this field different resampling techniques have been proposed, as the jackniffe in Feng, Baumgartner, and Svetnik (2013), but the analysis has been restricted to bootstrap implementations, very popular in the literature where they have been a common comparison to any proposal and obtaining consistently comparable results[4]. Both parametric

[1] Josep L. Carrasco et al. (2013)
[2] Feng (2020a)
[3] Fisher (1925)
[4] Barnhart, Haber, and Song (2002)

and non-parametric bootstrap[5] implementations were considered, and for their confidence intervals both percentile and Bias-Corrected-and-Accelerated[6] (Bca) were tested.

The Bayesian implementation is based (the `ccc_vc_bayes` function, which is also available as an .R file) in the approach in Feng, Baumgartner, and Svetnik (2015) (which also used the variance components estimate), which I extended to consider the longitudinal case and the non-longitudinal case with the subject-method interaction. I first developed the model all the random effects under a normal distributional assumption (information about the priors in the next section), and then I also considered the case where the subject effect followed a lognormal distribution (I extended the lognormal–normal linear mixed model proposed in Josep L. Carrasco et al. (2007)). Given that I expected that I already expected that the standard deviation of the log subject effect was going to be relatively small (« generally smaller than 1), I decided to choose a $\sigma_{\ln_\alpha} \sim \Gamma(1.5, 2)$ prior[7] in order to avoid boundary issues close to 0. The function acts as a wrapper of the **brms**[8] R package functions `stancode` and `standata`, which produce a linear mixed model *stan* file and adapt the data into a *stan* readable format, respectively, which are then internally edited to conform with the CCC definition as it can be seen in the *cccrm* R package (one of the key modifications that has to be made for the stan model to comply with the underlying LMM of the CCC derived through Variance Components is that the variance-covariance matrix of the random effects is assumed to be a block diagonal matrix with an identity condition).

5.2 Design

Concordance in each simulation set has been evaluated at 5 levels (at 0.1, 0.3, 0.5, 0.7, and at 0.9), in order to evaluate whether there is some differential performance close to the coefficient's bounds. Each simulation contains one thousand replicates. In all cases the number of bootstrap resamples has been five hundred. The following estimation methods were evaluated: one asymptotic approach with four methods for interval estimation (No transformation, Fisher's transformation, F2 transformation, Konishi-Gupta's transformation), two bootstrap approaches (parametric and non-parametric) with two methods for interval estimation (empirical and BCa), the U-statistics approach, and a Bayesian Normal-Normal approach[9]. The four simulation sets included evaluating the default longitudinal case of two observers in two times plus two replicates per combination of observer and time (which leads to a complete design where each subject was evaluated eight times). The first two simulation sets were generated under a Normal-Normal model (both the subjects' effect and the response are normally distributed), with the objective to review the impact of sample size (30 vs 120 individuals).

[5] B. Efron (1979)
[6] Bradley Efron (1987)
[7] Based on a recommendation by Gelman. See http://www.stat.columbia.edu/~gelman/presentations/wipnew2_handout.pdf.
[8] Bürkner (2017b)
[9] Its convergence was assessed during implementation through the Gelman-Rubin statistic.

The third and fourth simulation sets reflect on the inclusion of a lightly and a heavily skewed subjects' effect through a Lognormal-Normal design in the overall estimation.

For all simulation sets, the bayesian method's simulations were run on a ten thousand to one thousand iterations to burnin split, in order to produce a meaningful effective sample size for the coefficient. Nonetheless and upon review, while convergence is guaranteed, divergent transitions still have appeared. On further testing the problem seems to be solved by increasing the iterations to be discarded (burnin) to fifteen hundred.

5.2.1 First set

Longitudinal design. N-N model, 1000 simulations. 30 individuals with 2 observers, 2 times, and 2 replicates.

Table 5.1: Coverage (%) of each method at each concordance level. (continued below)

	AN	AF	AF2	AKG	BN	U	BPB	BPE
CCC_01	95.21	95.49	94.66	95.67	99	83.8	98.82	81.87
CCC_03	94.97	94.86	95.34	94.06	96.8	78.8	95.7	90.11
CCC_05	94.15	95.28	94.37	94.68	97.4	77.6	95.8	90.86
CCC_07	95.77	94.89	95.06	95.08	99.1	72.3	94.52	91.01
CCC_09	95.95	94.83	94.63	94.14	99.7	68.8	92.99	92.05

	NBPB	NBPE
CCC_01	96.5	79.53
CCC_03	91.74	87.16
CCC_05	89.95	88.43
CCC_07	90.2	88.38
CCC_09	90.22	89.31

AN: Asymptotic method (No transformation), AF: Asymptotic (Fisher's transformation), AF2: Asymptotic (F2 transformation), AKG: Asymptotic (Konishi-Gupta transformation), BN: Bayesian Normal estimate, U: U-statistics estimate, BPB: Parametric Bootstrap (BCa), BPE: Parametric Bootstrap (Empirical), NPBP: Non-parametric Bootstrap (BCa), NPBE: Non-parametric Bootstrap (Empirical).

Every confidence interval method based on the asymptotic estimate achieved the nominal coverage, while for the bootstrap approaches, which resample the asymptotic estimate, only under the Bias Corrected and accelerated method the nominal coverage was achieved. Both the robust U-statistics approach and the Bayesian N-N method did not fail to produce an estimate in any of the simulations, nonetheless, showed by far the worst coverage while the latter was the best performer. Comparatively the asymptotic model failed in 5% of the simulations for the 0.1 CCC and in 1% of the simulations for the 0.9 CCC. The bootstrap methods had a higher degree of convergence failure across the board, only being comparable at 0.1 with a failure rate of 5%, which grew to 22% failure rate for a CCC of 0.9. The non-parametric approach with the available version on the package failed at a larger rate, which will be updated. The main cause of the failures of the methods that relied on the asymptotic estimate were due to failures of the `lme` model in computing the Hessian matrix.

5.2.2 Second set

Longitudinal design. N-N model, 1000 simulations. 120 individuals.

Table 5.3: Coverage (%) of each method at each concordance level. (continued below)

	AN	AF	AF2	AKG	BN	U	BPB	BPE
CCC_01	95.95	94.59	93.87	95.88	96.8	84.1	96.31	92.64
CCC_03	94.4	94.22	95.05	95.31	98.5	81.5	94.15	93.92
CCC_05	95.31	95.42	95.17	93.89	97.6	77.5	94.01	92.31
CCC_07	95.13	95.52	94.96	95.44	98.6	76.1	95.82	94.69
CCC_09	95.24	95.18	94.98	94.85	91.1	75.4	94.41	94.85

	NBPB	NBPE
CCC_01	94.7	89.8
CCC_03	95.4	92.9
CCC_05	93.79	91.99
CCC_07	93.09	92.79
CCC_09	94.19	91.4

Under this set the U-statistics method still does not achieve its nominal coverage while the empirical bootstrap estimates show a minor improvement (even if they still fail to reach a single time the nominal coverage). The bootstrap approaches showed high levels of failure rates close to the bounds (between 11 and 13% at 0.1, and between 66% and 69% at 0.9), an issue that was later addressed in the implementation.

5.2.3 Third set

LN-N model, 1000 simulations. 30 individuals.

For the remaining two simulation sets, the failure rate of the bootstrap implementation was meaningfully addressed (at 0.1 the failure rate dropped to 5% and at 0.9 it dropped to 22%, see Appendix), which nonetheless still remains as a key issue regarding this implementation. The inclusion of the lognormal random effect (strongly skewed itself, but with the response still being lightly skewed) does not seem to meaningfully affect the methods' performance, whose coverages are comparable to those seen previously and carrying the same problems.

Table 5.5: Coverage (%) of each method at each concordance level. (continued below)

	AN	AF	AF2	AKG	BN	U	BPB	BPE
CCC_01	92.71	95.16	94.14	94.11	98.26	83.8	97.77	78.22
CCC_03	95.2	95.83	95.03	94.83	97.2	79.2	94.93	89.94
CCC_05	95.25	93.57	95.27	95.85	98.2	77.5	94.5	91.23
CCC_07	95.99	95.99	94.97	95.58	98.8	72	95.43	90.32
CCC_09	95.36	94.93	95.25	94.23	99.7	70.3	93.19	92.64

	NBPB	NBPE
CCC_01	96.09	76.85
CCC_03	91.07	88.5
CCC_05	90.01	89.4
CCC_07	91.23	89.04
CCC_09	90.97	89.25

5.2.4 Fourth set

Longitudinal design. LN-N model, 1000 simulations. 30 individuals.

For this final set the skewness of the lognormal effect was increased as such that the response was also heavily skewed (see Appendix for a detailed explanation on the procedure). All methods fail across the board in reaching the nominal coverage, but both the Bayesian and the non-parametric bootstrap Bca showed better coverage than the rest.

Table 5.7: Coverage (%) of each method at each concordance level. (continued below)

	AN	AF	AF2	AKG	BN	U	BPB	BPE
CCC_01	7.573	7.48	8.205	5.787	30.3	7.1	7.211	4.888
CCC_03	11.55	10.91	9.218	9.71	39.5	10.1	14.71	7.708
CCC_05	15.23	15.93	15.5	13.43	56.7	13.3	18.07	10.17
CCC_07	18.4	18.7	20.1	17.4	76.6	14	22.87	11.32
CCC_09	25.93	25.68	24.12	23.72	87.1	18.3	27.24	12.6

	NBPB	NBPE
CCC_01	36.56	5.632
CCC_03	53.44	8.021
CCC_05	61.43	10.47
CCC_07	67.23	11.82
CCC_09	74.36	13.48

5.3 Results

The four confidence interval construction methods for the asymptotic coefficient showed comparable coverage across the entire set. Given that the F2 transformation was proposed only for two replicates per individual, while both the Fisher's and the Konishi-Gupta transformation were generalized for any number of consistent replicates across individuals, it is still an important finding that the F2 still held for eight replicates. The robust U-statistics approach failed to provide its nominal coverage in all four sets and performed only comparable with the asymptotic approach in the heavily skewed case. The bootstrap approaches, which resampled the asymptotic CCC estimate, were the second most computationally expensive and those that showed the largest failure rate with regards of the estimation, and even after a revision to include parallelized batching and better error handling, while the improvement was notable the issues still remained.. The Bayesian approach was the most computationally expensive by a wide margin, and it requires a rather large number of iterations (and iterations to discard) in order to converge in a reasonable effective sample size, but showed some of the best performance (only outperformed by the non-parametric bootstrap in specific cases) even when failing to reach its nominal value. The use of a MAP (Maximum a posteriori probability) estimate as a point estimate did not produce a larger bias than the asymptotic or U-statistics point estimate, and it had the second lowest failure rate with regards of estimation. A Bayesian Lognormal-Normal estimate was also implemented but it requires further testing, and the approach could then be extended to handle other distributions.

Tables with the failure rates, mean skewness and average bias have been provided in the appendix.

6 Case Study 1: Microbioal tooth data

While simulation methods offer a computationally expensive but insightful approach on the misspecification of the distributional assumption and its impact, which couldn't have been easily derived from real data, both for its required volume and the large spectrum of parameters to be tested; the review of such a strong and powerful assumption, and its possible misspecification, with real world data is indispensable, as it currently is the field of application of the coefficient. Thus, two datasets have been provided and reviewed. The first has been used as an example of extreme non-normality in a non-longitudinal context, with the assessment on the impact of the transformation of the response variable, while the second dataset offers a longitudinal example with a relevant degree of drop-out, and a certain degree of skewness under one of the methods, whose impact has been explored for the estimates.

The following dataset comes from a pending to be published study[1], where the bacterial concentration information of a pair of molars of a total of 33 pediatic patients was collected. These molars were assigned their treatment method following a split mouth design, and the DNA information was collected for each molar using a Qiagen Kit, which allowed to measure the count of colony forming units in the sample, which in the case of this example dataset was of anaerobic bacteria. The dataset also contains information on the solution dilution volume used in the DNA extraction of each sample, and the total count of anaerobic bacteria found, which are necessary parameters to compute the CFU value, and as such they cannot be considered as candidate covariates.

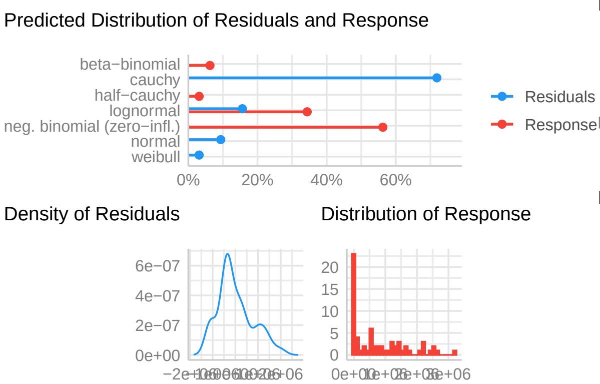
Given the highly non-normal, nature of the Colony Forming Units value, the estimation methods' robustness to a strong violation of the normality assumption has been tested, and the usefulness of the log-transformation, one of the main methods suggested to further normalize non-normal data[2], has also been assessed.

In order to assess the degree of failure of the distributional assumption for the response variable and the error term under the proposed Linear Mixed Model, I have used the `check_distribution` function from the R package **performance**[3], a large library focused on model testing tools to review model quality and fit, in order to provide an alternative perspective on the ill-definition of the distribution assumption, beyond visual inspection and the common statistical tests for normality (Shapiro-Wilks, and the non-parametric Kolgomorov-Smirnov). The relies on a Random Forest model in order to give a prediction on

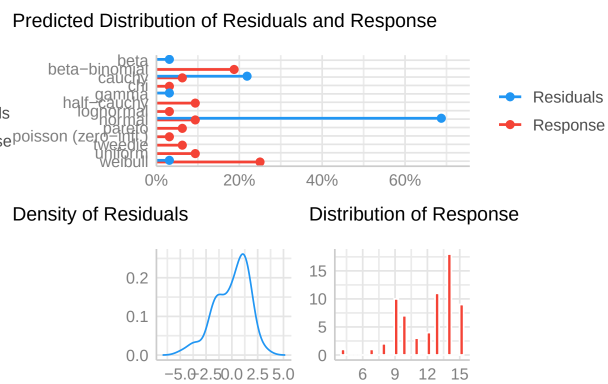
[1] The reference will be theb updated.
[2] Pek, Wong, and Wong (2018) .
[3] Lüdecke et al. (2021) .

the distribution of the data (both response and model residuals) given the key statistics (e.g., mode, mean, median, kurtosis, or skewness) of such data.

It is already apparent, given the histogram of the CFU response, that its distribution is rather closer to an exponential than to a normal distribution (with the exponential being the epitome of non-normal distributions). Also the distributional assumption seems to fail for the model residuals, which also goes against one of the core assumptions with the formulation generally derived for the Concordance Correlation Coefficient when derived from a Linear Mixed Model, where the inference about the error term comes from the interpretation that it represents the measurement error, and thus it should be normally distributed. Then the application of the log-transformation (a monotone transformation) for the CFU values has been a common resource used to further normalize the data before carrying statistical analysis[4], and not without its fair share of criticism[5] [6]. In the right figure, the assessment for the log-transformed response and for the model's error term is shown. While the data seems to be further normalized, more clearly in the residuals case, the log CFU histogram shows a rather bi-modal distribution, which might even question the LMM approach for the concordance correlation coefficient, and thus its validity.



(a) CFU distribution under the Linear Mixed Model



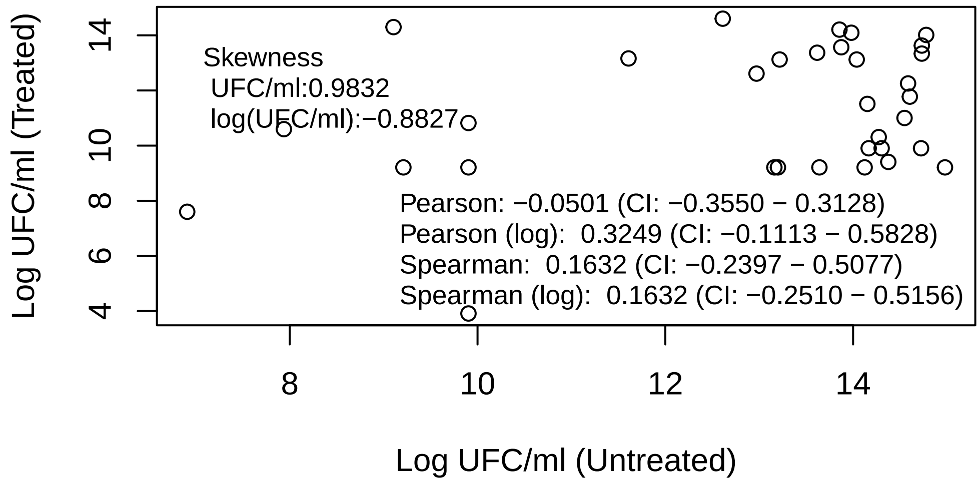
(a) log CFU distribution under the Linear Mixed Model

Response and residuals predicted distributions

Given the definition by L. I.-K. Lin (1989) of the Concordance Correlation Coefficient as a product between the Pearson's correlation coefficient ρ and the bias correction factor $0 < C_b \leq 1$ ($CCC = \rho C_b$), the CCC can be viewed as a stricter version of the correlation coefficient, with which it shares the normality assumption. As such non-robust point estimates of the CCC might suffer the identical problem as the correlation coefficient, which for the non-log transformed response variable finds the methods uncorrelated while for the transformed response the point estimate of correlation is low to medium. In comparison the rank-based

[4] Shirato et al. (2022) .
[5] (n.d.) .
[6] Wheatland (2022) .

Spearman correlation coefficient provides by design a transformation-invariant estimate of the correlation, which can be interpreted as the upper bound of the CCC point estimates, given its definition. It is also poignant to indicate that the log-transformation has not reduced the magnitude of the moderate skewness that was already present in the untransformed data, which will limit the reviewed methods' performances, as seen in the simulation section.



The rest of the review of this case study centers not on the appropriateness of the transformation, but on the study of its impact for the reviewed estimates.

6.1 Results

When working with the untransformed variable, the variance components obtained through the model are themselves levels of magnitude apart (the estimated subject variance was in the order of `e+04` while the estimated methods variance and error term were in the order of `e+11`), which leads to rather extremely small estimated values for the Concordance Correlation Coefficient (for which in the case of the U-statistic estimate even gives a negative value outside of the coefficient's range). The issues with the estimation of the CCC in this dataset also continue when providing 95% confidence intervals (or 95% highest density intervals for the bayesian case), for which the asymptotic method both in its base form and when it relies upon Fisher's Z-transformation, it provides negative lower bound estimates (outside of the bounds of the coefficient), which is also the case for the empirical bootstrap intervals and the U-statistic estimate. The problem observed with the confidence interval after applying the Konishi-Gupta transformation, is that the coefficient estimate lies outside of the CI provided. In that sense, the only non-ill-behaved estimates correspond to the bootstrap BCa estimates (both parametric and non-parametric), and the Bayesian model estimate provided. The other Bayesian implementation reviewed (Normal-Normal and Skewnormal-Normal from the agRee[7]

[7]Feng (2020b)

R package) also fails to produce estimates within the bounds of the coefficient.

Table 6.1: CCC for CFU/ml model

	CCC	LL CI 95%	UL CI 95%	SE CCC
Asymptotic	5.073e-08	-1.344e-05	1.354e-05	6.885e-06
Asymptotic (Fisher's Z)	5.073e-08	-1.344e-05	1.354e-05	6.885e-06
Asymptotic (Z, m=2)	5.073e-08	-1.344e-05	1.354e-05	6.885e-06
Asymptotic (KG transf)	5.073e-08	0.01514	0.01516	6.885e-06
Param Boot BCa	5.073e-08	1.124e-08	0.3248	0.07331
Param Boot Emp	5.073e-08	-0.2381	9.681e-08	0.07221
Non-Param Boot BCa	5.073e-08	1.486e-08	0.2524	0.05034
Non-Param Boot Emp	5.073e-08	-0.1784	9.553e-08	0.05043
U-stat	-0.03049	-0.2377	0.1794	0.108
N-N Bayesian (MAP)	0.0005366	2.027e-09	0.2384	0.08048
N-N Bayesian (agRee)	-0.0288	-0.2468	0.1935	NA
Skew-N Bayesian (agRee)	-0.004061	-0.1704	0.1693	NA

After applying the log-transformation, the estimated variance components are of a similar magnitude, and the point estimate for the coefficient has grown across the board (independent of the estimation method). Nonetheless, the lower bound provided by the asymptotic estimate (even after applying Fisher's Z-transform) still falls outside the coefficient's domain. A second point of relevance has been the stabilization of the estimate of the standard deviation of the coefficient, which has grown comparatively much more for the asymptotic method, while for the Bayesian and bootstrap cases only doubled, and it remained reasonably consistent for the U-statistic estimate. Finally, the MAP (Bayesian) estimate of the coefficient clearly diverges from the rest, even when it provides a similar upper bound to other methods, while it remains the most consistent with its estimate for the non-transformed case. The only other estimate close, but still over the samples' spearman correlation coefficient is the skewnormal-normal bayesian estimate, which would indicate that even after the transformation, which effectively shifted the skewness but failed to reduce its magnitude, the generalization approach for the normal distribution, might still not be appropriate.

Table 6.2: CCC for log(CFU/ml) model

	CCC	LL CI 95%	UL CI 95%	SE CCC
Asymptotic	0.2616	-0.0005759	0.5237	0.1337
Asymptotic (Fisher's Z)	0.2616	-0.0136	0.4999	0.1337
Asymptotic (Z, m=2)	0.2616	-0.0136	0.4999	0.1337
Asymptotic (KG transf)	0.2616	0.001547	0.5112	0.1337
Param Boot BCa	0.2616	0.01546	0.4862	0.1211

	CCC	LL CI 95%	UL CI 95%	SE CCC
Param Boot Emp	0.2616	0.006969	0.5231	0.1288
Non-Param Boot BCa	0.2616	7.028e-09	0.4886	0.1334
Non-Param Boot Emp	0.2616	0.01251	0.5231	0.1365
U-stat	0.2557	0.01516	0.4682	0.1175
N-N Bayesian (MAP)	0.002884	3.778e-11	0.4847	0.1503
N-N Bayesian (agRee)	0.254	0.001095	0.5077	NA
Skew-N Bayesian (agRee)	0.1776	-0.05301	0.4336	NA

7 Case Study 2

The following dataset offers a longitudinal profile of 139 patients with a drop-out mechanism (the minimum number of visits were 3 and the maximum were 12) for the comparison of two techniques, with the addition of potential covariates whose inclusion or lack there off, and even how they should be included raises interesting methodological questions.

It comes from the study Moreno et al. (2018), where the objective was to evaluate the evolution of possible gadolinium brain deposits in the patients under treatment through two MRI techniques, which have been seen superior in comparison with CT scans for distinguishing between soft tissues and malignancies, and in their lower inter-observer variability[1]. The first of the MRI techniques is T1 weighted image, which is based in the observation of differences in the T1 relaxation times of tissues after the application of the contrasting agent (which in this case was a gadolinium based contrast agent), while the second is the T2 weighted image, that as its name suggest captures a different pulse sequence on MRI (focusing on the long repetition and echo times), in this case the differences on the T2 relaxation times (which also reacts differently to the contrasting agent)[2]. While the T1Wi technique was the main technique through which the dentate nucleus identification was done, the T2Wi was used as an automatic check for the correct identification, and in the cases of non-concordance (those cases seen as doubtful) the decision was made by consensus.

The metric of interest in this case study will be the ratio of the weighted image across the dosages (which in the study, in a secondary analysis, was used to study the dentate nucleus visibility). The dataset also contains information of the machines used for each contrast and the practitioner that conducted the evaluation (plus whether the dentate nucleus was considered visible or not). While it would also be interesting to study the as methods the machine or the practitioner effect, both metrics lack the full design in comparison with the MRI techniques, which were both applied in every patient in every visit.

[1]Bahloul et al. (2024)
[2]Preston and Shapiro (2009)

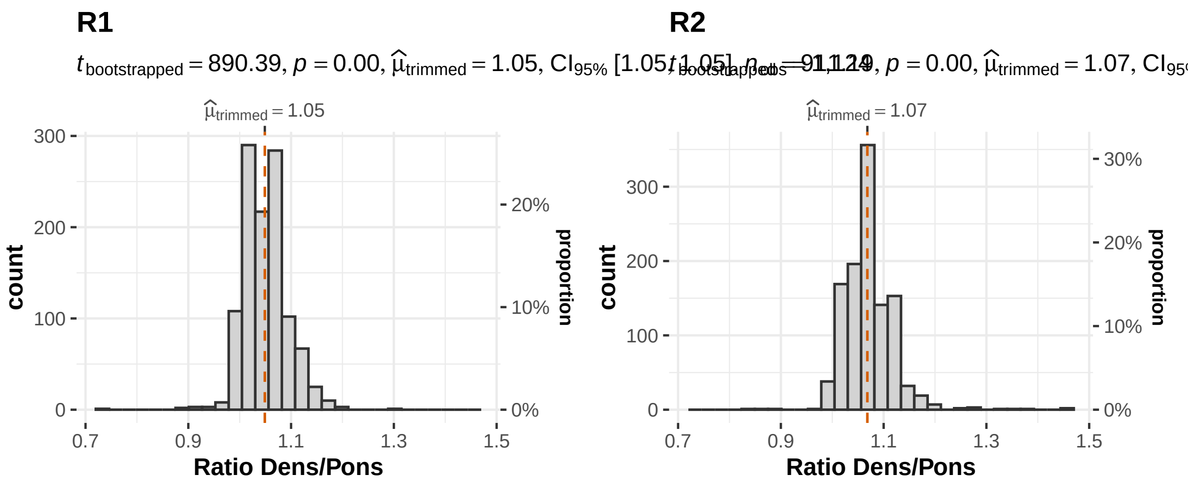


Figure 7.1: Distribution of the ratio between techniques

7.1 Results

First, a comment on the methods implementation and observed issues when computing the Concordance Correlation Coefficient. On their current implementation in the `cccrm` package, both the parametric bootstrap implementation from the `ccc_vc` function and the U-statistics estimate implemented in the `cccUst` function, showed problems handling missing observations. The issue for the parametric bootstrap case can be solved with dropping the missing observations and working with the available cases, while for the case of the U-statistics estimate, its implementation works under the assumption of equal number of observations per subject, and thus it requires a deeper review. Thus, I reflected the estimates for three levels of complete cases: those that went through at least 6 visits (106 patients), those that reached the 9th visit (72 patients), and those that reached the penultimate visit (only 17 patients). While the presence of a drop-out mechanism in the response variable should not by itself suggest any issue with the Concordance Correlation Coefficient estimate, in the degree that the non-observed samples (MNAR) or the omitted individuals (MAR) had shown a differential degree of concordance between T1Wi and TW2i, the Linear Mixed Models assumptions may not hold, and as it was used to provide the coefficients estimates these might also not hold.

When reviewing the results, beyond the limitation of the U-statistics implementation, first is noticeable how the MAP estimate[3] falls lower than the estimates provided by the other methods, while showing comparable (but larger) bounds for the coefficient. This is also can be noticed in the estimate of the standard error provided, which is clearly larger for the Bayesian

[3]With 10000 iterations and 1000 of the used for burnin, the CCC parameter obtained a potential scale reduction factor of 1.0013092, and an effective sample size of 464.9866581.

and the non-parametric bootstrap estimates, while the asymptotic method and the parametric bootstrap remain more consistent.

	CCC	LL CI 95%	UL CI 95%	SE CCC
Asymptotic	0.1942	0.1226	0.2658	0.03651
Asymptotic (Fisher's Z)	0.1942	0.1294	0.2722	0.03651
Asymptotic (Z, m=2)	0.1989	0.1257	0.27	0.03686
Asymptotic (KG transf)	0.1942	0.1313	0.2749	0.03651
Param Boot BCa (AC)	0.1942	0.1155	0.2616	0.03815
Param Boot Emp (AC)	0.1942	0.1273	0.2727	0.0375
Non-Param Boot BCa	0.1942	0.08988	0.2657	0.04436
Non-Param Boot Emp	0.1942	0.1228	0.2937	0.04542
U-stat (AC)	NA	NA	NA	NA
U-stat (CC-6)	0.2364	0.01214	0.4381	0.1102
U-stat (CC-9)	0.2581	0.04362	0.4498	0.105
U-stat (CC-11)	0.1488	-0.01856	0.308	0.08404
N-N Bayesian (MAP)	0.1308	0.09267	0.3339	0.0727

The inclusion of the two possible covariates, the practitioner and the machine used in the evaluation, while for all methods reduce the CCC point estimate, their inclusion as fixed effects do not dramatically shift the coefficient's confidence intervals. Given the experiment design it is up to discussion whether the standard inclusion of covariates as fixed effects was appropriate, or whether they should have been modeled as random effects.

	CCC	LL CI 95%	UL CI 95%	SE CCC
Asymptotic	0.1767	0.1033	0.2501	0.03744
Asymptotic (Fisher's Z)	0.1767	0.1112	0.2577	0.03744
Asymptotic (Z, m=2)	0.1841	0.1086	0.2575	0.03803
Asymptotic (KG transf)	0.1767	0.1129	0.2604	0.03744
Param Boot BCa (AC)	0.1767	0.09655	0.2493	0.03923
Param Boot Emp (AC)	0.1767	0.1072	0.2557	0.03852
Non-Param Boot BCa	0.1767	0.06093	0.2507	0.04985
Non-Param Boot Emp	0.1767	0.1031	0.279	0.04679
N-N Bayesian (MAP)	0.1058	0.05965	0.3079	0.07206

7.2 Missingness pattern

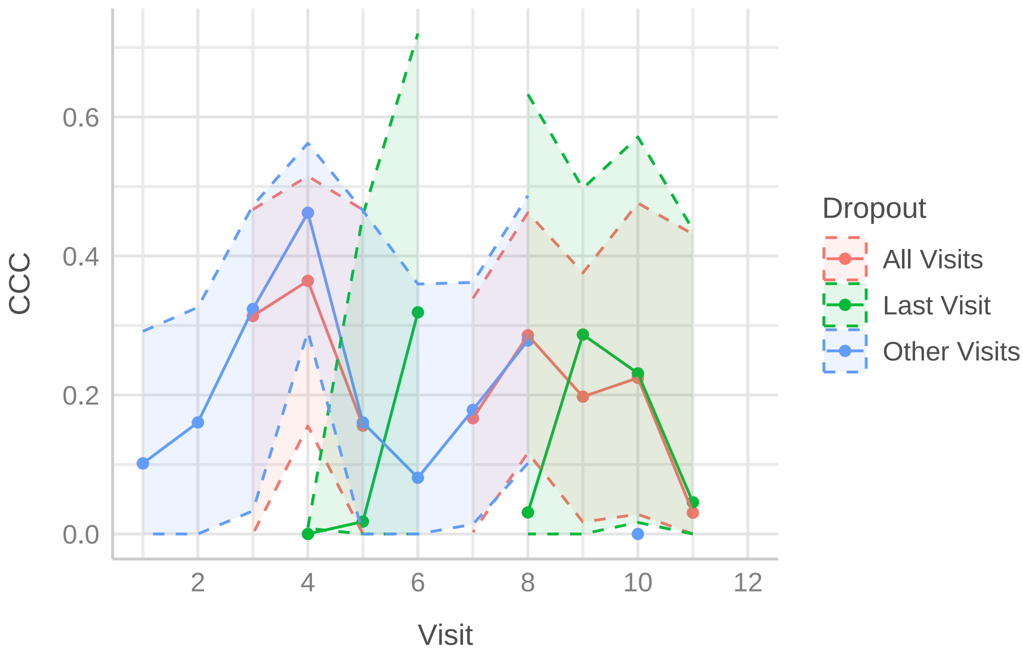
It felt rather important to review, for the adequacy of the application of the Complete Case Analysis for the concordance whether the dropout pattern in the dataset had any relationship with the degree of concordance between pairs, as if the observed pairs excluded in the complete case analysis presented a different degree of concordance (significantly more or less concordant) than the remaining pairs, we would be under a Missing At Random pattern, and the complete cases estimate would be biased. In order to review it, and with the awareness that the concordance was also time dependent, a comparison between the concordance of the pairs which were at the last visit of the subject versus the rest of the pairs was done for each visiting time. Then, the separation between the 95% confidence intervals at each time was assessed, with the only meaningful difference in Concordance happening at the fourth visit (where the nine subjects that dropped out had highly discordant samples). There was also no significant difference in concordance at each time when comparing the concordance of all available pairs versus the subset of pairs of continuing subjects, which would indicate that for this particular dataset the limitation of a Complete Case Analysis for the concordance correlation coefficient will not come from the missingness mechanism of the present dropout, but on the selection of the cut-off visit. To produce comparable estimates for methods that require complete cases to those under an Available Case framework, imputation strategies should be reviewed.

Table 7.3: Table continues below

Visit	Missing	Last_Visit	CCC	CCC_LV	CCC_Other
1	0	0	0.1016	-	0.1016
2	0	0	-	-	0.1605
3	0	2	0.3137	-	0.3237
4	2	9	0.3643	2.892e-10	0.4622
5	11	14	0.1558	0.01786	0.1605
6	25	13	-	0.3189	0.08091
7	38	7	0.1666	-	0.1785
8	45	17	0.2859	0.031	0.2784
9	62	31	0.1975	0.2869	-
10	93	27	0.2245	0.2311	7.853e-09
11	120	15	0.03061	0.04555	-
12	135	4	-	-	-

sig_ccc
-
-
-
Signficantly Less Concordant
No significant differences
No significant differences
-
No significant differences

sig_ccc
-
No significant differences
-
-



8 Limitations & Conclusions

8.1 Limitations

While the simulation sets reviewed represented a thorough analysis on the nominal coverage of the current estimation approaches for the Concordance Correlation Coefficient, which were complimented with the inclusion of two case studies, there were many relevant situations still pending to be tested. The analysis therefore could be expanded by selecting a different number of subjects, raters, and observation times, simulating with alternative weights for the variance components and other levels of skewness. With regards of the model structure, the reviewed CCC approaches could be compared with others that assume the fixed effects to be random, or that the skewness present in the longitudinal response is not only produced by the subjects-effect but also by the subject-time or subject-method interactions. Known approaches, such as skewnormal-normal models and generalized estimating equations, have yet to be extended for the repeated measures CCC, but have shown properties relevant to address moderate skewness.

The simulations were limited to longitudinal designs without missingness, but as seen in the second case study, the introduction of a drop-out pattern could hamper some of the available methods, and the impact of popular imputation techniques with regards to concordance has yet to be studied. Also, possible misspecification is not limited to the skewness produced by a log-normal effect, and the effect of under and overspecification on the underlying model was not reviewed in the present simulation sets.

8.2 Conclusions

The variance component approach is currently one of the most well developed and extensive approaches in order to assess the concordance between multiple raters in a repeated measures context, with particular emphasis in longitudinal outcomes. Most approaches reviewed were or have been included in the R package `cccrm`.

The application of transformations in the context of the presented simulation sets and the case studies do not seem to provide any sizable improvement over the default asymptotic approach. Both the empirical bootstrap and U-statistics methods generally failed to achieve its nominal coverage, while both the BCa bootstrap and the bayesian implementation based on the normal-normal model showed reasonable improvements in coverage, even if point estimate bias remained comparable. Another key insight is that the relevant misspecification is not the strength of skewness of the random effect itself, but of the response variable, which when it presents a strong skewness all tested methods fail to reach its nominal coverage. For the simulated right skewness better coverage was seen for high concordance values.

In the first case study it was observed that the application of the log-transformation on the response variable may not be an appropriate technique when the objective is to assess concordance, as most concordance estimates on the log-transformed response were greater than the estimated correlation, with the correlation being a natural upper bound for the CCC. Then, the second case study reflects that for the longitudinal assessment of concordance the selection of the cut-off points and covariates are meaningful choices, and when a drop-out mechanism is present and it is not at MNAR pattern, methods that model on available cases should be preferred.

Acknowledgements

This thesis was developed during an Erasmus stay in the Katholieke Universiteit Leuven, were professors Sophie Van Belle (University of Maastricht) and Emmanuel Lessafre (KU Leuven) acted as co-advisors, monitored the project, and provided key inputs for which I'm deeply grateful. I'm also very thankful for the support and time committed by professor Josep Lluís Carrasco in support of the project as the thesis director, from the its proposal and through its revisions.

Supplementary material

Both functions that have been implemented and used for the project have been made available in the following repository. The saved simulation results and the rest of the project code is also available.

References

n.d.

Armour, Marc, Joanne Michou, Imogen Schofield, and Karla Borland. 2024. "Agreement of Doppler Ultrasound and Visual Sphygmomanometer Needle Oscillation with Invasive Blood Pressure in Anaesthetised Dogs." *Animals* 14 (19): 2756. https://doi.org/10.3390/ani14192756.

Bahloul, Mohamed A., Saima Jabeen, Sara Benoumhani, Habib Abdulmohsen Alsaleh, Zehor Belkhatir, and Areej Al-Wabil. 2024. "Advancements in Synthetic CT Generation from MRI: A Review of Techniques, and Trends in Radiation Therapy Planning." *Journal of Applied Clinical Medical Physics* 25 (11). https://doi.org/10.1002/acm2.14499.

Barnhart, Huiman X., Michael Haber, and Jingli Song. 2002. "Overall Concordance Correlation Coefficient for Evaluating Agreement Among Multiple Observers." *Biometrics* 58 (4): 1020–27. https://doi.org/10.1111/j.0006-341x.2002.01020.x.

Bhattacharjee, Atanu, and Tapesh Bhattacharyya. 2015. "Bayesian Concordance Correlation Coefficient with Application to Repeatedly Measured Data." *Turkiye Klinikleri Journal of Biostatistics* 7 (2): 55–62. https://doi.org/10.5336/biostatic.2015-46702.

Brooks, Mollie E., Kasper Kristensen, Koen J. van, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler, and Benjamin M. Bolker. 2017. "{glmmTMB} Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling" 9. https://doi.org/10.32614/RJ-2017-066.

Bürkner, Paul-Christian. 2017b. "{Brms}: An {r} Package for {Bayesian} Multilevel Models Using {Stan}" 80. https://doi.org/10.18637/jss.v080.i01.

———. 2017a. "{Brms}: An {r} Package for {Bayesian} Multilevel Models Using {Stan}" 80. https://doi.org/10.18637/jss.v080.i01.

Carrasco, Josep L. 2009. "A Generalized Concordance Correlation Coefficient Based on the Variance Components Generalized Linear Mixed Models for Overdispersed Count Data." *Biometrics* 66 (3): 897–904. https://doi.org/10.1111/j.1541-0420.2009.01335.x.

Carrasco, Josep L., and Lluís Jover. 2003. "Estimating the Generalized Concordance Correlation Coefficient Through Variance Components." *Biometrics* 59 (4): 849–58. https://doi.org/10.1111/j.0006-341x.2003.00099.x.

Carrasco, Josep L., Lluis Jover, Tonya S. King, and Vernon M. Chinchilli. 2007. "Comparison of Concordance Correlation Coefficient Estimating Approaches with Skewed Data." *Journal of Biopharmaceutical Statistics* 17 (4): 673–84. https://doi.org/10.1080/10543400701329463.

Carrasco, Josep L., Tonya S. King, and Vernon M. Chinchilli. 2009. "The Concordance Correlation Coefficient for Repeated Measures Estimated by Variance Components." *Journal of*

Biopharmaceutical Statistics 19 (1): 90–105. https://doi.org/10.1080/10543400802527890.

Carrasco, Josep Lluis, and Gonzalo Peon Pena. 2024. "Cccrm: Concordance Correlation Coefficient for Repeated (and Non-Repeated) Measures."

Carrasco, Josep L., Brenda R. Phillips, Josep Puig-Martinez, Tonya S. King, and Vernon M. Chinchilli. 2013. "Estimation of the Concordance Correlation Coefficient for Repeated Measures Using SAS and R." *Computer Methods and Programs in Biomedicine* 109 (3): 293–304. https://doi.org/10.1016/j.cmpb.2012.09.002.

Chen, Chia-Cheng, and Huiman X. Barnhart. 2008. "Comparison of ICC and CCC for Assessing Agreement for Data Without and with Replications." *Computational Statistics & Data Analysis* 53 (2): 554–64. https://doi.org/10.1016/j.csda.2008.09.026.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46. https://doi.org/10.1177/001316446002000104.

———. 1968. "Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit." *Psychological Bulletin* 70 (4): 213–20. https://doi.org/10.1037/h0026256.

"Correction: A Note on the Concordance Correlation Coefficient." 2000. *Biometrics* 56 (1): 324–25.

Donner, Allan, and Guangyong Zou. 2002. "Testing the Equality of Dependent Intraclass Correlation Coefficients." *Journal of the Royal Statistical Society: Series D (The Statistician)* 51 (3): 367–79. https://doi.org/10.1111/1467-9884.00324.

Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7 (1). https://doi.org/10.1214/aos/1176344552.

Efron, Bradley. 1987. "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 82 (397): 171–85. https://doi.org/10.1080/01621459.1987.10478410.

Feng, Dai. 2020a. "agRee: Various Methods for Measuring Agreement." https://CRAN.R-project.org/package=agRee.

———. 2020b. "agRee: Various Methods for Measuring Agreement." https://CRAN.R-project.org/package=agRee.

Feng, Dai, Richard Baumgartner, and Vladimir Svetnik. 2013. "A Short Note on Jackknifing the Concordance Correlation Coefficient." *Statistics in Medicine* 33 (3): 514–16. https://doi.org/10.1002/sim.5931.

———. 2015. "A Bayesian Estimate of the Concordance Correlation Coefficient with Skewed Data." *Pharmaceutical Statistics* 14 (4): 350–58. https://doi.org/10.1002/pst.1692.

———. 2018. "A Bayesian Framework for Estimating the Concordance Correlation Coefficient Using Skew-Elliptical Distributions." *The International Journal of Biostatistics* 14 (1). https://doi.org/10.1515/ijb-2017-0050.

Fisher, R. A. 1925. *Statistical Methods for Research Workers, 11th Ed. Rev.* Edited by Oliver and Boyd. Edinburgh: Oliver; Boyd.

Gatti, Stefano, Emanuele Rezoagli, Fabiana Madotto, Giuseppe Foti, and Giacomo Bellani. 2024. "A Non-Invasive Continuous and Real-Time Volumetric Monitoring in Spontaneous Breathing Subjects Based on Bioimpedance—ExSpiron®Xi: A Validation Study in Healthy Volunteers." *Journal of Clinical Monitoring and Computing* 38 (2): 539–51. https://doi.org/10.1007/s10877-023-01107-0.

King, Tonya S., and Vernon M. Chinchilli. 2001a. "ROBUST ESTIMATORS OF THE CONCORDANCE CORRELATION COEFFICIENT." *Journal of Biopharmaceutical Statistics* 11 (3): 83–105. https://doi.org/10.1081/bip-100107651.

———. 2001b. "A Generalized Concordance Correlation Coefficient for Continuous and Categorical Data." *Statistics in Medicine* 20 (14): 2131–47. https://doi.org/10.1002/sim.845.

King, Tonya S., Vernon M. Chinchilli, and Josep L. Carrasco. 2007. "A Repeated Measures Concordance Correlation Coefficient." *Statistics in Medicine* 26 (16): 3095–3113. https://doi.org/10.1002/sim.2778.

Konishi, Sadanori, and A. K. Gupta. 1987. "Inferences about Interclass and Intraclass Correlations from Familial Data." In, 225–33. Springer Netherlands. https://doi.org/10.1007/978-94-009-4794-8_12.

Lin, Lawrence I-Kuei. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45 (1): 255. https://doi.org/10.2307/2532051.

Lin, Lawrence, A. S. Hedayat, and Wenting Wu. 2007. "A Unified Approach for Assessing Agreement for Continuous and Categorical Data." *Journal of Biopharmaceutical Statistics* 17 (4): 629–52. https://doi.org/10.1080/10543400701376498.

Lüdecke, Daniel, Mattan Ben-Shachar, Indrajeet Patil, Philip Waggoner, and Dominique Makowski. 2021. "Performance: An r Package for Assessment, Comparison and Testing of Statistical Models." *Journal of Open Source Software* 6 (60): 3139. https://doi.org/10.21105/joss.03139.

Mitra, Vikramjit, Hsiang-Yun Sherry Chien, Vasudha Kowtha, Joseph Yitan Cheng, and Erdrin Azemi. 2022. "Speech Emotion: Investigating Model Representations, Multi-Task Learning and Knowledge Distillation." *Interspeech 2022*, September, 4715–19. https://doi.org/10.21437/interspeech.2022-957.

Moreno, J, NB Vaz, JC Soler, JL Carrasco, S Podlipnick, S Puig, and Oleaga L. 2018. "High Signal Intensity in the Dentate Nucleus on Unenhanced T1-Weighted MR Images in Melanoma Patients Receiving Macrocyclic Gadolinium-Based Contrast." *Journal of Radiology and Diagnostic Methods* 1.

Pek, Jolynn, Octavia Wong, and Augustine C. M. Wong. 2018. "How to Address Non-Normality: A Taxonomy of Approaches, Reviewed, and Illustrated." *Frontiers in Psychology* 9 (November). https://doi.org/10.3389/fpsyg.2018.02104.

Preston, D. C., and B. E. Shapiro, eds. 2009. "Neuroimaging in Neurology: An Interactive CD." *American Journal of Neuroradiology* 30 (8): E125–25. https://doi.org/10.3174/ajnr.a1685.

Searle, Shayle R., George Casella, and Charles E. McCulloch. 1992. "Variance Components." *Wiley Series in Probability and Statistics*, March. https://doi.org/10.1002/9780470316856.

Shirato, Midori, Keisuke Nakamura, Taichi Tenkumo, Yoshimi Niwano, Taro Kanno, Keiichi Sasaki, Peter Lingström, and Ulf Örtengren. 2022. "Inhibition of Tooth Demineralization Caused by Streptococcus Mutans Biofilm via Antimicrobial Treatment Using Hydrogen Peroxide Photolysis." *Clinical Oral Investigations* 27 (2): 739–50. https://doi.org/10.1007/s00784-022-04821-2.

Stan Development Team. 2024. "{RStan}: The {r} Interface to {Stan}." https://mc-stan.org/.

Tsai, Miao-Yu, and Chao-Chun Lin. 2018. "Concordance Correlation Coefficients Estimated by Variance Components for Longitudinal Normal and Poisson Data." *Computational Statistics & Data Analysis* 121 (May): 57–70. https://doi.org/10.1016/j.csda.2017.12.003.

Vanbelle, Sophie, and Emmanuel Lesaffre. 2017. "Modeling Agreement on Bounded Scales." *Statistical Methods in Medical Research* 27 (11): 3460–77. https://doi.org/10.1177/0962280217705709.

Wheatland, Rand. 2022. "Pervasive Abuse of Log-Transformed Data in Viral Load Analyses." *Clinical Microbiology and Infection* 28 (12): 1669–70. https://doi.org/10.1016/j.cmi.2022.08.020.

A Additional Information on the simulation results

In this appendix, I have included some additional information on how the specific coefficients were derived for the simulations, the inclusion of the log-normal effects and its impact on the skewness, the rate of failure and the average bias of the point estimate of the evaluated methods, and of the mean skewness of the response variable in the simulations included in this thesis.

A.1 Concordance Levels

Given the standard longitudinal CCC formula derived through variance components $CCC = \frac{\sigma^2_\alpha + \sigma^2_{\alpha\gamma}}{\sigma^2_\alpha + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma} + \sigma^2_\epsilon}$, the numerator is composed by the subjects and the subject-time random effects' variance while the denominator also includes the variance that can be attributed to the fixed effects, the subject-method variance and the random error variance. For the five levels of CCC evaluated the variance of the subjects' effect was kept fixed at 0.5 and the total variance was also kept fixed at 10, and the different fractions were obtained by modifying the variance given to the subject-time effect and the random error.

A.2 Lognormal effect

A lognormal random variable X can be defined through a location μ and positive scale parameter σ in the following way $X = \exp(\mu + \sigma Z)$, with Z being a standard normal random variable. Its relevant moments for this thesis are its mean, which is $\mu_X = \exp(\mu + \sigma^2/2)$, variance $\sigma^2_X = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$, and skewness $\gamma_{1_X} = [\exp(\sigma^2) + 2]\sqrt{\exp(\sigma^2) - 1}$. Given that for the simulations σ^2_X was fixed at 0.5, the lognormal scale parameter can be derived with the following formula:

$$\sigma^2 = \log\left(\frac{1 + \sqrt{1 + 4\frac{\sigma^2_X}{\exp(2\mu)}}}{2}\right)$$

The location parameter μ then allows to control the degree of skewness for a same level of variance (as by reducing the location parameter the scale parameter increases). The two values used in the simulations were $-\log(1)$ and $-\log(10)$ in order to produce different levels of skewness (2.03 and 24.6).

While in both scenarios the subject effect presents a strong skewness, the skewness of the response variable (which is a sum of the fixed and random effects) is moderated by the other normally distributed variance components, given that the skewness be formulated as a ratio between the mean and the variance $\gamma_1 = \frac{\mu_3}{(\sigma^2)^{3/2}}$, and as the normal random effects are centered at zero they only contribute increasing the total variance of the response variable, and thus reducing the final skewness. The resulting skewness is further discussed in the next section.

A.3 Extra tables

As commented in the first set of simulations, both bayesian and U-statistic estimators did not present any significant rate of failure, while the asymptotic and the bootstrap approaches (the bootstrap resamples are generated with the asymptotic method) both show a meaningful failure rate close to the coefficient bounds. The bootstrap rate of failure can be mitigated (see LN-N Model tables) by suppressing the resamples where the asymptotic method failed, but nonetheless they still show the largest failure rates across the board (while now on par with the asymptotic methods for low values of the coefficient, they still show significantly larger failure rates at 0.9) even when taking this into account.

Table A.1: Failure rate under the N-N Model (continued below)

	AN_NA	AF_NA	AF2_NA	AKG_NA	BN_NA	U_NA	BPB_NA	BPE_NA
CCC_01	0.087	0.075	0.07	0.102	0	0	0.133	0.13
CCC_03	0.035	0.031	0.031	0.02	0	0	0.026	0.029
CCC_05	0.019	0.017	0.026	0.018	0	0	0.031	0.025
CCC_07	0.015	0.019	0.008	0.013	0	0	0.043	0.04
CCC_09	0.034	0.045	0.044	0.049	0	0	0.678	0.67

	NBPB_NA	NBPE_NA
CCC_01	0	0
CCC_03	0	0
CCC_05	0.001	0.001
CCC_07	0.001	0.001
CCC_09	0.002	0

Table A.3: Failure rate under the lightly skewed LN-N Model (continued below)

	AN_NA	AF_NA	AF2_NA	AKG_NA	BN_NA	U_NA	BPB_NA	BPE_NA
CCC_01	0.04	0.049	0.045	0.05	0.082	0	0.057	0.054
CCC_03	0.021	0.016	0.015	0.014	0	0	0.034	0.026
CCC_05	0.01	0.005	0.007	0.013	0	0	0.018	0.019
CCC_07	0.002	0.003	0.006	0.004	0	0	0.016	0.008
CCC_09	0.009	0.014	0.011	0.012	0	0	0.236	0.225

	NBPB_NA	NBPE_NA
CCC_01	0.054	0.054
CCC_03	0.026	0.026
CCC_05	0.019	0.019
CCC_07	0.008	0.015
CCC_09	0.225	0.228

Table A.5: Failure rate under the lightly highly LN-N Model (continued below)

	AN_NA	AF_NA	AF2_NA	AKG_NA	BN_NA	U_NA	BPB_NA	BPE_NA
CCC_01	0.036	0.024	0.025	0.015	0	0	0.057	0.059
CCC_03	0.004	0.001	0.002	0.001	0	0	0.035	0.04
CCC_05	0.002	0.002	0	0.002	0	0	0.004	0.007
CCC_07	0	0	0	0	0	0	0.003	0.002
CCC_09	0.001	0.003	0.001	0.005	0	0	0.086	0.095

	NBPB_NA	NBPE_NA
CCC_01	0.059	0.059
CCC_03	0.04	0.04
CCC_05	0.007	0.007
CCC_07	0.002	0.002
CCC_09	0.095	0.095

When assessing the bias of the point estimates, the use of the bayesian MAP as one did not produce a larger bias than the other discussed approaches, which were generally on par. It was also observed that it the average bias grew (diverged from zero) jointly with the mean skewness.

Table A.7: Average bias under the N-N Model (continued below)

	AN_b	AF_b	AF2_b	AKG_b	BN_b	U_b	BPB_b
CCC_01	0	-0.001	-0.003	-0.001	0.001	0.003	0
CCC_03	0	0.001	0.001	0.002	0.003	0.001	0.002
CCC_05	0.002	0.002	0.002	-0.001	0.008	0.002	-0.001
CCC_07	0.002	0	0.001	0.001	0.011	0.004	0.002
CCC_09	0.001	0	0	0.001	0.018	0.001	0.002

	BPE_b	NBPB_b	NBPE_b
CCC_01	-0.002	0	0
CCC_03	-0.001	-0.001	-0.001
CCC_05	0.001	0	0
CCC_07	0.001	0.001	0.001
CCC_09	0.001	0.001	0

Table A.9: Average bias under the lightly skewed LN-N Model (continued below)

	AN_b	AF_b	AF2_b	AKG_b	BN_b	U_b	BPB_b
CCC_01	-0.005	-0.009	-0.003	-0.004	0.005	0	-0.007
CCC_03	-0.001	0	0.001	-0.002	0.007	0.009	0.004
CCC_05	0.004	0.002	0.005	0.002	0.003	0.015	0.007
CCC_07	0.003	0.004	0.004	0.005	-0.002	0.01	0.001
CCC_09	0.004	0.003	0.002	0.002	-0.002	0.007	0.003

	BPE_b	NBPB_b	NBPE_b
CCC_01	-0.008	-0.008	-0.008
CCC_03	0.001	0.001	0.001
CCC_05	0.006	0.006	0.006
CCC_07	0.002	0.002	0.005
CCC_09	0.004	0.004	0.006

Table A.11: Average bias under the lightly highly LN-N Model (continued below)

	AN_b	AF_b	AF2_b	AKG_b	BN_b	U_b	BPB_b
CCC_01	-0.508	-0.511	-0.504	-0.522	-0.524	-0.51	-0.509

	AN_b	AF_b	AF2_b	AKG_b	BN_b	U_b	BPB_b
CCC_03	-0.397	-0.405	-0.403	-0.406	-0.402	-0.399	-0.393
CCC_05	-0.284	-0.284	-0.279	-0.287	-0.286	-0.275	-0.281
CCC_07	-0.173	-0.172	-0.169	-0.171	-0.172	-0.167	-0.176
CCC_09	-0.056	-0.056	-0.057	-0.057	-0.06	-0.054	-0.058

	BPE_b	NBPB_b	NBPE_b
CCC_01	-0.49	-0.49	-0.49
CCC_03	-0.405	-0.405	-0.405
CCC_05	-0.277	-0.277	-0.277
CCC_07	-0.172	-0.172	-0.172
CCC_09	-0.059	-0.059	-0.059

Given how the simulated datasets were constructed, where the subjects' random effect variance, which was the lognormal component of the model, was set at 0.5 (the total variance was 10), it can be seen that under light variance conditions for this lognormal effect it does not produce relevant skewness in the response variable, and it requires a extremely large skewness to meaningfully affect the estimation under this conditions. It follows that if the weight of the lognormal effect was larger of the total variance, its skewness would be less moderated.

Table A.13: Mean skewness under the lightly skewed LN-N Model

	AN_s	AF_s	AF2_s	AKG_s	BN_s
CCC_01	0	0	0	0	0.02
CCC_03	0	0	0	0	0.013
CCC_05	0	0	0	0	0.021
CCC_07	0	0	0	0	0.025
CCC_09	0	0	0	0	0.009

Table A.14: Mean skewness under the lightly highly LN-N Model (continued below)

	AN_s	AF_s	AF2_s	AKG_s	BN_s	U_s	BPB_s	BPE_s
CCC_01	1.472	1.508	1.463	1.547	1.636	1.526	1.467	1.362
CCC_03	1.523	1.572	1.547	1.556	1.55	1.554	1.473	1.571
CCC_05	1.535	1.534	1.451	1.533	1.544	1.475	1.463	1.474
CCC_07	1.556	1.53	1.541	1.487	1.547	1.501	1.607	1.56
CCC_09	1.508	1.509	1.532	1.507	1.624	1.474	1.628	1.634

	NBPB_s	NBPE_s
CCC_01	1.362	1.362
CCC_03	1.571	1.571
CCC_05	1.474	1.474
CCC_07	1.56	1.56
CCC_09	1.634	1.634