Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 145–167
http://www.ssca.org.in/journal

A Modified Measurement Error Model for Replicated Method Comparison Data with Skewness and Heavy Tails

Jeevana Duwarahan[1,2] and Lakshika S. Nawarathna[3]
[1]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka*
[2]*Department of Mathematics and Statistics, University of Jaffna, Jaffna, Sri Lanka*
[3]*Department of Statistics and Computer Science*
Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka

Received: 24 January 2023; Revised: 21 July 2023; Accepted: 06 December 2023

Abstract

Measurement error models (MEMs) provide a flexible framework to model the method comparison data by incorporating measurement errors. However, these models often rely on normality assumptions, which are frequently violated in practice due to skewness and heavy tails. Furthermore, repeated data with measurement errors (MEs) are often observed in medical research, epidemiological studies, economics, and the environment. Thus, this research aims to assess the extent of similarity and agreement between the two methods using the replicated measurement error model (RMEM) under asymmetric and heavy-tailed distributions with a matching degree for true covariate and errors. The expectation-maximization (EM) approach is applied to fit the model. A simulation study is used to test the proposed methodology, demonstrated by evaluating subcutaneous fat data. The Total Deviation Index (TDI) and Concordance Correlation Coefficient (CCC) were used to further assess the agreement between the methods. Our suggested model works well for analyzing replicated method comparison data with measurement errors, skewness, and heavy tails.

Key words: Agreement; Heavy-tailed distributions; Replicated measurement error model; EM algorithm; Concordance correlation; Total deviation index.

1. Introduction

Method comparison study refers to comparing two or more methods that analyze the outcome for understanding the agreement between the methods. Generally, a comparison is made between the already established methods and the new methods to see whether there is enough agreement between them. If the method comparison study of the continuous variables is agreeable to each other or similar, it reflects that both methods can be interchangeably used. With the vast development in the field of health and sciences, method comparisons play a vital role in determining the better of the existing practices and new innovative methods that are put into use. The methods include an assay, equipment, medical device,

Corresponding Author: Lakshika S. Nawarathna
Email: lakshikas@sci.pdn.ac.lk

observation, measurement techniques, and variables of interest such as blood pressure, pulse rate, level of cholesterol, the level of concentration of the chemical used, *etc.* Currently, we have new techniques evolving in the health sector as a result of advancement, and these new techniques might be much more effective, less invasive, economical, faster, and easy to handle. However, the medical practitioner needs analysis of these more recent methods that are to be compared with the already existing methods or standards to understand the outcome.

Numerous method comparison studies are being conducted in the field of health and science to evaluate the techniques used. Comparing the measurements of continuous variables helps us determine the better of the prevailing methods or if they can be used interchangeably. In method comparison studies, every subject has at least one measurement from each method. Our focus in this research is where measurements are replicated. The first step in the methodology is to model the method comparison data where the mixed-effects model is commonly used. It is to be noted that the model assumes independent normal distribution for both random effects and errors, and the measurement variability is constant over the whole measurement range (Bland and Altman, 1999, 2007; Carrasco and Jover, 2003; Carstensen *et al.*, 2008; Carrasco *et al.*, 2009; Hedayat *et al.*, 2009; Choudhary, 2008). Secondly, the evaluation of agreement between the methods is conducted using inference on one or more measures of agreements that quantify how much they agree well. When the difference in measurements is small, it reflects a good agreement between the two methods. There are numerous agreement measures, including the CCC (Lin, 1989; Barnhart *et al.*, 2007; Nawarathna and Choudhary, 2013, 2015) and the TDI (Lin, 1989; Nawarathna and Choudhary, 2013, 2015; Choudhary, 2009; Choudhary and Yin, 2010) have attracted the greatest attention in the statistical literature.

In many real-world situations, accurately measuring the true value of a variable is challenging. Instead, we can only observe it with some degree of error. This discrepancy between the observed and true values is known as "Measurement Error (ME)". Imagine trying to hit a target with a bow and arrow. The true bullseye represents the actual value we aim to measure, while the observed values are scattered around it due to measurement error. These errors can arise from various factors, such as different measurement methods, instruments, human error, or external influences. Ignoring these errors can lead to biased estimates and increased variability in statistical inferences. Therefore, it is essential to consider measurement errors to ensure accurate and reliable statistical analysis.

MEMs, which have been discussed in Nawarathna and Choudhary (2015), Dunn and Roberts (1999), Alanen (2010), typically assume normality for both the true covariate and error terms. However, in practice, the method comparison data often reflects skewness and heavy tails, indicating departures from normality. This is exemplified by the subcutaneous fat data discussed in Carstensen *et al.* (2020), demonstrating these characteristics. While data transformations can be used to achieve normality, limiting transformations to (natural) logarithmic transformations in method comparison studies is generally advised, as Bland and Altman (1999) recommend. However, the log transformation may not always be successful. In such cases, alternative approaches should be considered. Nonparametric methods, as suggested by King and Chinchilli (2001), King *et al.* (2007), and Choudhary (2010), do not rely on distributional assumptions. Generalized Estimating Equations (GEE), discussed by Barnhart *et al.* (2002, 2005), and Lin *et al.* (2007), offer a semiparametric approach

by directly modeling the moments of the data without assuming a specific distribution. Additionally, parametric models can be utilized based on distributions other than the normal distribution, as explored by Sengupta *et al.* (2015). These alternative approaches provide flexibility in modeling method comparison data, accounting for its specific characteristics beyond the assumptions of normality.

The parametric mixed-effects model approach developed by Sengupta *et al.* (2015) offers a methodology for analyzing method comparison data with skewness and heavy tails. In the context of MEM, Duwarahan and Nawarathna (2022) modified the STcT-MEM (Tomaya and de Castro, 2018) specifically for unreplicated method comparison data with known error variances. However, no MEM model is designed for replicated method comparison data. Inspired by this gap, we aim to modify a model within the MEM framework to analyze replicated method comparison data with skewed and heavy-tailed features. In our approach, building upon the work of Cao *et al.* (2017), we consider MEMs for replicated data under scale mixtures of skew-normal (SMSN) distributions for the true covariate and scale mixtures of normal (SMN) distributions for the error terms with the matching degree. Specifically, we use the skew-t (ST) distribution for the true covariate and the t distribution for the error term. We also consider the skew-normal (SN) and normal (N) distributions for comparative purposes. The primary objective of this paper is to modify the above model to analyze method comparison data, assess the agreement between the two methods, and determine if they can be used interchangeably.

Additionally, our proposed methodology provides a unified framework that can handle various types of data, including normally distributed, skewed, and heavy-tailed data. It encompasses the N-RMEM (normal-distributed replicated measurement error model) and SN-RMEM (skew-normal-distributed replicated measurement error model) as special cases. Specifically, when the degrees of freedom reach infinity, the SN-RMEM turns into a special case of the ST-RMEM (skew-t-distributed replicated measurement error model). Similarly, when the degrees of freedom tends to infinity and the skewness parameter is zero, the N-RMEM becomes a special case of the ST-RMEM. This flexibility allows for comprehensive analysis and comparison of different types of method comparison data.

The remainder of the paper is organized as follows. Section 2 introduces the ST-RMEM for method comparison data. Section 3 discusses the proposed methodology for evaluating similarity and agreement under ST-RMEM. Section 4 investigates the proposed model's performance using simulated studies. Section 5 illustrates our method using subcutaneous fat data, and the concluding section summarizes the findings and conclusions. The statistical program R (R Core Team, 2021) was used to perform all of the computations given in this research.

2. Framework for method comparison data

This section describes a framework for analyzing research that compares two methods when taking several measurements on each subject. Let $Y_{ijk}, k = 1, 2, \ldots, n_j, j = 1, 2, i = 1, 2, \ldots, m$ denote the k^{th} replicate measurement of the j^{th} method on the i^{th} subject. Here m is the number of subjects in the study, and n_j is the number of measurements on method j. It is to be noted that $n_j \geq 2$. Here Method 1 is the reference method, while Method 2 is the test method. Let $n = n_1 + n_2$ represent the total number of measurements taken on the

subject and $N = nm$ represent the total number of measurements in the dataset.

If multiple measurements are found on each subject, it is referred to as 'repeated measurements data' and categorized as unlinked, linked, and longitudinal data. These categories are essential as it influences the way the data are modeled. In this research, we focus on unlinked data. Unlinked data refers to the measurements obtained from the two methods separately, and multiple measurements on a subject taken by a method are independent replications of the same underlying measurement. In this case, it is not mandatory for the methods to have the same number of replications on a subject. As always, measurements from various subjects are presumed to be independent.

Let $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, and $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ indicate the p dimensional N, SN, t, and ST distributions, respectively. Here, $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location vector, $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite scale matrix, $\boldsymbol{\lambda} \in \mathbb{R}^p$ is a vector of skewness parameters, and $\nu(> 0)$ is degrees of freedom. Let $G(\alpha, \beta)$ represent the gamma distribution with parameters $\alpha(> 0)$ and $\beta(> 0)$, and $HN(0, \sigma^2)$ represent the half-normal $(0, \sigma^2)$ distribution. Let $\mathbf{I}_p$ denote a $p \times p$ identity matrix. The symbol $\boldsymbol{\Sigma}^{1/2}$ represents a square root of the symmetric and positive definite matrix $\boldsymbol{\Sigma}$. This implies that $\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})^T = \boldsymbol{\Sigma}$, where the symbol T denotes transposition. The inverse of $\boldsymbol{\Sigma}$ is denoted as $\boldsymbol{\Sigma}^{-1}$.

2.1. Definition of ST-RMEM

The classical replicated measurement error model is

$$\begin{aligned} X_{ik} &= b_i + \delta_{ik}; \quad k = 1, 2, \ldots, p \text{ and} \\ Y_{il} &= y_i + \epsilon_{ik}; \quad l = 1, 2, \ldots, q \\ y_i &= \alpha + \beta b_i + e_i; \quad i = 1, 2, \ldots, m \end{aligned} \tag{1}$$

where b_i, y_i be the unobserved true covariate and response, and they are observed p and q times, respectively; α is the fixed bias; slope β is its proportional bias; δ_{ik}, ϵ_{ik} are measurement errors of X_{ik} and Y_{il}, respectively; e_i is the equation error, which indicates that the true variables b_i and y_i are not completely connected if other factors other than b_i are also involved in the variation in y_i, and δ_{ik}, ϵ_{ik}, e_i are uncorrelated with each other. Moreover, e_i is known as 'method-subject interaction' in a mixed-effects model. When a measurement error model is used, it may be noted that they are frequently incorporated in the testing method but not in the standard method. However, when a mixed-effects model is used, they are always included in both methods. A slope β with a non-unit value suggests a difference in the proportionate biases (or scales) of the methods.

Consider a $(p+q)$ dimensional random vector $\boldsymbol{Z}_i = (\boldsymbol{X}_i^T, \boldsymbol{Y}_i^T)^T$, where $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T$ is a p dimensional random vector and $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iq})^T$ is a q dimensional random vector. From (1), the model can be written as

$$\boldsymbol{Z}_i = \boldsymbol{A} + \boldsymbol{B} b_i + \boldsymbol{\psi}_i \tag{2}$$

where $\mathrm{A} = \begin{bmatrix} \mathbf{0}_p \\ \alpha \mathbf{1}_q \end{bmatrix}$, $\mathrm{B} = \begin{bmatrix} \mathbf{1}_p \\ \beta \mathbf{1}_q \end{bmatrix}$, $\boldsymbol{\psi}_i = \begin{bmatrix} \boldsymbol{\delta}_i \\ e_i \mathbf{1}_q + \boldsymbol{\epsilon}_i \end{bmatrix}$ with $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{ip})^T$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iq})^T$. It is standard to assume that b_i, e_i, δ_{ik} and ϵ_{ik} are independent and

$$b_i \sim N_1(\mu_b, \phi_b), \quad e_i \sim N_1(0, \phi_e), \quad \delta_{ik} \sim N_1(0, \phi_\delta), \quad \text{and} \quad \epsilon_{ik} \sim N_1(0, \phi_\epsilon). \tag{3}$$

Normality assumption is sometimes unfeasible due to the skewness, heavy-tailed ness, and outliers. To overcome this problem, Cao *et al.* (2017) developed the ST-RMEM by considering ST for true covariate and t distribution for error terms with the same degrees of freedom. It follows

$$b_i \sim ST_1(\mu_b, \phi_b, \lambda_b, \nu), \quad e_i \sim t_1(0, \phi_e, \nu), \quad \delta_{ik} \sim t_1(0, \phi_\delta, \nu), \quad \text{and} \quad \epsilon_{ik} \sim t_1(0, \phi_\epsilon, \nu). \tag{4}$$

It can be hierarchically represented as

$$\begin{aligned} \boldsymbol{Z}_i \mid b_i, U_i = u_i &\sim N_n(\boldsymbol{A} + \boldsymbol{B}b_i, \boldsymbol{\Sigma}_1/U_i), \\ b_i \mid U_i = u_i, V_i = v_i &\sim N_1(\mu_b + \gamma_b v_i, \tau_b/U_i), \\ V_i \mid U_i = u_i &\sim HN(0, 1/U_i), \\ U_i &\sim G(\frac{\nu}{2}, \frac{\nu}{2}). \end{aligned} \tag{5}$$

where $n = p + q$, $\gamma_b = \phi_b^{1/2}\delta_b$, $\delta_b = \frac{\lambda_b}{\sqrt{1+\lambda_b^2}}$, $\tau_b = \phi_b(1-\delta_b^2)$, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} \phi_\delta \mathbf{I}_p & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \phi_e \mathbf{1}_q \mathbf{1}_q^T + \phi_\varepsilon \mathbf{I}_q \end{bmatrix}$.

The mean vector and variance matrix of $\boldsymbol{Z}_i$ are as follows.

$$\begin{aligned} E(\boldsymbol{Z}_i) &= \boldsymbol{A} + \boldsymbol{B}E(b_i), \quad \nu > 1 \quad \text{and} \\ Var(\boldsymbol{Z}_i) &= \frac{\nu}{\nu-2}\phi_b \boldsymbol{B}\boldsymbol{B}^T - \zeta^2 \boldsymbol{B}\gamma_b\gamma_b^T\boldsymbol{B}^T + \frac{\nu}{\nu-2}\boldsymbol{\Sigma}_1, \quad \nu > 2 \end{aligned} \tag{6}$$

where $E(b_i) = \mu_b + \zeta\gamma_b$, with $\zeta = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$ and $\Gamma(\cdot)$ denotes the gamma function, and its conditional distribution is expressed as

$$\boldsymbol{Z}_i \mid U_i \sim SN_n(\boldsymbol{A} + \boldsymbol{B}\mu_b, \boldsymbol{\Sigma}/U_i, \boldsymbol{\lambda}), \tag{7}$$

where $\boldsymbol{\Sigma} = \phi_b \boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{\Sigma}_1$ and $\boldsymbol{\lambda} = \frac{\lambda_b \phi_b \boldsymbol{\Sigma}^{-1/2}\boldsymbol{B}}{\sqrt{\phi_b + \lambda_b^2 \Lambda_b}}$ with $\Lambda_b = \frac{\phi_b}{c}$, $c = 1 + \phi_b \boldsymbol{B}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{B}$.
Cao *et al.* (2017) used an EM algorithm to estimate the parameters due to the complexity of the likelihood function.

2.2. ST-RMEM for method comparison data

Let $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \ldots, Y_{ijn_j})^T$ denote the n_j measurement vector from method $j(= 1, 2)$. The vector $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}^T, \boldsymbol{Y}_{i2}^T)^T$ denote all measurements on subject i. Let $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \tilde{Y}_2)^T$ represent paired observations from the two methods on a randomly chosen subject from the population. The basic ST-RMEM can now be used flexibly to model replicated method comparison data. This model implies that Method 1 is a well-known method used as a reference method in the comparison. It is of the form

$$\begin{aligned} &Y_{i1k} = b_i + \delta_{i1k}; \\ &Y_{i2k} = \alpha + \beta b_i + e_i + \epsilon_{i2k}; \quad i = 1, 2, \ldots, m \quad k = 1, 2, \ldots, n_j. \end{aligned} \tag{8}$$

where α and β are the fixed bias and proportional bias of method 2, respectively, b_i denotes the true unobservable measurement for the i^{th} subject, e_i is the equation error, and δ_{i1k}, ϵ_{i2k} are random errors. Both fixed and proportional biases result in systematic measuring

mistakes. However, Method 1 does not assume a fixed or proportionate bias for identifiability reasons.

This model can be expressed in the matrix notation of (2) by setting $\boldsymbol{Z}_i = \boldsymbol{Y}_i, \boldsymbol{\psi}_i = \begin{bmatrix} \boldsymbol{\delta}_{i1} \\ e_i \mathbf{1}_{n_2} + \boldsymbol{\epsilon}_{i2} \end{bmatrix}$ where $\boldsymbol{\delta}_{i1} = (\delta_{i11}, \ldots, \delta_{i1n_1})^T$, $\boldsymbol{\epsilon}_{i2} = (\epsilon_{i21}, \ldots, \epsilon_{i2n_2})^T$, and $(p, q) = (n_1, n_2)$. It further assumes that

$$b_i \sim ST_1(\mu_b, \phi_b, \lambda_b, \nu), \quad e_i \sim t_1(0, \phi_e, \nu), \quad \delta_{i1k} \sim t_1(0, \phi_\delta, \nu), \quad \text{and} \quad \epsilon_{i2k} \sim t_1(0, \phi_\epsilon, \nu). \tag{9}$$

where b_i, e_i, δ_{i1k}, and ϵ_{i2k} are mutually independent. This model is a modification of the ST-RMEM, previously mentioned in this section. It can be considered when the data shows skewness and heavy-tailedness in method comparison. It can handle two or more measuring methods, replicated and un-replicated measurements, as well as balanced and unbalanced designs. In the un-replicated case ($i.e., n_j = 1$) there is no need to include the equation error term. The unknown parameter vector of the model (9) is denoted by $\boldsymbol{\theta} = (\alpha, \beta, \mu_b, \phi_b, \lambda_b, \phi_e, \phi_\delta, \phi_\varepsilon)^T$, and we use the EM algorithm to obtain the maximum likelihood estimates (MLEs) of these parameters. The SN-RMEM gets to be a special case of the ST-RMEM (9) when $\nu \to \infty$.

$$b_i \sim SN_1(\mu_b, \phi_b, \lambda_b), \quad e_i \sim N_1(0, \phi_e), \quad \delta_{i1k} \sim N_1(0, \phi_\delta), \quad \text{and} \quad \epsilon_{i2k} \sim N_1(0, \phi_\epsilon) \tag{10}$$

When the skewness parameter $\lambda_b = 0$ and the degrees of freedom parameter $\nu \to \infty$, it is a standard N-RMEM.

$$b_i \sim N_1(\mu_b, \phi_b), \quad e_i \sim N_1(0, \phi_e), \quad \delta_{i1k} \sim N_1(0, \phi_\delta), \quad \text{and} \quad \epsilon_{i2k} \sim N_1(0, \phi_\epsilon) \tag{11}$$

3. Assessment of similarity and agreement

3.1. Similarity measures

A method comparison study aims to assess the similarity of measuring methods and their agreement. This evaluation is performed by drawing conclusions based on similarity and agreement measures, which are functions of the model parameters. Evaluation of similarity is a comparison of characteristics, including biases, precisions, and scales of the methods, to find out how the methods differ. In the case of the model (8), the similarity is assessed by analyzing biases with intercept (α) and slope (β). The scales of the methods are the same if the slope is 1. In addition, the true values of the methods are also the same if the intercept is zero. Method precisions can be determined using the ratio, denoted as $\lambda = \frac{\text{error variance of Method 1}}{\text{error variance of Method 2}}$. If $\lambda = 1$, methods 1 and 2 are equally accurate, but if $\lambda < 1$, Method 1 is more accurate than Method 2, and if $\lambda > 1$, Method 2 is more accurate than Method 1. However, this necessitates that these methods be on the same scale. For example, the precisions of two thermometers measured in Fahrenheit and Celsius cannot be compared until one is converted to the same scale. Hence, the test method's scale can be adjusted to equal that of the reference method by dividing $\tilde{Y}_2$ by the slope β. As a result, the precision ratio is $\beta^2\lambda$ and is referred to as the 'squared sensitivity ratio'.

3.2. Agreement measures

The evaluation of similarity is just a comparison of the methods' marginal distributions. Evaluation of agreement is an analysis of the methods' joint distribution, including

their marginal distributions. Further, the closeness of the two methods' measurements is referred to as agreement. When their measurements are identical, the methods agree perfectly. In this ideal case, the bivariate distribution of $\tilde{Y}_1$ and $\tilde{Y}_2$ is concentrated on the 45° line; as a result, the joint distribution becomes degenerate at zero.

In practice, we use measures of an agreement to quantify the extent of the agreement. In spite of the fact that a number of agreement measures are available (Barnhart *et al.*, 2007), two among them, to be specific, the CCC and the TDI, have received the foremost consideration in the statistical literature. These are explained below.

3.2.1. Concordance correlation coefficient

This measure was introduced by Lin (1989), and it is defined as

$$CCC = \frac{2\text{cov}(\tilde{Y}_1, \tilde{Y}_2)}{[E(\tilde{Y}_1) - E(\tilde{Y}_2)]^2 + \text{Var}(\tilde{Y}_1) + \text{Var}(\tilde{Y}_2)} \tag{12}$$

It lies in $[-1, 1]$, and a high CCC score indicates good agreement. A score of 1 indicates perfect positive agreement, whereas -1 denotes excellent negative agreement. More information on this measure's properties and generalizations to various data types, and models can be found in Barnhart *et al.* (2007) and Lin *et al.* (2012).

3.2.2. Total Deviation Index

Lin (2000) introduced this measure, defined as

$$TDI(p) = 100\, p^{\text{th}}\text{percentile of} |\tilde{D} = \tilde{Y}_1 - \tilde{Y}_2| \text{ for a specified } p. \tag{13}$$

In practice, the value of p is assumed to be between 0.80 and 0.95. It is a non-negative measure, with smaller values indicating higher agreement and zero indicating perfect agreement. The confidence bounds of this measure reflect how significant a measurement difference may be in a given large fraction of the population. As a result, if all of the discrepancies in this interval are acceptable from a practical standpoint, the methods are said to be in satisfactory agreement. Lin *et al.* (2002) and Choudhary (2008) provided extensive information on this measure.

In order to evaluate the agreement between methods, we first fit a model to the method comparison data Y_{ijk}, $k = 1, \ldots, n_j$, $j = 1, 2$, $i = 1, \ldots, m$, using the maximum likelihood (ML) approach, as indicated in section (2.2). Let $\hat{\boldsymbol{\theta}}$ be the ML estimator of the parameter vector of $\boldsymbol{\theta}$. According to asymptotic theory, when n is large, the sampling distribution of $\hat{\boldsymbol{\theta}}$ approximately follows a multivariate normal distribution with mean $\boldsymbol{\theta}$ and variance $\boldsymbol{I}^{-1}$ under specific regularity constraints, where $\boldsymbol{I}$ is the observed information matrix Lehmann (1998).

Consider φ as a scalar measure of agreement between two methods. Its ML estimator $\hat{\varphi}$ is produced by substituting $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$. When φ is a differentiable function of $\boldsymbol{\theta}$, the delta method can be used to estimate the sample distribution of $\hat{\varphi}$, expressed as $\hat{\varphi} \sim N(\varphi, \boldsymbol{D}^T \boldsymbol{I}^{-1} \boldsymbol{D})$, where $\boldsymbol{D} = \frac{\partial \varphi}{\partial \boldsymbol{\theta}}$ is the Jacobian matrix evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and they are typically estimated numerically. The $100(1-\alpha)\%$ two-sided confidence bounds for the

agreement measure φ are $\hat{\varphi} \pm z_{1-\alpha}\text{SE}(\hat{\varphi})$, where $z_{1-\alpha}$ is the $(1-\alpha)^{\text{th}}$ percentile of $N_1(0,1)$ and $\text{SE}(\hat{\varphi}) = (\boldsymbol{D}^T\boldsymbol{I}^{-1}\boldsymbol{D})^{\frac{1}{2}}$. In specific, in case small values for φ infer good agreement (*e.g.*, TDI), at that point, require an upper bound. Though in case large values for φ infer good agreement (*e.g.*, CCC), at that point, require a lower bound. After applying a normalizing transformation, the confidence intervals are computed to make stride accuracy for parameters or parameter functions whose range does not span the entire real line. The results are rearranged back to the initial scale. Particularly, TDI is transformed using a log transformation, while CCC is transformed using Fisher's z-transformation. These confidence boundaries are then used to assess if the methods agree sufficiently.

This approach makes sense only if there is no proportionate bias in the test procedure. Thus, the test method needs to be adjusted such that its scale matches to that of the reference method before the agreement can be evaluated (Nawarathna and Choudhary, 2015; Choudhary and Nagaraja, 2017). Therefore, we first transform $\tilde{Y}_2$ as $\tilde{Y}_2^{\,*} = \tilde{Y}_2/\beta$ to make $\tilde{Y}_2$ on the same scale as $\tilde{Y}_1$. The measures of agreement in the transformed case are functions of parameters of the bivariate distribution of $(\tilde{Y}_1, \tilde{Y}_2^{\,*})$, respectively, and the equation of these agreement measures can be determined by inserting the moments from their respective bivariate distributions into their definitions. After this transformation, these measures follow from (12)-(13) that

$$CCC^* = \frac{2\text{cov}(\tilde{Y}_1, \tilde{Y}_2^{\,*})}{[E(\tilde{Y}_1) - E(\tilde{Y}_2^{\,*})]^2 + \text{Var}(\tilde{Y}_1) + \text{Var}(\tilde{Y}_2^{\,*})} \tag{14}$$

$$TDI^* = 100\,p^{\text{th}}\text{percentile of}\,|\tilde{D}^* = \tilde{Y}_1 - \tilde{Y}_2^{\,*}|\ \text{for a specified}\ p. \tag{15}$$

3.3. Agreement evaluation under different models

3.3.1. ST-RMEM

As previously mentioned, $\tilde{Y}_j$ denotes a single measurement using the j^{th} method $(j = 1, 2)$ on a randomly selected subject from the population to derive the expressions for measures of the agreement under the assumed ST-RMEM. Moreover, a companion model for $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \tilde{Y}_2)^T$ is generated from the model (8).

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A} + \boldsymbol{B}\tilde{b} + \tilde{\boldsymbol{\psi}} \tag{16}$$

where $\boldsymbol{A} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$; $\boldsymbol{B} = \begin{bmatrix} 1 \\ \beta \end{bmatrix}$; $\tilde{\boldsymbol{\psi}} = \begin{bmatrix} \delta_1 \\ e + \epsilon_2 \end{bmatrix}$.

Further, $\tilde{b} \sim \text{ST}_1(\mu_b, \phi_b, \lambda_b, \nu)$ and $\tilde{\boldsymbol{\psi}} \sim t_2(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{1}}, \nu)$ with $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{1}} = \begin{bmatrix} \phi_\delta & 0 \\ 0 & \phi_e + \phi_\varepsilon \end{bmatrix}$.

The mean vector and variance matrix of $\tilde{\boldsymbol{Y}}^* = (\tilde{Y}_1, \tilde{Y}_2^* = \tilde{Y}_2/\beta)$ are as follows from (16) that

$$\begin{aligned} E(\tilde{\boldsymbol{Y}}^*) &= \boldsymbol{A}^* + \boldsymbol{B}^* E(b_i), \quad \nu > 1 \,\text{and} \\ Var(\tilde{\boldsymbol{Y}}^*) &= \frac{\nu}{\nu-2}\phi_b \boldsymbol{B}^* \boldsymbol{B}^{*T} - \zeta^2 \boldsymbol{B}^* \gamma_b \gamma_b^T \boldsymbol{B}^{*T} + \frac{\nu}{\nu-2}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{1}}^{\,*}, \quad \nu > 2 \end{aligned} \tag{17}$$

where $\boldsymbol{A}^* = \begin{bmatrix} 0 \\ \alpha/\beta \end{bmatrix}$; $\boldsymbol{B}^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$; $E(b_i) = \mu_b + \zeta\gamma_b$ with $\zeta = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{1}}^{\boldsymbol{*}} = \begin{bmatrix} \phi_\delta & 0 \\ 0 & \frac{1}{\beta^2}(\phi_e + \phi_\varepsilon) \end{bmatrix}$.

Further, we can write using the hierarchical representation (7),

$$\tilde{\boldsymbol{Y}} = \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} | U \sim \mathrm{SN}_2(\boldsymbol{A} + \boldsymbol{B}\mu_b, \tilde{\boldsymbol{\Sigma}}/U, \tilde{\boldsymbol{\lambda}}) \tag{18}$$

where $\boldsymbol{A} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix}; \boldsymbol{B} = \begin{bmatrix} 1 \\ \beta \end{bmatrix}; \tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \phi_b + \phi_\delta & \beta\phi_b \\ \beta\phi_b & \beta^2\phi_b + \phi_e + \phi_\epsilon \end{bmatrix}$ and $\tilde{\boldsymbol{\lambda}} = \frac{\lambda_b \phi_b \tilde{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{B}}{\sqrt{\phi_b + \lambda_b^2 \Lambda_b}}$ are counterparts of $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$. After the transformation, it becomes

$$\tilde{\boldsymbol{Y}}^* = \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2^* \end{pmatrix} | U \sim \mathrm{SN}_2(\boldsymbol{A}^* + \boldsymbol{B}^*\mu_b, \tilde{\boldsymbol{\Sigma}}^*/U, \tilde{\boldsymbol{\lambda}}^*) \tag{19}$$

where $\tilde{\boldsymbol{\Sigma}}^* = \begin{bmatrix} \phi_b + \phi_\delta & \phi_b \\ \phi_b & \phi_b + \frac{1}{\beta^2}(\phi_e + \phi_\epsilon) \end{bmatrix}$ and $\tilde{\boldsymbol{\lambda}}^* = \frac{\lambda_b \phi_b \tilde{\boldsymbol{\Sigma}}^{*-1/2} \boldsymbol{B}^*}{\sqrt{\phi_b + \lambda_b^2 \Lambda_b}}$ with $\Lambda_b = \frac{\phi_b}{c}$,
$c = 1 + \phi_b \boldsymbol{B}^{*T} \tilde{\boldsymbol{\Sigma}}_1^{*-1} \boldsymbol{B}^*, \tilde{\boldsymbol{\Sigma}}_1^* = \begin{bmatrix} \phi_\delta & 0 \\ 0 & \frac{1}{\beta^2}(\phi_e + \phi_\epsilon) \end{bmatrix}$.

Furthermore, we know, if $Y \sim \mathrm{SN}_q(\mu_b, \phi_b, \lambda_b)$ and $\delta_b = \frac{\lambda_b}{(1+\lambda_b^2)^{\frac{1}{2}}}$, $\gamma_b = \phi_b^{\frac{1}{2}} \delta_b$, $\tau_b = \phi_b(1 - \delta_b^2)$. Then

$$\boldsymbol{m}^T Y \sim \mathrm{SN}_1 \left(\boldsymbol{m}^T \mu_b, \boldsymbol{m}^T \phi_b \boldsymbol{m}, \boldsymbol{m}^T \phi_b^{\frac{1}{2}} \delta_b / (\boldsymbol{m}^T \tau_b \boldsymbol{m})^{\frac{1}{2}} \right), \tag{20}$$

where $\boldsymbol{m} \in \mathbb{R}^q$ with at least one non-zero element (Sengupta *et al.*, 2015). It follows from (20) that the difference $\tilde{D} = \tilde{Y}_1 - \tilde{Y}_2$ is

$$\tilde{D}|U \sim \mathrm{SN}_1 \left(\boldsymbol{m}^T(\boldsymbol{A} + \boldsymbol{B}\mu_b), \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}} \boldsymbol{m}/U, \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^{1/2} \tilde{\boldsymbol{\delta}} / (\boldsymbol{m}^T \tilde{\boldsymbol{\Gamma}}\, \boldsymbol{m})^{1/2} \right), \tag{21}$$

where $\boldsymbol{m} = (1, -1)^T$, $\tilde{\boldsymbol{\delta}} = \tilde{\boldsymbol{\lambda}}/(1 + \tilde{\boldsymbol{\lambda}}^T \tilde{\boldsymbol{\lambda}})^{1/2}$ and $\tilde{\boldsymbol{\Gamma}} = \tilde{\boldsymbol{\Sigma}} - \tilde{\boldsymbol{\Sigma}}^{1/2} \tilde{\boldsymbol{\delta}} \tilde{\boldsymbol{\delta}}^T \tilde{\boldsymbol{\Sigma}}^{1/2}$.
When considering transformation, $\tilde{D}^* = \tilde{Y}_1 - \tilde{Y}_2^*$ is

$$\tilde{D}^*|U \sim \mathrm{SN}_1 \left(\boldsymbol{m}^T(\boldsymbol{A}^* + \boldsymbol{B}^*\mu_b), \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^* \boldsymbol{m}/U, \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^{*1/2} \tilde{\boldsymbol{\delta}}^* / (\boldsymbol{m}^T \tilde{\boldsymbol{\Gamma}}^*\, \boldsymbol{m})^{1/2} \right), \tag{22}$$

where $\tilde{\boldsymbol{\delta}}^* = \tilde{\boldsymbol{\lambda}}^*/(1 + \tilde{\boldsymbol{\lambda}}^{*T} \tilde{\boldsymbol{\lambda}}^*)^{1/2}$ and $\tilde{\boldsymbol{\Gamma}}^* = \tilde{\boldsymbol{\Sigma}}^* - \tilde{\boldsymbol{\Sigma}}^{*1/2} \tilde{\boldsymbol{\delta}}^* \tilde{\boldsymbol{\delta}}^{*T} \tilde{\boldsymbol{\Sigma}}^{*1/2}$.

We can now derive the equations for CCC and TDI under ST-RMEM (16) for transformed data. As described in (14), the CCC^* for transformed data can be computed as

$$CCC^* = \frac{2\left[\frac{\nu}{\nu-2}\phi_b - \zeta^2 \gamma_b \gamma_b^T\right]}{\left[\boldsymbol{m}^T(\boldsymbol{A}^* + \boldsymbol{B}^*\mu_b)\right]^2 + \left[\frac{\nu}{\nu-2}\phi_b - \zeta^2 \gamma_b \gamma_b^T + \frac{\nu}{\nu-2}\phi_\delta\right] + \left[\left(\frac{\nu}{\nu-2}\phi_b - \zeta^2 \gamma_b \gamma_b^T\right) + \frac{1}{\beta^2}\left(\frac{\nu}{\nu-2}\phi_\epsilon + \frac{\nu}{\nu-2}\phi_e\right)\right]} \tag{23}$$

Next, as we know from equation (15), the TDI^* is defined as the p^{th} quantile of $\tilde{D}^*$ with a given large probability of $0 < p < 1$, and it can be obtained by solving

$$TDI^* = P(|\tilde{D}^*| \leq t) = \int_0^\infty \{F^*(t) - F^*(-t)\} f(u|\nu)\, du, \quad t > 0 \tag{24}$$

where F^* is the distribution function of $\tilde{D}^*|U$ and $f(u|\nu)$ is the density of U.

3.3.2. SN-RMEM

For the model (10), the mean vector and variance matrix of $\tilde{\boldsymbol{Y}}^*$ are

$$\begin{aligned} E(\tilde{\boldsymbol{Y}}^*) &= \boldsymbol{A}^* + \boldsymbol{B}^* E(b_i), \quad \text{and} \\ Var(\tilde{\boldsymbol{Y}}^*) &= \phi_b \boldsymbol{B}^* \boldsymbol{B}^{*T} \left(1 - \frac{2\delta_b^T \delta_b}{\pi}\right) + \tilde{\boldsymbol{\Sigma}}_1^*, \end{aligned} \tag{25}$$

where $E(b_i) = \mu_b + \sqrt{\frac{2}{\pi}}\gamma_b$.

The hierarchical representation of $\tilde{\boldsymbol{Y}}$ is $\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} \sim \mathrm{SN}_2(\boldsymbol{A} + \boldsymbol{B}\mu_b, \tilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{\lambda}})$, and after the transformation, the marginal distribution of $(\tilde{Y}_1, \tilde{Y}_2^*)$ is $\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2^* \end{pmatrix} \sim \mathrm{SN}_2(\boldsymbol{A}^* + \boldsymbol{B}^*\mu_b, \tilde{\boldsymbol{\Sigma}}^*, \tilde{\boldsymbol{\lambda}}^*)$.

Then, CCC^* can be defined as

$$CCC^* = \frac{2\phi_b \left(1 - \frac{2\delta_b^2}{\pi}\right)}{\left(\frac{\alpha}{\beta}\right)^2 + \left[\phi_b \left(1 - \frac{2\delta_b^2}{\pi}\right) + \phi_\delta\right] + \left[\phi_b \left(1 - \frac{2\delta_b^2}{\pi}\right) + \frac{1}{\beta^2}(\phi_\varepsilon + \phi_e)\right]} \tag{26}$$

Next, $\tilde{D}^* = \tilde{Y}_1 - \tilde{Y}_2^*$ and $\boldsymbol{m} = (1, -1)^T$.

$$\tilde{D}^* \sim \mathrm{SN}_1\left(\alpha/\beta, \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^* \boldsymbol{m}, \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^{*1/2} \tilde{\boldsymbol{\delta}}^* / (\boldsymbol{m}^T \tilde{\boldsymbol{\Gamma}}^* \boldsymbol{m})^{1/2}\right). \tag{27}$$

The TDI* for SN-RMEM is

$$P(|\tilde{D}^*| \le t) = F^*(t) - F^*(-t); \quad t > 0 \tag{28}$$

where F^* is the distribution function of $\tilde{D}^*$.

3.3.3. N-RMEM

The mean vector and variance matrix of $\tilde{\boldsymbol{Y}}^*$ are, according to the standard model (11),

$$\begin{aligned} E(\tilde{\boldsymbol{Y}}^*) &= \boldsymbol{A}^* + \boldsymbol{B}^* \mu_b \quad \text{and} \\ Var(\tilde{\boldsymbol{Y}}^*) &= \phi_b \boldsymbol{B}^* \boldsymbol{B}^{*T} + \tilde{\boldsymbol{\Sigma}}_1^*. \end{aligned} \tag{29}$$

The marginal distribution of $(\tilde{Y}_1, \tilde{Y}_2^*)$ is $\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2^* \end{pmatrix} \sim \mathrm{N}_2(\boldsymbol{A}^* + \boldsymbol{B}^*\mu_b, \tilde{\boldsymbol{\Sigma}}^*)$.

Next, $\tilde{D}^* = \tilde{Y}_1 - \tilde{Y}_2^*$ can be represented as

$$\tilde{D}^* \sim \mathrm{N}_1\left(\alpha/\beta, \boldsymbol{m}^T \tilde{\boldsymbol{\Sigma}}^* \boldsymbol{m}\right). \tag{30}$$

The N-RMEM adaptation of CCC^* can now be defined as

$$CCC^* = \frac{2\phi_b}{\left(\frac{\alpha}{\beta}\right)^2 + [\phi_b + \phi_\delta] + \left[\phi_b + \frac{1}{\beta^2}(\phi_\varepsilon + \phi_e)\right]} \tag{31}$$

The TDI* under N-RMEM can be determined as

$$P(|\tilde{D}^*| \le t) = \Phi\left(\frac{t - E(\tilde{D}^*)}{\mathrm{sd}(\tilde{D}^*)}\right) - \Phi\left(\frac{-t - E(\tilde{D}^*)}{\mathrm{sd}(\tilde{D}^*)}\right) \tag{32}$$

where Φ denotes the cumulative distribution function (CDF) of a standard N distribution.

4. Simulation study

A simulation study is performed to evaluate the performance of the MLEs under the ST-RMEM, SN-RMEM, and N-RMEM models designed for analyzing method comparison data. We generated the data for two different methods, considering sample sizes of $m = 20$, 50, and 100, using models (5) and (8) that incorporated ST distribution for true covariate and t distribution for the error term ($\nu = 5$). The initial values of the parameters are $\mu_b = 1$, $\alpha = 0.02$, $\beta = 0.96$, $\log(\phi_b) = 0.03$, $\log(\phi_\delta) = -5$, $\log(\phi_\varepsilon) = -6$, $\log(\phi_e) = -6$, and we set $\lambda_b = 5$ and 10 for comparison. These findings were inspired by the ML estimate from the real data set studied in Section 5. Furthermore, we assume that the repeated number of observations per method is three. We then compute the MLEs using the EM algorithm on the sample data using the ST, SN, and N distributions with equation error, respectively. For the assessment of the estimations, we compute the sample bias (BIAS), standard deviation (SD), root mean square error (RMSE), and coverage probability (CP) after 1000 repeats. Table 1 summarizes the findings. The R programming language (R Core Team, 2021) was used to do all calculations.

Table 1 shows BIAS, SD, and RMSE values for the ST distribution are lower in all circumstances. As a result, the performance of the ST distribution is better than that of the SN and N distributions, which may be due to their skewed and heavy-tailed characteristics. Additionally, the estimates become more exact when the sample size rises from 20 to 100. When $m = 100$, all coverage probabilities (CPs) are near the nominal value of 95 percent. For smaller and moderate sample sizes, most of the CPs are also around 95 percent, and some are considerably lower. However, the CPs for all cases rise as the sample size increases. As a result, whether the skewness is moderate or heavy, we may state that the ST-RMEM CPs outperform other models.

Table 2 presents the efficiencies of ST-RMEM-based estimators in relation to the SN-RMEM and N-RMEM models calculated by dividing the MSE under the SN-RMEM and N-RMEM models by the MSE under the ST-RMEM. Notice that the relative efficiencies are greater than one in all situations, meaning that ST-RMEM is more accurate than SN-RMEM and N-RMEM. Furthermore, when n rises, the relative efficiencies improve.

We also compute the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values when the data is produced via ST-RMEM. These values are shown in Table 3, and the findings reveal that ST-RMEM performs better than other models since it has lower values. Furthermore, as the sample size rises, the estimates become more exact. Table 4 presents estimated type I error probabilities for the 5% level Likelihood Ratio (LR) test, where the null hypothesis claims that a smaller model (SN-RMEM or N-RMEM) gives a good fit and the alternative hypothesis states that a larger model (ST-RMEM) provides a good fit. For the small sample size, values are close to 5%, showing the minimal difference between the two models. The values are fewer than 5% for moderate and large sample sizes, indicating that ST-RMEM is preferable. In summary, the ST-RMEM performs much better than the N-RMEM and SN-RMEM in the presence of skewed and heavy-tailedness.

5. Application to fat data

The subcutaneous fat thickness (Carstensen *et al.*, 2020) was measured in centimeters at the Steno Diabetes Center to compare the measurements of two experienced observers, 'KL' (Method 1) and 'SL' (Method 2). The study includes 43 persons (subjects), and the measurements (cm) from each method are repeated three times on each subject. The three replicates are interchangeable within the subject and method, and the repeated measurements are unlinked. The design is balanced with 43×3×2=258 observations, and measurements vary from 0.39 to 4.20 cm. Figure 1

depicts a histogram and normal Q-Q plot for subcutaneous fat, revealing that the data is positively skewed and heavy-tailed. The trellis plot of this data, shown in Figure 2, reveals considerable overlap in the measurements given by the methods. At the same time, it is evident that SL values are lower than KL for the majority of persons. A few cases show quite substantial disparities, implying a skewed distribution of differences. The measures show significant within-subject variation, although it is small when compared to between-subject variation. The dataset is homoscedastic, and there are no obvious outliers.

Figure 3 shows scatterplots and Bland-Altman plots of randomly chosen and averaged over replications. The scatter plots reveal a high correlation between the methods, confirming that KL readings are greater than SL readings since most points are above the line of equality, and the Bland-Altman plots indicate that the scales of the methods may differ. Moreover, it should be noted that the data were obtained on persons from a Diabetes Center, and numerous factors, such as a person's food habits and laboratory conditions, might influence a measurement. As a result, these measures are prone to inaccuracy. Thus, the measurement error model gives a better fit for this data.

The modeling of data is the preliminary step in the analysis. Initially, we fit the data using the modified ST-RMEM (9), where b_i follows ST distribution, measurement errors $(\delta_{ik}, \epsilon_{ik})$, and equation errors (e_i) follows multivariate t distribution. In this case, the degree of freedom (ν) is treated as a known parameter, determined by the Schwarz information criteria (Schwarz, 1978). There are a total of eight parameters in this model. The *numDeriv* package (Gilbert and Varadhan, 2019) in R is used to compute the required numerical derivatives. Secondly, we fit the SN-RMEM (10) where b_i follows SN distribution, measurement errors $(\delta_{ik}, \epsilon_{ik})$, and equation errors (e_i) follow multivariate N distribution. This model also has eight unknown parameters. Next, we fit the N-RMEM (11), which has seven unknown parameters. We then compute the MLEs of parameter $\boldsymbol{\theta}$ using the EM algorithm and their standard errors (SEs) under the above models.

Table 5 provides these parameter estimates, SEs, and 95% confidence limits for the above RMEMs. AIC and BIC values based on the RMEM model under the above distributions are shown in Table 6. The model is better when the AIC value is small, and we find that the AIC value is small for ST-RMEM. Furthermore, the LR test is used to determine if the null hypothesis H_0 : SN-RMEM model is preferred to the alternative hypothesis H_1 : ST-RMEM model is preferable. It is important to test the hypothesis H_0 to see if the inclusion of the degrees of freedom (ν) is meaningful. The p-value for this LR test is < 0.0001. Therefore, the parameter (ν) must be taken into account. Thus, the modified ST-RMEM fits significantly better than N-RMEM and SN-RMEM.

The examination of similarity is the second step in the analysis. The proportionate bias estimate (β) is 0.97 (SE = 0.02), and the 95% confidence interval is [0.93, 0.99]. As a result, there is evidence of a minor downward proportionate bias, but it is marginal. Furthermore, the estimated fixed bias (α) is 0.02 (SE = 0.03), with a 95% confidence range of [-0.04, 0.08]. Although this interval includes 0, it also demonstrates a minor fixed bias. Because there is evidence of small bias, the methods have unequal scales. Thus, their precision is measured using a squared sensitivity ratio, and the value is 1.16 (> 1), indicating that Method 2 (SL) is more precise than Method 1(KL).

The next step is an assessment of the agreement. As previously indicated, due to a little bias in SL measurement, we rescaled its measurement $\tilde{Y_2}$ as $\tilde{Y_2}^* = \tilde{Y_2}/\beta$. The estimated transformation is $\tilde{Y_2}^* = \tilde{Y_2}/0.97$. We compute the estimates and 95% one-sided confidence limits for the agreement measures examined in Section 3.2 for the converted data, which are also shown in Table 6. To obtain these estimates, first, perform Fisher's z-transformation for CCC^* and the log transformation for

TDI^*. The CCC^* estimate for ST-RMEM is 0.990, as defined in (23), and its lower bound is 0.984; both are close to one, suggesting good agreement amongst the methods. Next, we make an inference on the agreement measure TDI^* (with $p = 0.90$), which is given by (24). It has an estimate of 0.034 and an upper bound of 0.050. This upper bound indicates that 90% of discrepancies in measurements from the methods lie within -0.05 to 0.05 with 95% confidence. Since the true value range is around 4, this discrepancy may be regarded as acceptable. As a result, we may conclude that the methods are in good agreement. This obviously suggests that the KL and rescaled SL methods agree sufficiently to be deemed interchangeable.

6. Conclusions

This paper develops the methodology for analyzing replicated method comparison data using the MEM framework with the ST distribution for true covariate and the t distribution for errors. We considered the same degree for true covariates and errors. This methodology is sufficient enough to accommodate normally distributed, skewed, heavy-tailed data and both together. The main advantage of this model is that it can assess similarity and agreement between methods, regardless of whether or not the methods use the same nominal unit of measurement. We concentrated here on a comparison of two methods. However, the model may be expanded to include more than two methods. Simulation experiments and the use of subcutaneous fat data confirmed the efficiency and reliability of findings under the ST-RMEM model. Furthermore, we determined that the ST-RMEM model performs best with skewed and heavy-tailed data. Our proposed model would yield appropriate results for method comparison data with measurement error, skewness, and heavy tails, which are frequent in many fields such as economics, health, and the environment.

Data availability statement

The subcutaneous fat dataset is available in Carstensen *et al.* (2020).

References

Alanen, E. (2010). Everything all right in method comparison studies? *Statistical Methods in Medical Research*, **21**, 297–309.

Barnhart, H. X., Haber, M., and Song, J. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics*, **58**, 1020–1027.

Barnhart, H. X., Lokhnygina, Y., Kosinski, A. S., and Haber, M. (2007). Comparison of concordance correlation coefficient and coefficient of individual agreement in assessing agreement. *Journal of Biopharmaceutical Statistics*, **17**, 721–738.

Barnhart, H. X., Song, J., and Haber, M. J. (2005). Assessing intra, inter and total agreement with replicated readings. *Statistics in Medicine*, **24**, 1371–1384.

Bland, J. M. and Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, **8**, 135–160.

Bland, J. M. and Altman, D. G. (2007). Agreement between methods of measurement with multiple observations. *Journal of Biopharmaceutical Statistics*, **17**, 571–582.

Cao, C., Wang, Y., Shi, J. Q., and Lin, J. (2017). Measurement error models for replicated data under asymmetric heavy-tailed distributions. *Computational Economics*, **52**, 531–553.

Carrasco, J. L. and Jover, L. (2003). Estimating the generalized concordance correlation coefficient through variance components. *Biometrics*, **59**, 849–858.

Carrasco, J. L., King, T. S., and Chinchilli, V. M. (2009). The concordance correlation coefficient for repeated measures estimated by variance components. *Journal of Biopharmaceutical Statistics*, **19**, 90–105.

Carstensen, B., Gurrin, L., Ekstrøm, C., and Figurski, M. (2020). Methcomp: Analysis of agreement in method comparison studies. *R package version*, **1**.

Carstensen, B., Simpson, J., and Gurrin, L. C. (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics*, **4**, 1–26.

Choudhary, P. K. (2008). A tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements. *Journal of Statistical Planning and Inference*, **138**, 1102–1115.

Choudhary, P. K. (2009). *Methods and Applications of Statistics in the Life and Health Sciences*, chapter Interrater agreement. John Wiley Sons, New York.

Choudhary, P. K. (2010). A unified approach for nonparametric evaluation of agreement in method comparison studies. *The International Journal of Biostatistics*, **6**, 1–24.

Choudhary, P. K. and Nagaraja, H. N. (2017). *Measuring Agreement: Models, Methods, and Applications*. J. Wiley amp; Sons.

Choudhary, P. K. and Yin, K. (2010). Bayesian and frequentist methodologies for analyzing method comparison studies with multiple methods. *Statistics in Biopharmaceutical Research*, **2**, 122–132.

Dunn, G. and Roberts, C. (1999). Modeling method comparison data. *Statistical Methods in Medical Research*, **8**, 161–179.

Duwarahan, J. and Nawarathna, L. S. (2022). An improved measurement error model for analyzing unreplicated method comparison data under asymmetric heavy-tailed distributions. *Journal of Probability and Statistics*, **2**, 1–13.

Gilbert, P. and Varadhan, R. (2019). numderiv: Accurate numerical derivatives.

Hedayat, A. S., Lou, C., and Sinha, B. K. (2009). A statistical approach to assessment of agreement involving multiple raters. *Communications in Statistics - Theory and Methods*, **38**, 2899–2922.

King, T. S. and Chinchilli, V. M. (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Statistics in Medicine*, **20**, 2131–2147.

King, T. S., Chinchilli, V. M., Wang, K.-L., and Carrasco, J. L. (2007). A class of repeated measures concordance correlation coefficients. *Journal of Biopharmaceutical Statistics*, **17**, 653–672.

Lehmann, E. L. (1998). *Elements of Large-sample Theory*. Springer.

Lin, L., Hedayat, A., and Wu, W. (2012). *Statistical Tools for Measuring Agreement*. Springer.

Lin, L., Hedayat, A. S., Sinha, B., and Yang, M. (2002). Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, **97**, 257–270.

Lin, L., Hedayat, A. S., and Wu, W. (2007). A unified approach for assessing agreement for continuous and categorical data. *Journal of Biopharmaceutical Statistics*, **17**, 629–652.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.

Lin, L. I.-K. (2000). Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Statistics in Medicine*, **19**, 255–270.

Nawarathna, L. S. and Choudhary, P. K. (2013). Measuring agreement in method comparison studies with heteroscedastic measurements. *Statistics in Medicine*, **32**, 5156–5171.

Nawarathna, L. S. and Choudhary, P. K. (2015). A heteroscedastic measurement error model for method comparison data with replicate measurements. *Statistics in Medicine*, **34**, 1242–1258.

R Core Team, R. (2021). *R: A Language and Environment for Statistical Computing*.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Sengupta, D., Choudhary, P. K., and Cassey, P. (2015). Modeling and analysis of method comparison data with skewness and heavy tails. In Choudhary, P., Nagaraja, C. H., and Ng, H. K. T., editors, *Ordered Data Analysis, Modeling and Health Research Methods*, pages 169–187, Cham. Springer International Publishing.

Tomaya, L. C. and de Castro, M. (2018). A heteroscedastic measurement error model based on skew and heavy-tailed distributions with known error variances. *Journal of Statistical Computation and Simulation*, **88**, 2185–2200.

ANNEXURE

Table 1: Simulated bias, sample standard deviation (SD), root mean square error (RMSE), and coverage probability (CP) of ML for the ST-RMEM, SN-RMEM, and N-RMEM

λ	m	Parameter	ST-RMEM				SN-RMEM				N-RMEM			
			BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP
5	20	α	0.000	0.042	0.042	0.968	-0.002	0.054	0.054	0.932	-0.001	0.053	0.053	0.944
		β	0.000	0.022	0.022	0.974	0.002	0.029	0.029	0.890	0.000	0.028	0.028	0.904
		μ_b	0.015	0.080	0.082	0.794	-0.022	0.099	0.101	1.000	0.003	0.214	0.214	0.902
		$\log(\phi_b)$	-0.044	0.152	0.159	0.876	0.097	0.215	0.236	0.960	-0.244	0.593	0.641	0.682
		λ_b	0.003	0.405	0.405	0.802	0.347	1.367	1.410	1.000	-	-	-	-
		$\log(\phi_\delta)$	-0.015	0.215	0.216	0.982	-0.097	0.357	0.370	0.740	-0.095	0.353	0.366	0.746
		$\log(\phi_\varepsilon)$	0.002	0.258	0.258	0.970	-0.059	0.380	0.385	0.748	-0.059	0.380	0.385	0.748
		$\log(\phi_e)$	-0.259	0.617	0.669	0.996	-0.513	0.906	1.041	0.976	-0.438	0.802	0.914	0.980
	50	α	0.000	0.024	0.024	0.984	-0.003	0.038	0.038	0.908	0.000	0.036	0.036	0.924
		β	0.000	0.013	0.013	0.988	0.001	0.021	0.021	0.868	0.000	0.020	0.020	0.882
		μ_b	0.003	0.039	0.039	0.884	-0.034	0.058	0.067	0.996	-0.004	0.133	0.133	0.900
		$\log(\phi_b)$	-0.011	0.067	0.068	0.908	0.109	0.141	0.178	0.942	-0.126	0.445	0.462	0.612
		λ_b	0.001	0.175	0.175	0.884	0.195	0.699	0.726	0.996	-	-	-	-
		$\log(\phi_\delta)$	0.001	0.114	0.114	0.994	-0.036	0.235	0.238	0.772	-0.036	0.232	0.234	0.784
		$\log(\phi_\varepsilon)$	-0.003	0.140	0.140	0.982	-0.037	0.246	0.248	0.758	-0.037	0.246	0.248	0.758
		$\log(\phi_e)$	-0.065	0.256	0.264	0.998	-0.176	0.560	0.587	0.968	-0.149	0.457	0.481	0.980
	100	α	-0.001	0.015	0.015	0.988	-0.005	0.030	0.031	0.904	-0.002	0.030	0.030	0.914
		β	0.001	0.008	0.008	0.984	0.003	0.016	0.017	0.848	0.001	0.016	0.016	0.856
		μ_b	0.002	0.020	0.021	0.934	-0.033	0.038	0.050	0.996	0.000	0.094	0.094	0.934
		$\log(\phi_b)$	-0.005	0.032	0.033	0.938	0.100	0.088	0.133	0.948	-0.067	0.330	0.337	0.614
		λ_b	-0.003	0.099	0.099	0.934	0.097	0.359	0.372	0.998	-	-	-	-
		$\log(\phi_\delta)$	-0.003	0.069	0.069	1.000	-0.020	0.193	0.194	0.724	-0.022	0.190	0.192	0.732
		$\log(\phi_\varepsilon)$	-0.001	0.093	0.093	0.984	-0.023	0.187	0.189	0.708	-0.023	0.187	0.189	0.708
		$\log(\phi_e)$	-0.027	0.136	0.138	0.998	-0.104	0.363	0.378	0.952	-0.089	0.303	0.316	0.964

Table 1: Continued

λ	m	Parameter	ST-RMEM				SN-RMEM				N-RMEM			
			BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP
10	20	α	0.000	0.043	0.043	0.972	-0.002	0.055	0.055	0.934	-0.001	0.055	0.055	0.932
		β	0.000	0.023	0.023	0.972	0.001	0.029	0.029	0.900	0.000	0.029	0.029	0.906
		μ_b	0.006	0.033	0.034	0.752	-0.006	0.041	0.041	0.946	0.004	0.211	0.211	0.890
		$log(\phi_b)$	-0.015	0.057	0.059	0.944	0.026	0.097	0.101	0.954	-0.258	0.608	0.660	0.666
		λ_b	0.026	0.548	0.549	0.758	0.291	2.083	2.103	0.950	-	-	-	-
		$log(\phi_\delta)$	-0.015	0.210	0.211	0.986	-0.095	0.355	0.368	0.746	-0.095	0.353	0.366	0.746
		$log(\phi_\varepsilon)$	0.001	0.256	0.256	0.968	-0.059	0.380	0.385	0.748	-0.059	0.380	0.385	0.748
		$log(\phi_e)$	-0.247	0.597	0.646	0.994	-0.478	0.915	1.032	0.976	-0.437	0.804	0.915	0.976
	50	α	0.000	0.024	0.024	0.982	-0.003	0.038	0.038	0.914	0.000	0.038	0.038	0.920
		β	0.000	0.013	0.013	0.988	0.001	0.021	0.021	0.866	0.000	0.021	0.021	0.878
		μ_b	0.001	0.015	0.015	0.798	-0.008	0.023	0.024	0.994	-0.004	0.130	0.130	0.892
		$\log(\phi_b)$	-0.003	0.024	0.024	0.934	0.027	0.060	0.066	0.978	-0.131	0.456	0.474	0.588
		λ_b	0.007	0.230	0.230	0.796	0.086	0.888	0.892	0.990	-	-	-	-
		$log(\phi_\delta)$	0.003	0.112	0.112	0.992	-0.035	0.233	0.235	0.780	-0.036	0.232	0.235	0.784
		$log(\phi_\varepsilon)$	-0.001	0.138	0.138	0.980	-0.037	0.246	0.248	0.758	-0.037	0.246	0.248	0.760
		$log(\phi_e)$	-0.063	0.247	0.255	0.994	-0.167	0.511	0.538	0.970	-0.150	0.454	0.478	0.980
	100	α	-0.001	0.015	0.015	0.990	-0.005	0.029	0.030	0.918	-0.002	0.029	0.030	0.928
		β	0.001	0.008	0.008	0.986	0.003	0.016	0.016	0.854	0.001	0.016	0.016	0.858
		μ_b	0.001	0.008	0.008	0.844	-0.008	0.013	0.016	0.998	0.000	0.093	0.093	0.934
		$\log(\phi_b)$	-0.001	0.012	0.012	0.946	0.023	0.025	0.034	0.996	-0.073	0.339	0.346	0.596
		λ_b	0.001	0.126	0.126	0.844	0.001	0.381	0.381	1.000	-	-	-	-
		$log(\phi_\delta)$	-0.002	0.067	0.067	1.000	-0.019	0.192	0.193	0.722	-0.022	0.190	0.191	0.732
		$log(\phi_\varepsilon)$	-0.001	0.092	0.092	0.984	-0.023	0.187	0.189	0.708	-0.023	0.187	0.189	0.708
		$log(\phi_e)$	-0.024	0.129	0.131	0.998	-0.091	0.313	0.326	0.958	-0.089	0.309	0.321	0.962

Table 2: Relative efficiencies of ST-RMEM-based estimators relative to the N-RMEM and SN-RMEM

m	Quantity	$\lambda = 5$		$\lambda = 10$	
		MSE_{SN}/MSE_{ST}	MSE_N/MSE_{ST}	MSE_{SN}/MSE_{ST}	MSE_N/MSE_{ST}
20	α	1.669	1.564	1.645	1.615
	β	1.697	1.591	1.689	1.664
	μ_b	1.542	6.880	1.494	39.476
	$log(\phi_b)$	2.214	16.298	2.899	125.044
	λ_b	12.140	-	14.691	-
	$log(\phi_\delta)$	2.941	2.869	3.041	3.004
	$log(\phi_\epsilon)$	2.230	2.230	2.261	2.261
	$log(\phi_e)$	2.420	1.864	2.552	2.008
50	α	2.518	2.353	2.478	2.436
	β	2.599	2.427	2.542	2.479
	μ_b	2.987	11.626	2.784	79.822
	$log(\phi_b)$	6.929	46.536	7.531	384.150
	λ_b	17.174	-	15.026	-
	$log(\phi_\delta)$	4.310	4.195	4.420	4.386
	$log(\phi_\epsilon)$	3.144	3.145	3.244	3.244
	$log(\phi_e)$	4.950	3.322	4.451	3.514
100	α	4.281	4.007	4.079	4.016
	β	4.466	4.172	4.272	4.160
	μ_b	5.886	20.898	4.224	150.331
	$log(\phi_b)$	16.624	106.344	8.354	860.975
	λ_b	14.090	-	9.108	-
	$log(\phi_\delta)$	8.006	7.790	8.370	8.245
	$log(\phi_\epsilon)$	4.114	4.114	4.215	4.215
	$log(\phi_e)$	7.444	5.211	6.147	5.975

Table 3: Results of model selection criteria when the ST-RMEM is the data generating model

Set	m	Criterion	Models		
			ST-RMEM	SN-RMEM	N-RMEM
$\lambda = 5$	20	AIC	-127.966	-114.127	-114.254
		BIC	-120.001	-106.161	-107.284
	50	AIC	-334.621	-291.235	-286.153
		BIC	-319.325	275.939	-272.769
	100	AIC	-681.815	-585.752	-571.630
		BIC	-660.974	-564.911	-553.394
$\lambda = 10$	20	AIC	-130.099	-115.218	-115.177
		BIC	-122.133	-107.252	-108.206
	50	AIC	-341.021	-295.815	-288.066
		BIC	-325.725	-280.519	-274.681
	100	AIC	-695.410	-596.829	-575.334
		BIC	-674.569	-575.987	-557.098

Table 4: Estimated type I error probabilities for 5% level likelihood ratio test

Set	m	H_0 : SN-RMEM model is preferable H_1 : ST-RMEM model is preferable	H_0 : N-RMEM model is preferable H_1 : ST-RMEM model is preferable
$\lambda = 5$	20	0.067	0.044
	50	0.008	0.002
	100	< 0.001	< 0.001
$\lambda = 10$	20	0.062	0.032
	50	0.008	0.001
	100	< 0.001	< 0.001

Table 5: ML estimates, their SEs, and 95% confidence intervals for parameters of subcutaneous fat data

Parameter	ST-RMEM				SN-RMEM				N-RMEM			
	Estimate	SE	95% CI		Estimate	SE	95% CI		Estimate	SE	95% CI	
			LCL	UCL			LCL	UCL			LCL	UCL
α	0.018	0.029	-0.039	0.075	0.032	0.034	-0.035	0.099	0.033	0.035	-0.037	0.102
β	0.967	0.017	0.934	0.999	0.958	0.017	0.925	0.991	0.957	0.017	0.923	0.992
μ_b	0.959	0.202	0.564	1.355	1.237	0.503	0.252	2.223	1.830	0.134	1.567	2.093
$log(\phi_b)$	0.027	0.339	-0.638	0.691	0.105	0.426	-0.729	0.939	-0.258	0.216	-0.682	0.166
λ_b	1.687	0.842	0.036	3.338	0.988	1.288	-1.537	3.513	-	-	-	-
$log(\phi_\delta)$	-5.318	0.180	-5.671	-4.966	-5.134	0.151	-5.430	-4.837	-5.135	0.151	-5.431	-4.840
$log(\phi_\epsilon)$	-5.532	0.186	-5.896	-5.168	-5.251	0.152	-5.549	-4.952	-5.251	0.152	-5.548	-4.953
$log(\phi_e)$	-5.754	0.470	-6.676	-4.832	-5.139	0.354	-5.832	-4.445	-5.030	0.344	-5.705	-4.356

Table 6: Model selection criteria and measures of agreement for transformed subcutaneous fat data. Lower bound for CCC^* and upper bound for TDI^* are presented

Models	AIC	BIC	CCC^*		TDI^*	
			Estimate	95% Bound	Estimate	95% Bound
ST-RMEM	-278.806	-259.171	0.990	0.984	0.034	0.050
SN-RMEM	-261.149	-241.514	0.987	0.977	0.056	0.083
N-RMEM	-264.469	-247.289	0.987	0.980	0.232	0.263

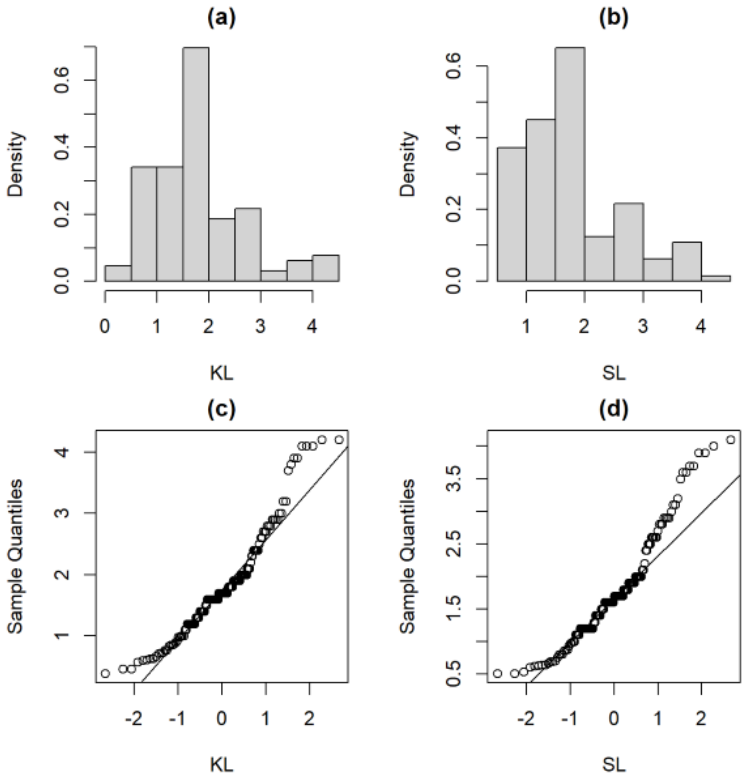

Figure 1: Histogram (a-b) and normal Q-Q plot (c-d) of the subcutaneous fat data. Left panel for 'KL' observer and right panel for 'SL' observer

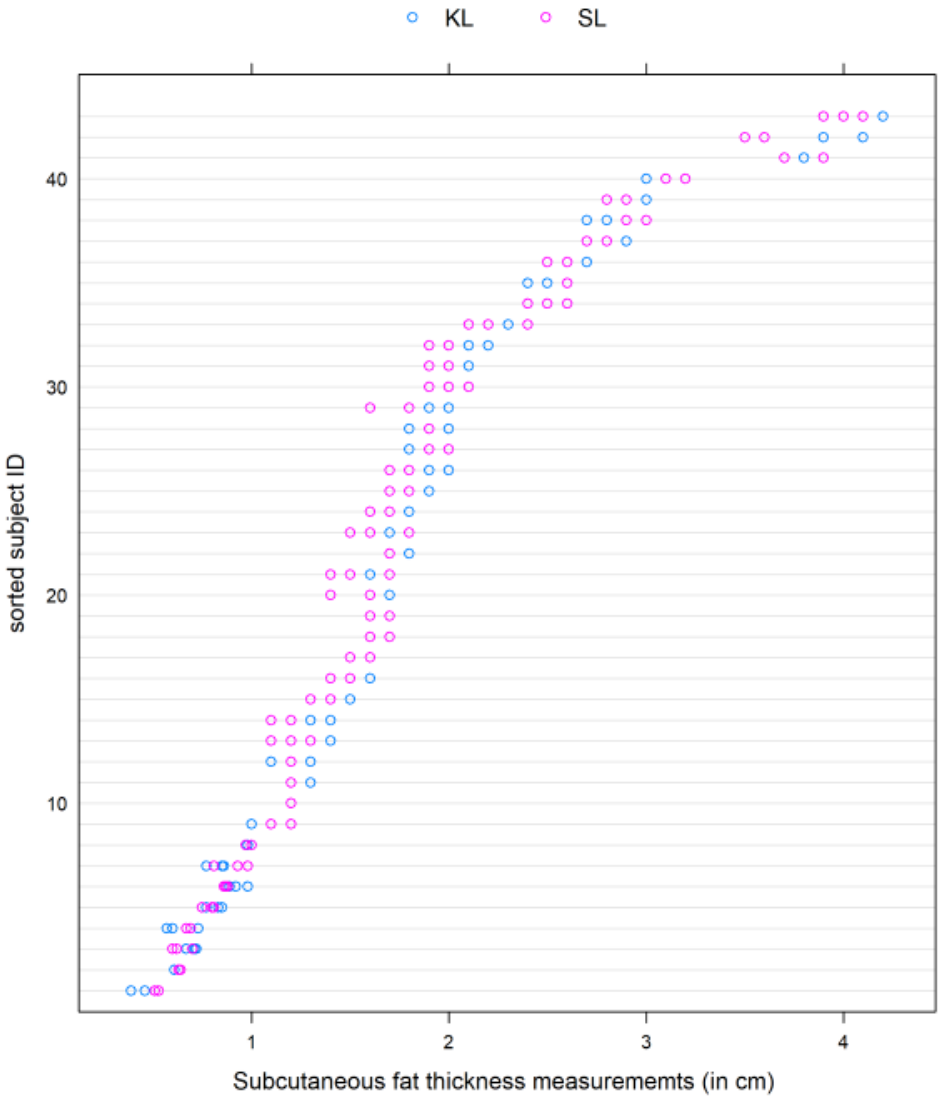

Figure 2: Trellis plot for subcutaneous fat data

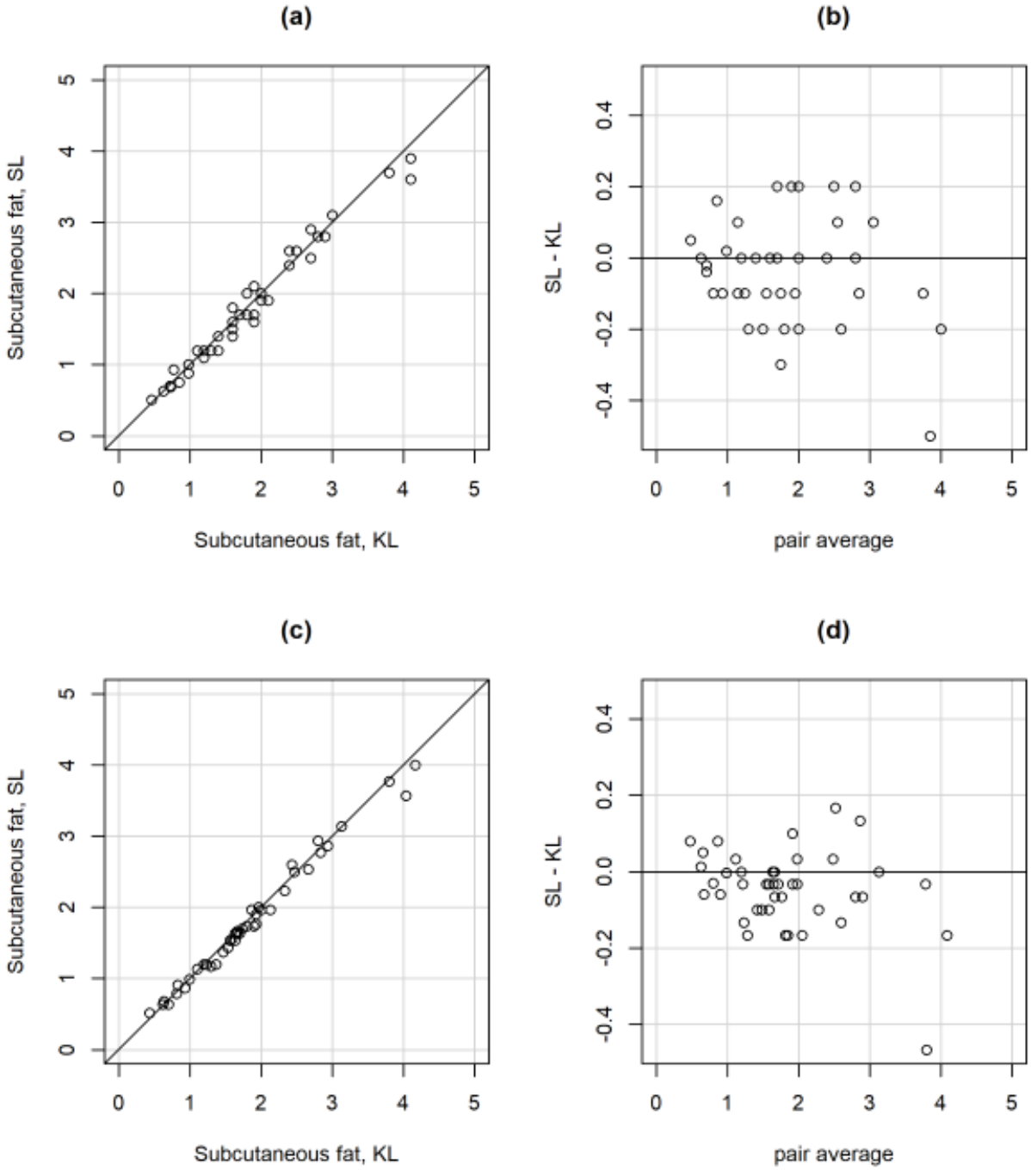

Figure 3: Scatterplot with line of equality (left) and Bland-Altman plots with zero line (right) for subcutaneous fat thickness measurements. One measurement per method from each of the 43 subjects is randomly selected for this plot. Same as the top panel but based on 43 average measurements