Journal of Biopharmaceutical Statistics

ISSN: 1054-3406 (Print) 1520-5711 (Online) Journal homepage: http://www.tandfonline.com/loi/lbps20

Bayesian inference for agreement measures

Ignacio Vidal & Mário de Castro

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
http://dx.doi.org/10.1080/10543406.2016.1226323

ORIGINAL ARTICLES

Bayesian inference for agreement measures

Ignacio Vidal[a] and Mário de Castro[b]

[a]Instituto de Matemática y Física, Universidad de Talca, Talca, Chile; [b]Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos, SP, Brasil

ABSTRACT

The agreement of different measurement methods is an important issue in several disciplines like, for example, Medicine, Metrology, and Engineering. In this article, some agreement measures, common in the literature, were analyzed from a Bayesian point of view. Posterior inferences for such agreement measures were obtained based on well-known Bayesian inference procedures for the bivariate normal distribution. As a consequence, a general, simple, and effective method is presented, which does not require Markov Chain Monte Carlo methods and can be applied considering a great variety of prior distributions. Illustratively, the method was exemplified using five objective priors for the bivariate normal distribution. A tool for assessing the adequacy of the model is discussed. Results from a simulation study and an application to a real dataset are also reported.

ARTICLE HISTORY
Received 28 September 2015
Accepted 9 August 2016

KEYWORDS
Accuracy; concordance correlation coefficient; coverage probability; precision; total deviation index

1. Introduction

Many approaches have been proposed to quantify the strength of agreement between measurements carried out with two different methods. A comprehensive account can be found in Lin et al. (2002), Dunn (2004), Barnhart et al. (2007a), Barnhart et al. (2007b), Choudhary & Nagaraja (2007), and Lin et al. (2012). In this work, we focus on some tools discussed by Lin et al. (2002), but from a Bayesian point of view. In particular, the concordance correlation coefficient (hereafter, CCC) proposed by Lin (1989) aims to quantify agreement by assessing the departure of the measurements from the identity line. The CCC has gained attention and reached widespread acceptance among practitioners in diverse application areas.

Lin et al. (2002) and Lin et al. (2012) give a thorough exposition of tools and methods related to agreement assessment. It is worth to recall that for some measures, inference based on frequentist methods is rather intractable, as pointed out by Lin et al. (2002). We present an alternative for making inference in agreement studies from a Bayesian point of view. Our approach is built on objective priors (Berger & Sun, 2008). An interpretation for Bayesian inference using objective priors is based on finding a good overall prior, defined as one that yields reasonable frequentist coverage properties when applied to quantities of interest (Berger et al., 2015, Section 1.2) such as an agreement measure. The procedure is effective, computationally simple, and does not require asymptotic approximations. Therefore, by using objective priors, our proposal can be seen as a general method for assessing the agreement between two measures.

Previous works on agreement measures from a Bayesian point of view include Yin et al. (2008), Broemeling (2009), Schluter (2009), and Choudhary and Yin (2010), among others. They approach several problems related to inferences on agreement measures not covered in our work, like for example sample size determination, design with repeated measurements, categorical outcomes, and more than two measurement methods. Their Bayesian solutions depend on Markov Chain Monte Carlo (MCMC) methods. Our proposal does not rely on asymptotic results and MCMC methods are

CONTACT Ignacio Vidal ividal@utalca.cl Universidad de Talca, 2 Norte 685 Talca, Chile.

not required at all. The sampler in this work delivers exact samples of the posterior distribution, so that we do not have to deal with issues such as starting points for the chains, burn-in samples, thinning, and convergence of chains.

The remainder of the article unfolds as follows. In Section 2, we describe some measures of agreement from the literature. In Section 3, we discuss different prior distributions for the parameters of the bivariate normal distribution and describe simple schemes to draw samples from the posterior distribution under five prior distributions. Furthermore, a graphical device for model checking is presented. A simulation study is carried out in Section 4. In Section 5, the proposed methodology is further illustrated with a real dataset from the literature. We conclude with some general remarks in Section 6.

2. Assessing agreement

In this section, we present some well-known methods for assessing agreement between a set of paired observations. Details about such methods can be found in a broad existing literature (see, for example, Lin et al., 2002, 2012). In this section, we give a brief account of some agreement and related measures.

A way to assess agreement between two random variables X and Y is the mean squared deviation (MSD) given by:

$$\mathrm{MSD} = \mathrm{E}[(X-Y)^2] = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}, \tag{1}$$

assuming that the joint distribution of (X, Y) has finite second moments with mean vector $(\mu_X, \mu_Y)'$ and covariance matrix with elements $\mathrm{Var}(X) = \sigma_X^2$, $\mathrm{Var}(Y) = \sigma_Y^2$ and $\mathrm{Cov}(X, Y) = \sigma_{XY}$.

Another measure of agreement between two random variables is the CCC, which is defined by:

$$\rho_c = 1 - \frac{\mathrm{MSD}}{\mathrm{MSD}|_{\sigma_{XY}=0}} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2},$$

where $\mathrm{MSD}|_{\sigma_{XY}=0}$ means that the expectation in (1) is computed under the assumption of null correlation between X and Y. The CCC can also be written as $\rho_c = \chi_a \rho$, where ρ is the Pearson correlation coefficient ($\rho = \sigma_{XY}/\sigma_X\sigma_Y$):

$$\chi_a = \frac{2}{\bar{\omega} + 1/\bar{\omega} + \nu^2}, \bar{\omega} = \frac{\sigma_Y}{\sigma_X} \quad \text{and} \quad \nu^2 = \frac{(\mu_X - \mu_Y)^2}{\sigma_X \sigma_Y}.$$

Notice that if the marginal distributions of X and Y are the same, then $\chi_a = 1$. Also, $\chi_a = 1$ implies $\mu_X = \mu_Y$ and $\sigma_X = \sigma_Y$. Hence, χ_a can be seen as an accuracy coefficient. Obviously, since ρ quantifies the correlation between X and Y, then ρ is a precision coefficient. Moreover, since $\rho_c = \chi_a \rho$, it is easy to see that a perfect agreement ($X \overset{d}{=} Y$) implies $\rho_c = 1$, a perfect disagreement ($X \overset{d}{=} -Y$) implies $\rho_c = -1$, and $\rho_c = 0$ corresponds to no agreement, where $X \overset{d}{=} Y$ means that the variables X and Y have the same distribution.

The total deviation index (TDI), proposed by Lin (2000) (see also Lin et al., 2015; Perez-Jaume & Carrasco, 2015), is an intuitive measure of agreement. This index gives an upper bound for the difference between two random variables within a given probability range. The TDI is the solution of the equation:

$$p_0 = \mathrm{P}(|X - Y| < \kappa_{p_0}) \tag{2}$$

with respect to κ_{p_0}, where p_0 is a fixed probability value. If we assume that the distribution of $(X, Y)'$ is bivariate normal, as in Section 3, then:

$$p_0 = \mathrm{P}(|X - Y| < \kappa_\pi) = \Phi\left(\frac{\kappa_\pi - \mu_D}{\sigma_D}\right) - \Phi\left(\frac{-\kappa_\pi - \mu_D}{\sigma_D}\right), \tag{3}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of the standard normal distribution, $D = X - Y \sim N(\mu_D, \sigma_D^2)$, with $\mu_D = \mu_X - \mu_Y$ and $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$. Equivalently:

$$p_0 = \mathrm{P}\left((X - Y)^2 < \kappa_{p_0}^2\right) = \chi_1^2\left(\frac{\kappa_{p_0}^2}{\sigma_D^2}, \frac{\mu_D^2}{\sigma_D^2}\right), \tag{4}$$

where $\chi_1^2\left(\cdot, \mu_D^2/\sigma_D^2\right)$ stands for the cdf of the noncentral Chi-squared distribution with 1 degree of freedom and noncentrality parameter μ_D^2/σ_D^2.

From expressions (2)–(4), we can obtain the coverage probability (CP) p_κ for a given κ. The CP is the probability that the absolute difference $|X - Y|$ does not exceed a predetermined bound κ, which is easier to compute than the TDI. Notice that higher the p_κ, higher the agreement.

3. Bayesian inference

In this work, we assume that $(X_i, Y_i)' | \boldsymbol{\mu}, \boldsymbol{\Sigma} \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$, where:

$$\boldsymbol{\mu} = (\mu_X, \mu_Y)' \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}. \tag{5}$$

In order to draw samples from the posterior distribution of the measures in Section 2, we take advantage of samplers developed for the parameters in (5). Bayesian inferences for $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$ and several functions of them have been widely studied in the literature. The main goal of our work is to implement Bayesian inference for the quantities mentioned in Section 2. The works by Berger & Sun (2008) and Berger et al. (2015) give a methodology to carry out Bayesian inference for a broad class of parameter functions under a general class of objective prior distributions. These works serve as our starting point.

For several objective prior distributions $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Berger & Sun (2008) present the posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For example, if $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\mu}|\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma})$ with $\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \propto 1$, then:

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}, \mathbf{y}) \propto \phi_2\left(\boldsymbol{\mu}|(\bar{x}, \bar{y})'^{-1}\boldsymbol{\Sigma}\right)|\boldsymbol{\Sigma}|^{-(n-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(S\boldsymbol{\Sigma}^{-1})\right)\pi(\boldsymbol{\Sigma}),$$

where $\phi_2(\cdot|\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ denotes the density function of the bivariate normal distribution, $\bar{x} = \sum_{i=1}^n x_i/n$, $\bar{y} = \sum_{i=1}^n y_i/n$, $|\boldsymbol{\Sigma}| = \sigma_X^2\sigma_Y^2(1 - \rho^2)$, tr(A) indicates the trace of the matrix $\mathbf{A}$ and

$$S = \sum_{i=1}^n \begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \end{pmatrix}\begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \end{pmatrix}' = \begin{pmatrix} s_{11} & r(s_{11}s_{22})^{1/2} \\ r(s_{11}s_{22})^{1/2} & s_{22} \end{pmatrix},$$

with $r = s_{12}/(s_{11}s_{22})^{1/2}$, $s_{11} = \sum_{i=1}^n (x_i - \bar{x})^2$, $s_{22} = \sum_{i=1}^n (y_i - \bar{y})^2$ and $s_{12} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$. There are two well-known cases of $\pi(\mu, \boldsymbol{\Sigma}|x, y)$:

3.1. *Jeffreys' prior*

If we take the Jefreys' prior $\pi_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \sigma_X^{-2}\sigma_Y^{-2}(1 - \rho^2)^{-2}$, then:

$$\boldsymbol{\mu} |\ \boldsymbol{\Sigma}, x, y \sim N_2\left((\bar{x}, \bar{y})', n^{-1}\ \boldsymbol{\Sigma}\right) \quad \text{and} \quad \boldsymbol{\Sigma} | \mathbf{x}, \mathbf{y} \sim IW_2\left(S^{-1}, n\right), \tag{6}$$

where the density function of the bivariate inverse Wishart distribution, $IW_2(A^{-1}, \nu)$, has kernel $|\ \boldsymbol{\Sigma}|^{-(\nu+3)/2} \exp(-\mathrm{tr}(A\ \boldsymbol{\Sigma}^{-1})/2)$.

3.2. *Independence Jeffreys' prior*

If we consider the independence Jeffreys' prior $\pi_{IJ}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \propto \sigma_X^{-1}\sigma_Y^{-1}(1-\rho^2)^{-3/2}$, then:

$$\boldsymbol{\mu} |\ \boldsymbol{\Sigma}, x, y \sim N_2\left((\bar{x}, \bar{y})', n^{-1}\ \boldsymbol{\Sigma}\ \right) \quad \text{and} \quad \boldsymbol{\Sigma}\ | x, y \sim IW_2\left(S^{-1}, n-1\right). \tag{7}$$

More generally, Bayesian inference for μ and $\boldsymbol{\Sigma}$ from the objective prior family:

$$\pi_{ab}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \propto \frac{1}{\sigma_X^{3-a}\sigma_Y^{2-b}(1-\rho^2)^{2-b/2}} \tag{8}$$

is detailed by Berger & Sun (2008), noticing that the Jeffreys' and independence Jeffreys' priors correspond to π_{10} and π_{21} in (8), respectively. The objective prior family $\pi_{ab}(\ \boldsymbol{\mu},\ \boldsymbol{\Sigma})$ includes several objective priors, which are mentioned in Berger et al. (2015) as overall objective priors. That is, objective posterior distributions based on these overall objective priors are useful to produce reasonable inferences on all parameters of the model or functions of them.

The right-Haar prior $\pi_H(\boldsymbol{\mu},\ \boldsymbol{\Sigma}\ ) \propto \sigma_X^{-2}(1-\rho^2)^{-1}$ is another special case of (8) given by $\pi_H = \pi_{12}$. Samples from the posterior distributions of σ_X, σ_Y, ρ, μ_X, and μ_Y under π_H are drawn repeating the steps in expressions (16)–(19) in Berger & Sun (2008). These distributions are called constructive posterior distributions. For the sake of completeness, the steps are presented in Appendix A.

Two additional priors recommended by Berger & Sun (2008) Berger et al. (2015) are given by:

$$\pi_{R\rho}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \propto \frac{1}{\sigma_X\sigma_Y(1-\rho^2)} \quad \text{and} \quad \pi_{R\sigma}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \propto \frac{(1+\rho^2)^{1/2}}{\sigma_X\sigma_Y(1-\rho^2)}, \tag{9}$$

which correspond to reference priors for ρ and σ_X (or σ_Y) in (5), respectively. Samples from the posterior distributions of σ_X, σ_Y, ρ, μ_X, and μ_Y under $\pi_{R\rho}$ and $\pi_{R\sigma}$ can be obtained by running the acceptance–rejection scheme described in Berger & Sun (2008). Let π_R denote either $\pi_{R\rho}$ or $\pi_{R\sigma}$. Suppose that $M = \sup_{(\sigma_X,\sigma_Y,\rho)} \pi_R(\boldsymbol{\mu},\ \boldsymbol{\Sigma})/\pi_{IJ}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \in (0,\infty)$. In fact, $M = 1$ for the distributions in (9) (see Berger & Sun, 2008, Table 3). The steps for generating samples from the posterior distribution of the model parameters are as follows:

Simulation step. Generate $(\sigma_X, \sigma_Y, \rho)$ from the IW_2 distribution in (7) and independently generate $u \sim \mathrm{Uniform}(0,1)$.

Rejection step. If $u \leq \pi_R(\boldsymbol{\mu},\ \boldsymbol{\Sigma})/[M\pi_{IJ}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})]$, then accept $(\sigma_X, \sigma_Y, \rho)$; else, return to the simulation step. Upon acceptance, generate (μ_X, μ_Y) from the N_2 distribution in (7). For the priors in (9), the inequalities in the rejection step are $u \leq (1-\rho^2)^{1/2}$ and $u \leq (1-\rho^4)^{1/2}$, respectively. For more details, the reader is referred to Berger & Sun (2008, Section 2.1). Notice that the reference priors in (9) require simulation and rejection steps that are not necessary when we use the other three priors in this work.

Therefore, once we have samples from the posterior distribution of μ_X, μ_Y, σ_X, σ_Y, and ρ, we can obtain Bayesian inferences for the measures in Section 2, which are functions of μ_X, μ_Y, σ_X, σ_Y, and ρ. As can be seen in Section 4, this strategy leads to estimators with desirable properties. We stress that the implementation of the required steps is straightforward using free software; e.g., R (R Core Team, 2015). The computational codes are available on request from the authors.

3.3. *Model assessment*

Before drawing inferences on quantities of interest, the goodness of fit of the model should be assessed. To deal with this point, we adopt a procedure based on a pivotal quantity, as advocated by Johnson (2007). Taking into account the distribution of the pair $Z_i = (X_i, Y_i)'$ in (5), it follows that $(Z_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(Z_i - \boldsymbol{\mu}) \overset{iid}{\sim} \chi^2_2$, $i = 1, \ldots, n$. Therefore, under the postulated model, $Q(Z, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n}(Z_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(Z_i - \boldsymbol{\mu})$ is distributed as a central χ^2_{2n} random variable. The pivotal quantity $Q(Z, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a key element for model checking. If $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ denote the data-generating values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}_{\text{post}}$ and $\boldsymbol{\Sigma}_{\text{post}}$ are drawn from the posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given Z, Johnson (2007) proves that $Q(Z, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $Q(Z, \boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$ have the same distribution. A simple and useful device for model assessment can be based on graphical comparisons of the posterior distribution of this pivotal quantity and its reference distribution. This graphical diagnostics may reveal model inadequacy.

4. Simulations

In this section, we show results from an extensive simulation study comprising different scenarios, which were analyzed by Feng et al. (2015). These scenarios are described in Table 1.

For each parameter combination given in Table 1, we generate 1000 datasets with sample size $n(= 30, 100)$ and estimate MSD, CCC, precision (ρ), accuracy (χ_a), TDI, CP, and the proportion of 95% highest posterior density (HPD) intervals that include the true values of them and their corresponding average lengths. HPD intervals were computed following the method of Chen et al. (2000, Ch. 7). For the sake of comparison, we also applied the frequentist methods explained in Lin et al. (2002).

Table 2 shows the simulation results for some of the scenarios considered. For the other combination of parameters considered, the results obtained yield good performance like those ones shown in Table 2. The third column of Table 2 presents the theoretical value corresponding to each one of the agreement measures. The next columns give the estimates for those theoretical values corresponding to two sample sizes, three of the five prior distributions analyzed, and frequentist estimates. Bayesian estimates are averages based on 1000 posterior medians that were computed from 5000 samples from

Table 1. Combination of parameters used to generate bivariate normal data.

Combination	μ_X	μ_Y	σ^2_X	σ^2_Y	ρ
1	100	100	100	100	0.99
2					0.9
3					0.7
4					0.5
5	100	100	100	125	0.99
6					0.9
7					0.7
8					0.5
9	100	105	100	100	0.99
10					0.9
11					0.7
12					0.5
13	100	105	100	125	0.99
14					0.9
15					0.7
16					0.5

Table 2. Averages of the posterior median of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

			Jeffreys		Independence	Jeffreys	Right-Haar		Frequentist methods	
Comb.	Measure	True value	n = 30	100	30	100	30	100	30	100
1	MSD	2	2.197	2.043	2.275	2.063	2.348	2.084	2.075	2.018
	CCC	0.99	0.989	0.990	0.989	0.990	0.988	0.989	0.989	0.990
	Precision	0.99	0.990	0.990	0.990	0.990	0.989	0.990	0.990	0.990
	Accuracy	1	0.999	1.000	0.999	1.000	0.999	1.000	0.998	1.000
	TDI	2.326	2.344	2.322	2.386	2.334	2.425	2.346	2.348	2.331
	CP	0.9	0.885	0.897	0.879	0.895	0.873	0.893	0.895	0.899
6	MSD	23.754	26.008	24.328	26.930	24.572	27.792	24.820	24.630	23.922
	CCC	0.894	0.882	0.890	0.882	0.890	0.875	0.888	0.885	0.892
	Precision	0.9	0.900	0.899	0.900	0.899	0.894	0.898	0.897	0.899
	Accuracy	0.994	0.983	0.991	0.983	0.991	0.982	0.990	0.961	0.971
	TDI	8.017	8.069	8.014	8.215	8.055	8.348	8.095	8.089	8.025
	CP	0.9	0.885	0.896	0.879	0.894	0.873	0.893	0.895	0.899
11	MSD	85	90.471	86.668	92.842	87.285	94.357	87.711	86.996	85.683
	CCC	0.622	0.604	0.617	0.606	0.617	0.594	0.614	0.612	0.616
	Precision	0.7	0.700	0.701	0.700	0.701	0.688	0.697	0.699	0.698
	Accuracy	0.889	0.870	0.882	0.873	0.883	0.871	0.882	0.854	0.877
	TDI	15.136	15.430	15.221	15.368	15.276	15.771	15.314	15.141	15.147
	CP	0.9	0.891	0.897	0.886	0.896	0.883	0.895	0.897	0.899
16	MSD	138.197	146.321	141.773	150.700	142.955	153.361	143.687	140.488	139.358
	CCC	0.447	0.428	0.438	0.430	0.439	0.417	0.435	0.435	0.444
	Precision	0.5	0.499	0.498	0.499	0.498	0.487	0.495	0.497	0.500
	Accuracy	0.894	0.868	0.883	0.871	0.884	0.868	0.883	0.833	0.865
	TDI	19.336	19.693	19.523	19.989	19.605	20.170	19.655	19.309	19.364
	CP	0.9	0.891	0.896	0.886	0.894	0.883	0.894	0.898	0.899

the posterior distributions obtained by the different prior distributions considered in Section 3. Frequentist estimates are averages based on 1000 estimates computed following the methods given by Lin et al. (2002). From Table 2, we can see very good estimates for all priors considered and also for the frequentist methods, which were better for greater sample size. As we can see, the results are consistent among the three priors, so that within our study, the inferences are minimally sensitive to prior specification. However, less precise estimates were obtained under the right-Haar prior distribution, but the best estimates were obtained with the Jeffreys' and independence Jeffreys' priors. From a practical perspective, due to its simplicity, the posterior distribution arising from the Jeffreys' prior in (6) might be recommended as the prior of choice.

Table 3 shows the estimates of the CP of the 95% HPD intervals and their corresponding average lengths for MSD, CCC, precision, accuracy, TDI, and CP under the Jeffreys' prior. Results from the remaining prior distributions are presented in the Supplemental Web Materials (Appendix B).

Since the support of the posterior distribution of χ_a is the interval $(0, 1]$, there is no sense measuring the CP when the true value of χ_a is equal to 1. That happens for the combinations 1 to 4 in Table 1. However, estimates for χ_a were very good, see for example the parameter combination 1 in Table 2. As we expected, better results were obtained for greater sample size. In general, the CPs are near to the nominal value (0.95). However, the worst results were obtained for estimating ρ (precision) when $n = 30$ was used. The behavior of the estimated CP was in general similar for the five prior distributions considered, as can be seen in the tables in the Supplemental Web Materials (Appendix B). For estimating ρ, the best results were not achieved with the Jeffreys' prior.

Table 4 shows frequentist estimates for the agreement measures (Lin et al., 2002). It displays the CP of the 95% asymptotic confidence intervals and their average lengths based on 1000 replications. As a general pattern, Tables 3 and 4 reveal that the CPs of the HPD intervals are closer to the nominal value when compared to the asymptotic intervals. Indeed, there are some particular cases where Bayesian estimates are clearly better like, for example, in the estimation of the accuracy coefficient and in the estimation of TDI for the scenarios 9 and 13. Moreover, in scenarios 8 with

Table 3. Jeffreys' prior.

Comb.	n	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.957 (2.378)	0.946 (0.017)	0.927 (0.015)	–	0.943 (1.233)	0.939 (0.177)
	100	0.949 (1.153)	0.953 (0.008)	0.943 (0.008)	–	0.948 (0.649)	0.943 (0.094)
2	30	0.948 (23.363)	0.940 (0.150)	0.917 (0.139)	–	0.938 (3.867)	0.934 (0.175)
	100	0.952 (11.548)	0.952 (0.077)	0.943 (0.075)	–	0.944 (2.054)	0.943 (0.094)
3	30	0.969 (70.249)	0.923 (0.368)	0.918 (0.360)	–	0.951 (6.700)	0.947 (0.175)
	100	0.951 (34.644)	0.950 (0.200)	0.948 (0.198)	–	0.942 (3.556)	0.945 (0.094)
4	30	0.953 (117.820)	0.939 (0.510)	0.939 (0.521)	–	0.942 (8.675)	0.940 (0.176)
	100	0.945 (57.502)	0.934 (0.288)	0.937 (0.290)	–	0.940 (4.580)	0.936 (0.094)
5	30	0.957 (4.281)	0.950 (0.021)	0.922 (0.015)	0.951 (0.013)	0.945 (1.653)	0.941 (0.176)
	100	0.946 (2.092)	0.952 (0.010)	0.949 (0.008)	0.945 (0.006)	0.939 (0.874)	0.938 (0.094)
6	30	0.954 (28.189)	0.948 (0.153)	0.926 (0.140)	0.964 (0.048)	0.946 (4.247)	0.941 (0.177)
	100	0.961 (13.721)	0.964 (0.079)	0.951 (0.075)	0.957 (0.020)	0.956 (2.239)	0.956 (0.094)
7	30	0.953 (81.561)	0.930 (0.376)	0.914 (0.369)	0.950 (0.107)	0.940 (7.215)	0.936 (0.177)
	100	0.951 (39.894)	0.939 (0.200)	0.947 (0.199)	0.972 (0.040)	0.948 (3.817)	0.938 (0.095)
8	30	0.956 (134.794)	0.939 (0.513)	0.927 (0.525)	0.968 (0.145)	0.949 (9.289)	0.942 (0.178)
	100	0.951 (64.911)	0.950 (0.287)	0.942 (0.289)	0.961 (0.055)	0.944 (4.868)	0.947 (0.094)
9	30	0.946 (10.634)	0.935 (0.118)	0.926 (0.015)	0.936 (0.110)	0.951 (1.412)	0.937 (0.162)
	100	0.946 (5.693)	0.938 (0.063)	0.957 (0.008)	0.933 (0.059)	0.950 (0.754)	0.947 (0.091)
10	30	0.948 (39.497)	0.947 (0.221)	0.936 (0.138)	0.943 (0.158)	0.948 (4.413)	0.938 (0.164)
	100	0.944 (20.863)	0.948 (0.120)	0.946 (0.075)	0.941 (0.086)	0.950 (2.371)	0.943 (0.091)
11	30	0.954 (89.771)	0.950 (0.384)	0.931 (0.364)	0.948 (0.221)	0.951 (7.456)	0.947 (0.171)
	100	0.945 (45.943)	0.943 (0.213)	0.948 (0.199)	0.923 (0.126)	0.944 (3.974)	0.936 (0.093)
12	30	0.947 (137.374)	0.945 (0.489)	0.940 (0.521)	0.954 (0.258)	0.943 (9.434)	0.932 (0.171)
	100	0.948 (69.040)	0.949 (0.277)	0.947 (0.290)	0.938 (0.153)	0.948 (4.987)	0.942 (0.093)
13	30	0.950 (14.480)	0.940 (0.113)	0.931 (0.015)	0.940 (0.106)	0.945 (1.891)	0.935 (0.162)
	100	0.941 (7.736)	0.944 (0.061)	0.947 (0.008)	0.942 (0.057)	0.935 (1.011)	0.928 (0.091)
14	30	0.957 (44.115)	0.951 (0.216)	0.929 (0.138)	0.951 (0.153)	0.955 (4.787)	0.942 (0.164)
	100	0.954 (23.580)	0.950 (0.119)	0.944 (0.075)	0.947 (0.085)	0.953 (2.590)	0.951 (0.092)
15	30	0.958 (99.081)	0.936 (0.383)	0.937 (0.364)	0.945 (0.218)	0.953 (7.878)	0.949 (0.170)
	100	0.950 (51.102)	0.951 (0.213)	0.947 (0.200)	0.950 (0.124)	0.952 (4.227)	0.947 (0.094)
16	30	0.957 (151.530)	0.933 (0.492)	0.918 (0.523)	0.954 (0.251)	0.955 (9.960)	0.945 (0.171)
	100	0.954 (77.693)	0.948 (0.278)	0.945 (0.292)	0.919 (0.152)	0.952 (5.305)	0.949 (0.094)

Coverage probability (and average length) of the 95% HPD interval for some measures based on 1000 replications and 5000 samples from the posterior distribution.

$n = 30$ and 16 with $n = 100$ Bayesian and frequentist estimates of TDI gave the same CP, but the average length of the intervals was shorter for the Bayesian estimates. A similar behavior was observed for the other prior distributions considered, as can be seen in the Supplemental Web Materials (Appendix B).

5. Example

In this section, we apply the methodology for estimating the six measures presented in Section 2 to a real dataset. The data in our example were presented by Luiz et al. (2003) and Liao (2015) and consist of two different measurements of the inferior pelvic infundibular angle for 52 kidneys, which were taken by tomography (T) (ranging from 40° to 108°, with sample mean and standard deviation equal to 79.5° and 17.2°, respectively) and urography (U) (ranging from 40° to 105°, with sample mean and standard deviation equal to 75.8° and 15.3°, respectively). Since the tomography measurements are more expensive, the idea is to assess how much agreement there is between measurements taken by urography and tomography. If attainable, reliable measurements taken by urography are convenient for diagnosis and treatment of renal lithiasis. All results are based on 5000 samples from the posterior distribution.

Due to its simplicity and taking into account the results of the simulation study in Section 4, we focus on the Jeffreys' prior. In Figure 1, we show the histogram of the posterior samples of $Q\left(Z, \boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}\right)$ in Section 3.3 under the Jeffreys' prior together with the χ^2_{104} density function

Table 4. Frequentist estimates.

Comb.	n	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.946 (2.273)	0.948 (0.018)	0.952 (0.017)	0.006 (0.050)	0.946 (1.243)	0.959 (0.200)
	100	0.953 (1.145)	0.950 (0.008)	0.956 (0.008)	0.007 (0.037)	0.953 (0.655)	0.965 (0.100)
2	30	0.968 (22.691)	0.945 (0.158)	0.957 (0.157)	0.008 (0.191)	0.968 (3.932)	0.980 (0.201)
	100	0.965 (11.401)	0.972 (0.078)	0.972 (0.078)	0.006 (0.096)	0.965 (2.067)	0.972 (0.100)
3	30	0.956 (68.970)	0.945 (0.386)	0.954 (0.391)	0.006 (0.329)	0.956 (6.853)	0.967 (0.202)
	100	0.944 (34.380)	0.942 (0.202)	0.942 (0.203)	0.003 (0.173)	0.944 (3.587)	0.950 (0.100)
4	30	0.951 (113.497)	0.948 (0.530)	0.959 (0.547)	0.004 (0.416)	0.951 (8.790)	0.968 (0.201)
	100	0.947 (57.210)	0.938 (0.293)	0.938 (0.296)	0.008 (0.203)	0.947 (4.628)	0.957 (0.100)
5	30	0.963 (4.090)	0.949 (0.021)	0.952 (0.017)	0.105 (0.024)	0.963 (1.670)	0.975 (0.200)
	100	0.949 (2.083)	0.941 (0.010)	0.948 (0.008)	0.003 (0.012)	0.949 (0.883)	0.958 (0.101)
6	30	0.943 (26.987)	0.939 (0.159)	0.959 (0.156)	0.455 (0.145)	0.943 (4.282)	0.963 (0.199)
	100	0.949 (13.572)	0.944 (0.079)	0.941 (0.077)	0.272 (0.044)	0.950 (2.254)	0.960 (0.100)
7	30	0.962 (78.065)	0.949 (0.386)	0.958 (0.392)	0.471 (0.293)	0.962 (7.289)	0.970 (0.201)
	100	0.955 (39.156)	0.947 (0.203)	0.949 (0.204)	0.459 (0.114)	0.955 (3.830)	0.966 (0.100)
8	30	0.949 (129.975)	0.945 (0.524)	0.947 (0.544)	0.437 (0.384)	0.949 (9.402)	0.959 (0.201)
	100	0.953 (64.630)	0.944 (0.293)	0.943 (0.297)	0.472 (0.164)	0.953 (4.919)	0.961 (0.100)
9	30	0.972 (12.905)	0.939 (0.127)	0.946 (0.017)	0.938 (0.118)	0.014 (1.999)	0.947 (0.173)
	100	0.968 (6.066)	0.951 (0.065)	0.956 (0.008)	0.951 (0.061)	0.000 (0.954)	0.952 (0.090)
10	30	0.939 (41.712)	0.948 (0.235)	0.958 (0.154)	0.936 (0.179)	0.930 (4.886)	0.935 (0.176)
	100	0.947 (21.545)	0.948 (0.125)	0.953 (0.079)	0.944 (0.091)	0.917 (2.596)	0.933 (0.092)
11	30	0.943 (90.475)	0.932 (0.397)	0.941 (0.385)	0.874 (0.271)	0.940 (7.647)	0.940 (0.178)
	100	0.948 (46.308)	0.947 (0.217)	0.946 (0.204)	0.922 (0.137)	0.947 (4.065)	0.938 (0.092)
12	30	0.952 (135.971)	0.939 (0.504)	0.951 (0.545)	0.854 (0.343)	0.952 (9.491)	0.951 (0.180)
	100	0.950 (70.146)	0.947 (0.280)	0.953 (0.295)	0.920 (0.171)	0.950 (5.066)	0.942 (0.093)
13	30	0.966 (16.516)	0.955 (0.120)	0.950 (0.017)	0.911 (0.121)	0.262 (2.474)	0.946 (0.174)
	100	0.945 (8.110)	0.955 (0.062)	0.952 (0.008)	0.835 (0.063)	0.002 (1.236)	0.936 (0.091)
14	30	0.952 (46.254)	0.933 (0.233)	0.938 (0.157)	0.863 (0.181)	0.945 (5.213)	0.945 (0.176)
	100	0.972 (23.780)	0.962 (0.121)	0.956 (0.077)	0.859 (0.093)	0.948 (2.764)	0.966 (0.091)
15	30	0.956 (98.855)	0.943 (0.396)	0.951 (0.386)	0.802 (0.270)	0.955 (8.030)	0.957 (0.177)
	100	0.955 (51.419)	0.952 (0.217)	0.947 (0.205)	0.845 (0.138)	0.954 (4.303)	0.949 (0.092)
16	30	0.945 (150.489)	0.949 (0.505)	0.945 (0.546)	0.802 (0.344)	0.945 (10.001)	0.951 (0.180)
	100	0.952 (77.570)	0.943 (0.281)	0.940 (0.295)	0.826 (0.170)	0.952 (5.337)	0.843 (0.092)

Coverage probability (and average length) of the 95% confidence interval for some measures based on 1000 replications.

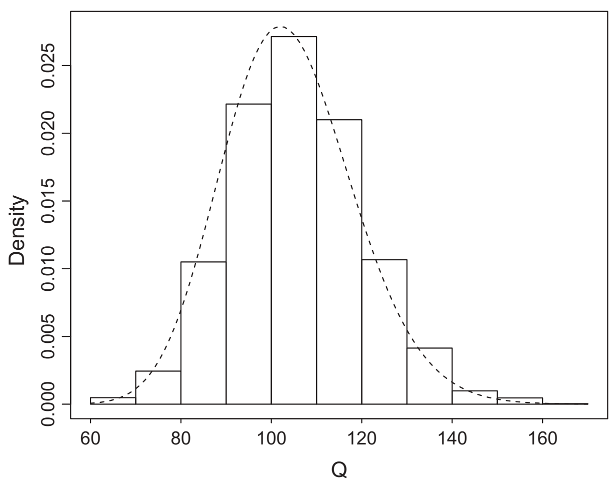

Figure 1. Histogram of the posterior samples of $Q(Z,\ \mu_{\text{post}},\ \Sigma_{\text{post}})$ under the Jeffreys' prior with the χ^2_{104} density function as reference.

in a dashed line. Since the histogram and the density function overlap, this plot does not suggest serious departures from the assumed model in our example.

Table 5. Some agreement measures: (a) posterior medians (and 95% HPD intervals) under the Jeffreys' prior based on 5000 samples from the posterior distribution and (b) estimates (and asymptotic 95% confidence intervals).

Measure	(a)	(b)
MSD	104.407	101.024
	(69.062, 150.126)	(68.309, 149.567)
CCC	0.808	0.810
	(0.704, 0.889)	(0.694, 0.885)
Precision	0.823	0.820
	(0.723, 0.901)	(0.704, 0.893)
Accuracy	0.985	0.969
	(0.954, 1.000)	(0.896, 0.991)
	20.027	19.699
$TDI_{95\%}$	(16.527, 24.164)	(16.199, 23.970)
	0.858	0.864
CP_{15°	(0.775, 0.925)	(0.767, 0.925)

Table 5 shows posterior summaries for the six measures analyzed in this work for the Jeffreys' prior and the frequentist estimates by using the methods described by Lin et al. (2002). The TDI was estimated picking a fixed CP equal to 0.95 and the last block in Table 5 shows the CP with respect to a difference between T and U of at most 15°. This difference represents a clinically meaningful threshold (Liao, 2015). The Bayesian estimates of CCC are close to the estimate of CCC computed by the method of Lin (1989). Furthermore, our posterior estimate of $TDI_{95\%}$ is

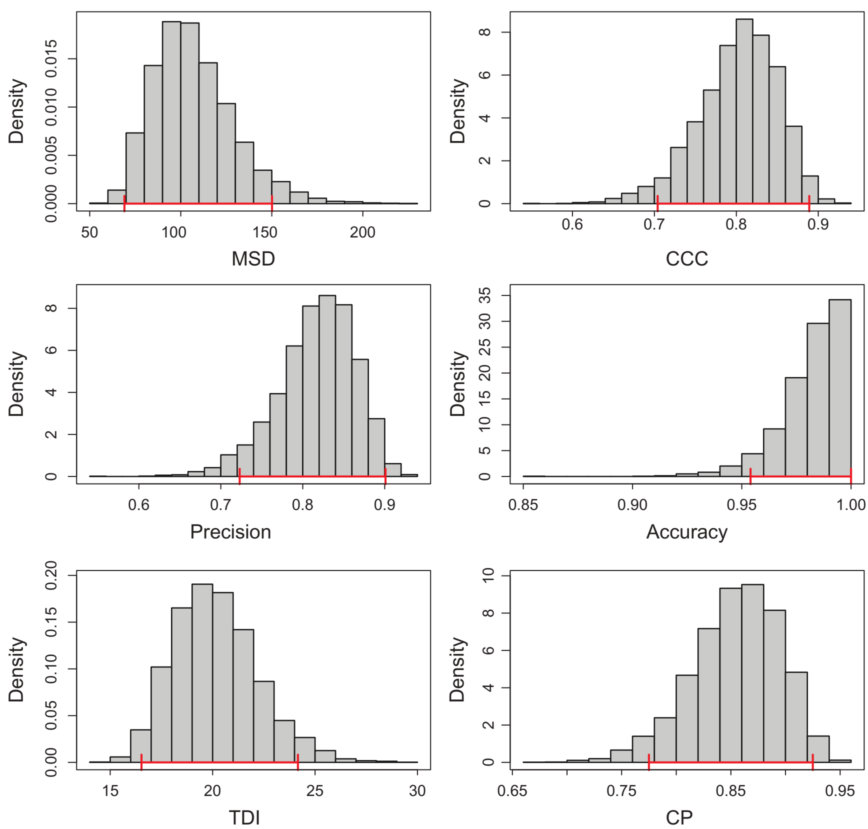

Figure 2. Histograms and 95% HPD intervals (in red) from 5000 samples from the posterior distribution for some measures under the Jeffreys' prior.

similar to that one given by Liao (2015), which is 19.699. The CP for an extreme difference of 15° between tomography and urography measures (CP_{15°) also shows a similar result when compared to that one obtained by Liao (2015). All the HPD intervals in Table 5 are slightly shorter than the asymptotic ones.

In Figure 2, we display the histograms and the 95% HPD intervals for the measures in Table 5 under the Jeffreys' prior. This figure conveys a pictorial description of the posterior distribution. We can see, for example, how strongly skewed to the left is the posterior distribution of the accuracy.

Finally, we can say that both measurement methods, tomography and urography, show a very good agreement. This claim is supported by the fact that the estimates of the CCC and the coverage probability CP_{15° are around 0.81 and 0.85, respectively.

6. Conclusion

The approach presented in this work is a simple way to draw Bayesian inferences for agreement measures under a bivariate normal model. Although simple, our proposal yields estimators with good performance, as reported in the simulation study in Section 4. We stress that asymptotic results, approximations (see, e. g., Lin et al., 2002, Section 2.4), and MCMC methods are not required at all. Hence, our method does not demand a particularly intricate construction of proposal distributions. First, we sample from the parameters of the bivariate normal distribution and through a reparameterization, samples from the posterior distribution for different measures are easily obtained. It is possible to use a great variety of prior distributions with well-studied properties, as presented by Berger & Sun (2008) and Berger et al. (2015). Moreover, application to other measures, like those ones in Carstensen (2010, Ch. 10) and Liao (2015), is straightforward and was not presented here for the sake of space.

Our proposal is based on the normal distribution, and a device for model checking is described in Section 3.3. The elliptical family of distributions includes the normal distribution and could be envisioned as an extension to the present work. Objective Bayesian inference on elliptical distributions is investigated by Fang & Li (1999). Multiple measurements were not considered in this work, but they could be approached based on objective priors for the multivariate normal distribution (Berger et al., 2015). In particular, the approach could be applied to the generalized CCC in Feng et al. (2015). We do not deal with categorical outcomes. However, the methodology could be applied, for example, to the well-known kappa coefficient by assuming that the counts in a contingency table follow a multinomial distribution. Objective priors for the multinomial model are presented in Berger et al. (2015).

Acknowledgments

We would like to thank an Associate Editor/Editorial Board and three referees for their comments, which have contributed to improve the article.

Funding

This work was partially supported by FONDECYT-1130375, Chile and CNPq, Brazil.

References

Barnhart, H. X., Haber, M. J., Lin, L. I. (2007a). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics* 17:529–569.

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

Berger, J. O., Sun, D. (2008). Objective priors for the bivariate normal model. *The Annals of Statistics* 36:963–982.

Berger, J. O., Bernardo, J. M., Sun, D. (2015). Overall objective priors. *Bayesian Analysis* 10:189–221.
Broemeling, L. D. (2009). *Bayesian Methods for Measures of Agreement*. Boca Raton, FL: Chapman & Hall/CRC.
Carstensen, B. (2010). *Comparing Clinical Measurement Methods: A Practical Guide*. Chichester, UK: Wiley.
Chen, M.-H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York, NY: Springer.
Choudhary, P. K., & Nagaraja, H. N. (2007). Tests for assessment of agreement using probability criteria. *Journal of Statistical Planning and Inference* 137:279–290.
Choudhary, P. K., & Yin, K. (2010). Bayesian and frequentist methodologies for analyzing method comparison studies with multiple methods. *Statistics in Biopharmaceutical Research* 2:122–132.
Dunn, G. (2004). *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies*, 2nd ed. London, UK: Arnold.
Fang, K.-T., & Li, R. (1999). Bayesian statistical inference on elliptical matrix distributions. *Journal of Multivariate Analysis* 70:66–85.
Feng, D., Baumgartner, R., & Svetnik, V. (2015). A robust Bayesian estimate of the concordance correlation coefficient. *Journal of Biopharmaceutical Statistics* 25:490–507.
Johnson, V. E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis* 2:719–734.
Liao, J. J. Z. (2015). Quantifying an agreement study. *International Journal of Biostatistics* 11:125–133.
Lin, L., Hedayat, A. S., Sinha, B., & 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., & Wu, W. (2012). *Statistical Tools for Measuring Agreement*. New York, NY: Springer.
Lin, L., Pan, Y., Hedayat, A. S., Barnhart, H. X., & Haber, M. (2015). A simulation study of nonparametric total deviation index as a measure of agreement based on quantile regression. *Journal of Biopharmaceutical Statistics* doi: 10.1080/10543406.2015.1094812.
Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268.
Lin, L. I. (2000). Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Statistics in Medicine* 19:255–270.
Luiz, R. R., Costa, A. J. L., Kale, P. L., & Werneck, G. L. (2003). Assessment of agreement of a quantitative variable: a new graphical approach. *Journal of Clinical Epidemiology* 56:963–967.
Perez-Jaume, S., & Carrasco, J. L. (2015). A non-parametric approach to estimate the total deviation index for non-normal data. *Statistics in Medicine* 34:3318–3335.
R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Schluter, P. J. (2009). A multivariate hierarchical Bayesian approach to measuring agreement in repeated measurement method comparison studies. *BMC Medical Research Methodology* 9(6). doi:10.1186/1471-2288-9-6
Yin, K., Choudhary, P. K., Varghese, D., & Goodman, S. T. (2008). A Bayesian approach for sample size determination in method comparison studies. *Statistics in Medicine* 27:2273–2289.

Appendix

A. Sampling under the right-haar prior

For the right-Haar prior, samples from the posterior distribution of the parameters σ_X, σ_Y, ρ, μ_X, and μ_Y are drawn by cycling the steps 1–6 below:

(1) Draw independent $Z_1, Z_2, Z_3 \sim N(0,1)$, $C_1 \sim \chi^2_{n-1}$, and $C_2 \sim \chi^2_{n-2}$,
(2) Compute $\sigma_X = (s_{11}/C_1)^{1/2}$,
(3) Compute

$$\sigma_Y = [s_{22}(1-r^2)]^{1/2}\left[\frac{1}{C_2} + \frac{1}{C_1}\left(\frac{Z_3}{C_2^{1/2}} - \frac{r}{(1-r^2)^{1/2}}\right)^2\right]^{1/2},$$

(4) Compute $\rho = \psi(W)$, where

$$W = -\frac{Z_3}{C_1^{1/2}} + \frac{C_2^{1/2}}{C_1^{1/2}}\frac{r}{(1-r^2)^{1/2}} \quad \text{and} \quad \psi(v) = \frac{v}{(1+v^2)^{1/2}},$$

(5) Compute $\mu_X = \bar{x} + Z_1 s_{11}^{1/2}/(nC_1)^{1/2}$, and
(6) Compute

$$\mu_Y = \bar{y} + \frac{Z_1 r s_{22}^{1/2}}{(nC_1)^{1/2}} + \left(\frac{Z_2}{C_2^{1/2}} - \frac{Z_1 Z_3}{(C_1 C_2)^{1/2}}\right)\left(\frac{s_{22}(1-r^2)}{n}\right)^{1/2}.$$

B. Supplemental web materials

In this web appendix, we present tables summarizing some results from our simulation study in Section 4 of the main text. The tables in this web appendix show the estimates of the CP of the 95% HPD intervals and their corresponding average lengths for MSD, CCC, precision, accuracy, TDI, and CP under the independence Jeffreys' prior, the right-Haar prior, and the reference priors for ρ and σ.

Table B1. Independence Jeffreys' prior.

Comb.	n	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.957 (2.378)	0.946 (0.017)	0.927 (0.015)	–	0.945 (1.280)	0.932 (0.185)
	100	0.949 (1.153)	0.953 (0.008)	0.943 (0.008)	–	0.949 (0.656)	0.947 (0.095)
2	30	0.956 (24.681)	0.945 (0.152)	0.925 (0.142)	–	0.951 (4.014)	0.944 (0.183)
	100	0.953 (11.718)	0.954 (0.077)	0.943 (0.075)	–	0.953 (2.074)	0.948 (0.096)
3	30	0.970 (74.178)	0.930 (0.374)	0.924 (0.367)	–	0.956 (6.955)	0.948 (0.183)
	100	0.954 (35.189)	0.952 (0.201)	0.949 (0.200)	–	0.949 (3.592)	0.947 (0.096)
4	30	0.956 (124.355)	0.947 (0.518)	0.942 (0.530)	–	0.947 (9.004)	0.939 (0.184)
	100	0.952 (58.459)	0.933 (0.290)	0.935 (0.291)	–	0.943 (4.632)	0.941 (0.095)
5	30	0.956 (4.513)	0.954 (0.021)	0.923 (0.015)	0.954 (0.013)	0.953 (1.713)	0.944 (0.183)
	100	0.944 (2.124)	0.955 (0.010)	0.950 (0.008)	0.944 (0.006)	0.943 (0.883)	0.936 (0.095)
6	30	0.961 (29.736)	0.949 (0.156)	0.933 (0.142)	0.966 (0.048)	0.954 (4.405)	0.945 (0.185)
	100	0.966 (13.924)	0.961 (0.079)	0.958 (0.076)	0.961 (0.020)	0.960 (2.261)	0.957 (0.096)
7	30	0.957 (86.056)	0.940 (0.382)	0.916 (0.375)	0.955 (0.108)	0.944 (7.489)	0.936 (0.185)
	100	0.951 (40.528)	0.941 (0.202)	0.946 (0.200)	0.969 (0.040)	0.948 (3.857)	0.944 (0.096)
8	30	0.957 (142.297)	0.944 (0.521)	0.934 (0.534)	0.968 (0.145)	0.951 (9.634)	0.940 (0.185)
	100	0.955 (65.914)	0.951 (0.288)	0.946 (0.291)	0.964 (0.055)	0.949 (4.918)	0.951 (0.095)
9	30	0.947 (10.853)	0.932 (0.117)	0.932 (0.015)	0.929 (0.109)	0.951 (1.451)	0.941 (0.167)
	100	0.948 (5.718)	0.937 (0.063)	0.955 (0.008)	0.932 (0.059)	0.954 (0.759)	0.950 (0.092)
10	30	0.956 (40.701)	0.949 (0.222)	0.942 (0.141)	0.947 (0.156)	0.953 (4.533)	0.940 (0.168)
	100	0.946 (21.078)	0.952 (0.120)	0.947 (0.075)	0.941 (0.086)	0.954 (2.392)	0.951 (0.092)
11	30	0.956 (93.536)	0.955 (0.390)	0.936 (0.370)	0.950 (0.220)	0.948 (7.678)	0.949 (0.177)
	100	0.948 (46.464)	0.944 (0.214)	0.947 (0.200)	0.921 (0.126)	0.947 (4.007)	0.943 (0.094)
12	30	0.954 (143.813)	0.945 (0.498)	0.940 (0.530)	0.954 (0.257)	0.947 (9.731)	0.937 (0.178)
	100	0.951 (69.927)	0.945 (0.278)	0.949 (0.292)	0.937 (0.153)	0.947 (5.029)	0.944 (0.094)
13	30	0.951 (14.803)	0.941 (0.112)	0.935 (0.015)	0.940 (0.105)	0.949 (1.942)	0.940 (0.166)
	100	0.949 (7.786)	0.947 (0.061)	0.946 (0.008)	0.944 (0.057)	0.943 (1.019)	0.937 (0.091)
14	30	0.962 (45.646)	0.954 (0.218)	0.933 (0.141)	0.954 (0.152)	0.956 (4.925)	0.950 (0.169)
	100	0.952 (23.816)	0.949 (0.119)	0.946 (0.076)	0.948 (0.084)	0.952 (2.613)	0.953 (0.093)
15	30	0.956 (103.388)	0.939 (0.389)	0.938 (0.371)	0.950 (0.217)	0.955 (8.121)	0.949 (0.176)
	100	0.957 (51.707)	0.951 (0.214)	0.953 (0.201)	0.947 (0.124)	0.952 (4.262)	0.948 (0.095)
16	30	0.967 (158.732)	0.937 (0.501)	0.923 (0.532)	0.960 (0.249)	0.959 (10.282)	0.948 (0.177)
	100	0.957 (78.643)	0.948 (0.279)	0.948 (0.293)	0.927 (0.151)	0.955 (5.349)	0.952 (0.095)

Coverage probability (and average length) of the 95% HPD interval for some measures based on 1000 replications and 5000 samples from the posterior distribution.

Table B2. Right-Haar prior.

Comb.	n	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.956 (2.629)	0.954 (0.018)	0.951 (0.016)	–	0.949 (1.320)	0.928 (0.191)
	100	0.949 (1.187)	0.955 (0.008)	0.951 (0.008)	–	0.948 (0.662)	0.944 (0.097)
2	30	0.966 (25.703)	0.942 (0.160)	0.933 (0.149)	–	0.956 (4.125)	0.952 (0.188)
	100	0.948 (11.868)	0.955 (0.078)	0.952 (0.076)	–	0.944 (2.091)	0.944 (0.097)
3	30	0.965 (76.322)	0.935 (0.382)	0.934 (0.375)	–	0.961 (7.080)	0.945 (0.187)
	100	0.955 (35.513)	0.953 (0.203)	0.952 (0.201)	–	0.949 (3.616)	0.947 (0.096)
4	30	0.958 (126.617)	0.947 (0.519)	0.953 (0.532)	–	0.949 (9.103)	0.939 (0.186)
	100	0.948 (58.759)	0.935 (0.290)	0.935 (0.292)	–	0.942 (4.647)	0.940 (0.096)
5	30	0.958 (4.570)	0.952 (0.023)	0.941 (0.017)	0.958 (0.013)	0.956 (1.726)	0.944 (0.185)
	100	0.944 (2.132)	0.955 (0.010)	0.954 (0.008)	0.943 (0.006)	0.943 (0.885)	0.939 (0.096)
6	30	0.959 (31.108)	0.947 (0.164)	0.943 (0.150)	0.964 (0.051)	0.954 (4.540)	0.942 (0.191)
	100	0.962 (14.144)	0.960 (0.080)	0.965 (0.077)	0.955 (0.021)	0.959 (2.286)	0.957 (0.097)
7	30	0.947 (89.388)	0.934 (0.390)	0.937 (0.384)	0.951 (0.113)	0.952 (7.673)	0.935 (0.190)
	100	0.951 (41.032)	0.940 (0.203)	0.947 (0.201)	0.966 (0.041)	0.941 (3.888)	0.939 (0.097)
8	30	0.956 (145.750)	0.942 (0.521)	0.942 (0.535)	0.952 (0.151)	0.950 (9.778)	0.938 (0.189)
	100	0.962 (66.505)	0.950 (0.289)	0.950 (0.291)	0.959 (0.057)	0.957 (4.946)	0.957 (0.096)
9	30	0.953 (11.061)	0.940 (0.119)	0.952 (0.016)	0.939 (0.110)	0.956 (1.486)	0.951 (0.171)
	100	0.949 (5.751)	0.939 (0.063)	0.960 (0.008)	0.933 (0.059)	0.955 (0.765)	0.951 (0.093)
10	30	0.962 (41.761)	0.954 (0.227)	0.952 (0.148)	0.949 (0.160)	0.961 (4.628)	0.947 (0.172)
	100	0.949 (21.212)	0.946 (0.121)	0.951 (0.076)	0.941 (0.086)	0.952 (2.405)	0.949 (0.093)
11	30	0.959 (95.716)	0.962 (0.395)	0.943 (0.379)	0.954 (0.223)	0.958 (7.799)	0.947 (0.180)
	100	0.948 (46.764)	0.943 (0.215)	0.952 (0.201)	0.926 (0.127)	0.947 (4.024)	0.943 (0.095)
12	30	0.957 (145.993)	0.951 (0.497)	0.946 (0.532)	0.957 (0.259)	0.949 (9.823)	0.943 (0.180)
	100	0.946 (70.270)	0.945 (0.278)	0.953 (0.292)	0.939 (0.153)	0.945 (5.044)	0.943 (0.094)
13	30	0.955 (14.907)	0.947 (0.114)	0.946 (0.016)	0.942 (0.106)	0.953 (1.954)	0.943 (0.167)
	100	0.945 (7.805)	0.947 (0.061)	0.953 (0.008)	0.940 (0.057)	0.942 (1.021)	0.937 (0.091)
14	30	0.963 (46.979)	0.957 (0.223)	0.950 (0.149)	0.955 (0.156)	0.958 (5.041)	0.953 (0.173)
	100	0.956 (23.998)	0.953 (0.120)	0.950 (0.077)	0.949 (0.085)	0.952 (2.631)	0.951 (0.094)
15	30	0.960 (106.639)	0.939 (0.393)	0.947 (0.379)	0.947 (0.221)	0.950 (8.292)	0.944 (0.180)
	100	0.954 (52.142)	0.952 (0.214)	0.952 (0.202)	0.949 (0.125)	0.949 (4.284)	0.952 (0.095)
16	30	0.965 (162.224)	0.937 (0.499)	0.930 (0.534)	0.957 (0.253)	0.958 (10.422)	0.951 (0.181)
	100	0.956 (79.197)	0.950 (0.279)	0.950 (0.293)	0.929 (0.152)	0.956 (5.371)	0.949 (0.096)

Coverage probability (and average length) of the 95% HPD interval for some measures based on 1000 replications and 5000 samples from the posterior distribution.

Table B3. Reference prior for ρ.

Comb.	n	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.952 (2.624)	0.953 (0.018)	0.952 (0.016)	–	0.942 (1.318)	0.929 (0.191)
	100	0.955 (1.187)	0.958 (0.008)	0.949 (0.008)	–	0.953 (0.662)	0.948 (0.097)
2	30	0.963 (25.666)	0.940 (0.160)	0.937 (0.150)	–	0.957 (4.120)	0.948 (0.188)
	100	0.951 (11.870)	0.958 (0.078)	0.950 (0.076)	–	0.944 (2.091)	0.941 (0.097)
3	30	0.967 (76.237)	0.934 (0.381)	0.937 (0.375)	–	0.962 (7.080)	0.947 (0.187)
	100	0.957 (35.498)	0.955 (0.203)	0.954 (0.201)	–	0.951 (3.615)	0.951 (0.096)
4	30	0.954 (126.185)	0.945 (0.519)	0.948 (0.532)	–	0.947 (9.074)	0.941 (0.186)
	100	0.948 (58.683)	0.930 (0.290)	0.935 (0.292)	–	0.944 (4.642)	0.942 (0.096)
5	30	0.959 (4.533)	0.949 (0.022)	0.943 (0.017)	0.956 (0.013)	0.954 (1.716)	0.944 (0.184)
	100	0.947 (2.124)	0.952 (0.010)	0.956 (0.008)	0.944 (0.006)	0.944 (0.882)	0.940 (0.095)
6	30	0.961 (30.720)	0.945 (0.163)	0.938 (0.150)	0.966 (0.049)	0.952 (4.500)	0.946 (0.189)
	100	0.963 (14.0878)	0.961 (0.080)	0.959 (0.077)	0.961 (0.020)	0.960 (2.278)	0.960 (0.097)
7	30	0.953 (88.191)	0.939 (0.389)	0.934 (0.383)	0.957 (0.109)	0.955 (7.601)	0.938 (0.188)
	100	0.949 (40.847)	0.939 (0.203)	0.942 (0.201)	0.967 (0.040)	0.946 (3.877)	0.942 (0.097)
8	30	0.953 (144.256)	0.942 (0.522)	0.941 (0.535)	0.971 (0.147)	0.953 (9.713)	0.940 (0.187)
	100	0.958 (66.286)	0.949 (0.288)	0.947 (0.291)	0.962 (0.056)	0.952 (4.932)	0.951 (0.096)
9	30	0.950 (11.054)	0.938 (0.119)	0.947 (0.016)	0.936 (0.110)	0.957 (1.485)	0.945 (0.171)
	100	0.949 (5.749)	0.938 (0.063)	0.957 (0.008)	0.937 (0.059)	0.953 (0.765)	0.949 (0.093)
10	30	0.958 (41.757)	0.952 (0.227)	0.955 (0.148)	0.949 (0.160)	0.955 (4.630)	0.945 (0.172)
	100	0.950 (21.211)	0.952 (0.121)	0.952 (0.076)	0.944 (0.086)	0.952 (2.406)	0.948 (0.093)
11	30	0.960 (95.645)	0.958 (0.394)	0.944 (0.378)	0.954 (0.223)	0.952 (7.792)	0.945 (0.180)
	100	0.949 (46.767)	0.942 (0.214)	0.954 (0.201)	0.923 (0.127)	0.948 (4.024)	0.942 (0.095)
12	30	0.955 (145.555)	0.949 (0.497)	0.945 (0.532)	0.956 (0.259)	0.953 (9.804)	0.942 (0.180)
	100	0.951 (70.219)	0.946 (0.278)	0.952 (0.292)	0.943 (0.153)	0.948 (5.042)	0.946 (0.094)
13	30	0.954 (14.866)	0.939 (0.114)	0.950 (0.016)	0.944 (0.106)	0.946 (1.946)	0.938 (0.167)
	100	0.946 (7.799)	0.941 (0.061)	0.951 (0.008)	0.947 (0.057)	0.947 (1.020)	0.936 (0.091)
14	30	0.963 (46.609)	0.955 (0.223)	0.948 (0.149)	0.956 (0.155)	0.957 (5.007)	0.953 (0.172)
	100	0.955 (23.962)	0.951 (0.120)	0.951 (0.077)	0.950 (0.085)	0.952 (2.626)	0.954 (0.094)
15	30	0.960 (105.427)	0.940 (0.393)	0.949 (0.379)	0.950 (0.220)	0.953 (8.222)	0.948 (0.179)
	100	0.953 (52.003)	0.953 (0.214)	0.952 (0.202)	0.953 (0.124)	0.955 (4.277)	0.949 (0.095)
16	30	0.966 (160.675)	0.938 (0.500)	0.935 (0.534)	0.960 (0.251)	0.957 (10.351)	0.949 (0.179)
	100	0.956 (78.895)	0.950 (0.279)	0.945 (0.293)	0.927 (0.152)	0.955 (5.356)	0.951 (0.095)

Coverage probability (and average length) of the 95% HPD interval for some measures based on 1000 replications and 5000 samples from the posterior distribution.

Table B4. Reference prior for σ.

Comb.	*n*	MSD	CCC	Precision	Accuracy	TDI	CP
1	30	0.951 (2.623)	0.951 (0.018)	0.949 (0.016)	–	0.949 (1.317)	0.930 (0.191)
	100	0.954 (l.187)	0.957 (0.008)	0.950 (0.008)	–	0.949 (0.662)	0.946 (0.097)
2	30	0.960 (25.467)	0.942 (0.158)	0.934 (0.148)	–	0.953 (4.096)	0.949 (0.187)
	100	0.947 (11.845)	0.959 (0.078)	0.948 (0.076)	–	0.946 (2.088)	0.941 (0.096)
3	30	0.970 (75.258)	0.936 (0.377)	0.935 (0.370)	–	0.960 (7.012)	0.948 (0.185)
	100	0.954 (35.359)	0.950 (0.202)	0.955 (0.200)	–	0.950 (3.604)	0.951 (0.096)
4	30	0.956 (125.017)	0.948 (0.516)	0.950 (0.528)	–	0.950 (9.023)	0.940 (0.185)
	100	0.950 (58.503)	0.934 (0.289)	0.933 (0.291)	–	0.949 (4.632)	0.944 (0.095)
5	30	0.958 (4.538)	0.952 (0.022)	0.940 (0.017)	0.956 (0.013)	0.954 (1.718)	0.942 (0.184)
	100	0.944 (2.127)	0.958 (0.010)	0.956 (0.008)	0.945 (0.006)	0.942 (0.883)	0.934 (0.096)
6	30	0.959 (30.543)	0.947 (0.161)	0.941 (0.148)	0.971 (0.050)	0.957 (4.482)	0.944 (0.189)
	100	0.966 (14.065)	0.962 (0.080)	0.963 (0.076)	0.959 (0.020)	0.962 (2.275)	0.956 (0.097)
7	30	0.956 (87.123)	0.938 (0.384)	0.928 (0.378)	0.962 (0.109)	0.948 (7.534)	0.935 (0.187)
	100	0.952 (40.687)	0.941 (0.202)	0.946 (0.200)	0.969 (0.040)	0.947 (3.867)	0.940 (0.097)
8	30	0.959 (142.977)	0.941 (0.520)	0.940 (0.532)	0.971 (0.146)	0.952 (9.654)	0.942 (0.186)
	100	0.959 (66.033)	0.950 (0.288)	0.951 (0.291)	0.965 (0.056)	0.950 (4.924)	0.953 (0.095)
9	30	0.952 (11.053)	0.940 (0.119)	0.946 (0.016)	0.939 (0.110)	0.957 (1.484)	0.950 (0.171)
	100	0.952 (5.749)	0.941 (0.063)	0.960 (0.008)	0.940 (0.059)	0.957 (0.765)	0.953 (0.093)
10	30	0.962 (41.568)	0.953 (0.226)	0.953 (0.147)	0.942 (0.159)	0.959 (4.610)	0.946 (0.172)
	100	0.949 (21.196)	0.948 (0.121)	0.948 (0.076)	0.943 (0.086)	0.954 (2.404)	0.944 (0.093)
11	30	0.955 (94.680)	0.957 (0.391)	0.944 (0.373)	0.953 (0.222)	0.953 (7.734)	0.947 (0.178)
	100	0.947 (46.598)	0.945 (0.214)	0.952 (0.200)	0.922 (0.126)	0.947 (4.012)	0.942 (0.094)
12	30	0.958 (144.451)	0.945 (0.496)	0.939 (0.529)	0.959 (0.258)	0.950 (9.753)	0.939 (0.179)
	100	0.949 (70.016)	0.947 (0.278)	0.951 (0.291)	0.939 (0.153)	0.950 (5.033)	0.944 (0.094)
13	30	0.955 (14.861)	0.940 (0.114)	0.951 (0.016)	0.943 (0.106)	0.959 (1.948)	0.945 (0.167)
	100	0.947 (7.795)	0.942 (0.061)	0.951 (0.008)	0.941 (0.057)	0.947 (1.019)	0.936 (0.091)
14	30	0.959 (46.434)	0.956 (0.222)	0.945 (0.147)	0.953 (0.155)	0.957 (4.992)	0.951 (0.171)
	100	0.954 (23.928)	0.954 (0.120)	0.946 (0.077)	0.946 (0.085)	0.952 (2.623)	0.952 (0.094)
15	30	0.956 (104.659)	0.942 (0.390)	0.947 (0.374)	0.952 (0.219)	0.954 (8.182)	0.949 (0.178)
	100	0.950 (51.896)	0.954 (0.214)	0.955 (0.201)	0.952 (0.124)	0.949 (4.271)	0.944 (0.095)
16	30	0.965 (159.315)	0.940 (0.498)	0.922 (0.530)	0.958 (0.250)	0.956 (10.294)	0.946 (0.178)
	100	0.957 (78.690)	0.948 (0.278)	0.947 (0.292)	0.930 (0.151)	0.953 (5.347)	0.952 (0.095)

Coverage probability (and average length) of the 95% HPD interval for some measures based on 1000 replications and 5000 samples from the posterior distribution.