2009

Confidence interval estimation for a difference between two correlated intraclass correlation coefficients with variable class sizes

Danuta Kowalik

Confidence interval estimation for a difference between two correlated intraclass correlation coefficients with variable class sizes

(Spine Title: CI for a difference between two correlated ICCs)

(Thesis Format: Monograph)

by

Danuta Kowalik

Graduate Program in Epidemiology & Biostatistics

Submitted in partial fulfillment
of the requirements for the degree of
Master of Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario
July 2009

ABSTRACT

The intraclass correlation coefficient (ICC) has application in several fields of research. Various confidence interval methods for a single ICC are available. However, statistical inference for multiple ICCs has primarily relied on hypothesis testing which cannot distinguish between statistical significance and practical importance. The focus of this thesis is to develop and evaluate confidence interval procedures for a difference between two correlated ICCs with variable class sizes. The strategy used in the thesis is to recover variance estimates needed for the confidence interval for the difference from the confidence limits for single ICCs. Simulation results show that the procedure based on inverse hyperbolic tangent transformation for single ICCs performs well. The Galton's 1886 dataset on siblings heights is used to illustrate the methodology.

Key Words: Intraclass correlation coefficients, confidence limits, random effects model, statistical significance, variance.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Guang Yong Zou for the support, guidance and inspiration throughout the course of my thesis. Without his enthusiasm, great ideas and positive attitude this work would have not been possible. I am also grateful to Dr. Yun-Hee Choi for her helpful comments and suggestions on the thesis.

I owe my loving thanks to my boyfriend Esteban Villacis for the emotional support, love and constant encouragement through the course of this work. Thank you for putting a smile on my face even in the toughest times. Last, but not least, I am indebted to my family for their optimistic and motivating words and endless love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The intraclass correlation coefficient (ICC), an index of resemblance, plays an important role in a wide range of disciplines. For example, in epidemiologic research, it is commonly used to measure the degree of resemblance among family members with respect to biological or environmental attributes. It can also be used as an index of reliability in biomedical and psychological research. Intraclass correlation coefficients also play an important role in the design and analysis of trials in which the units being randomized are in fact social clusters. So called cluster randomization trials, have become widespread in health research.

Comparison of two ICCs is sometimes a focus in biomedical and epidemiologic research. Despite the fact that inferential procedures for the problem of comparing ICCs from two or more independent samples have been developed, limited literature exists on the comparison of two ICCs obtained from the same sample.

Assuming a balanced one-way random effects model, Ramasundarahettige *et al.* (2009) provided a confidence interval for a difference of two correlated ICCs by applying a procedure developed by Zou and Donner (2008). Ramasundarahettige *et al.* (2009) showed by their simulation study that the method based on the F-distribution for single ICCs performs very well in small sample sizes. The general idea is to recover variance estimates needed for setting confidence limits for the difference from confidence limits for single ICCs. The advantage of this method is that it can reflect the underlying sampling distribution, in contrast to the conventional method of point estimate plus/minus a multiplier of critical values.

However, very often researchers encounter unbalanced studies. For instance, in family studies it is rare to have sibships of constant size. In reliability studies, variable number of observations per subject could arise due to missing data. Therefore, confidence interval procedures for the unbalanced case are needed. The purpose of this project is to apply the procedure proposed by Zou and Donner (2008) to a difference between two correlated ICCs with variable class sizes. Monte Carlo simulation is used to assess the performance of the proposed methods.

The rest of this chapter is organized as follows. The concept of ICC is introduced by providing its definition and interpretation in Section 1.1. An overview of statistical inference for ICC is provided in Section 1.2. A motivating example is given in Section 1.3. The chapter ends with the objective of the thesis and its organization presented in Sections 1.4 and Section 1.5, respectively.

1.1 Intraclass correlation coefficient

Suppose that observations on a single variable Y are arranged in k classes and the following one-way random effects model is assumed

$$Y_{ij} = \mu + a_i + e_{ij}, \tag{1.1}$$

where Y_{ij} represents an observation on the j^{th} member of the i^{th} class, $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, k$. The overall mean is denoted as μ, the class effects a_i are normally distributed with mean 0 and variance σ_A^2, the residual errors e_{ij} are normally distributed with mean 0 and variance σ_e^2. In addition, a_i and e_{ij} are independent. The variance of Y_{ij} is given by $\sigma^2(Y_{ij}) = \sigma_Y^2 = \sigma_A^2 + \sigma_e^2$. Model (1.1) sometimes is referred as "components of variance model" (Donner, 1986) since the total variance, σ_Y^2, is a sum of two variance components, σ_e^2 and σ_A^2. Under model (1.1), the covariance of observations in the same class, Y_{ij}, $Y_{ij'}$ and $j \neq j'$, is

$$\begin{aligned}
\mathrm{cov}(Y_{ij}, Y_{ij'}) &= \mathrm{E}[(Y_{ij} - \mathrm{E}(Y_{ij}))(Y_{ij'} - \mathrm{E}(Y_{ij'}))] \\
&= \mathrm{E}[(a_i + e_{ij})(a_i + e_{ij'})] \\
&= \mathrm{E}[a_i^2] \\
&= \sigma_A^2
\end{aligned}$$

and the covariance of observations from different classes is

$$\mathrm{cov}(Y_{ij}, Y_{i'j'}) = 0, \ i \neq i'.$$

Thus, the correlation between any two observations Y_{ij} and $Y_{ij'}$ is given by

$$\rho(Y_{ij}, Y_{ij'}) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2} \tag{1.2}$$

since by definition of correlation, ρ is given by

$$\rho = \mathrm{cor}(Y_{ij}, Y_{ij'}) = \frac{\mathrm{cov}(Y_{ij}, Y_{ij'})}{\sqrt{\mathrm{var}(Y_{ij})\mathrm{var}(Y_{ij'})}} = \frac{\sigma_A^2}{\sqrt{\sigma_Y^2 \sigma_Y^2}} = \frac{\sigma_A^2}{\sigma_Y^2} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}.$$

The intraclass correlation coefficient, ρ, can be interpreted as the proportion of the total variability that is attributed to the variability among classes.

Another commonly used model is a common correlation model under which ρ is defined explicitly. Specifically, observations Y_{ij} are assumed to be distributed with the same mean μ and with the same variance σ_Y^2. Any two members within a class have a common correlation but and any two members from different classes are uncorrelated. This latter model is more general than model (1.1) since it also allows values of ρ to be negative. When in fact, ρ is nonnegative, the two models become equivalent.

In this thesis, ρ is assumed to be positive with possible values range from 0 to 1. The lowest possible value of $\rho = 0$, which implies independence of observations among individuals within a class. On the other hand, values of $\rho > 0$ indicate that individuals within a class are more similar with respect to the attribute Y than to individuals from a different class. If the value of ρ is near zero, differences in mean observations between classes are relatively small as compared to the differences in observations within each class. On the contrary, if ρ is close to 1, then much of the total variability is attributed to differences between classes.

1.2 Statistical inference on intraclass correlation coefficient

Statistical inference for a single ICC have been well established and applied to various areas of research. There exist several methods to estimate ρ. The most commonly used in practice (Donner and Eliasziw, 1991) include: the ANOVA estimator; the pairwise estimator; and the maximum likelihood estimator (MLE). Although, the latter is efficient for all values of ρ, it cannot be obtained explicitly, but requires iterative procedures at least when class sizes are variable. Several other estimators of ρ have been proposed including variations of the ANOVA estimator and the pairwise estimator implementing different weighting schemes, others rely on algebraic combinations of two estimators with an appropriate choice of weights. Further discussion of point estimation is provided in the next chapter.

While statistical inference for ρ frequently focuses on point estimation and hypothesis testing, confidence interval estimation is a more informative approach. There are several methods for constructing confidence intervals for ρ. The most intuitive approach is to implement a simply asymptotic method of point estimate plus/minus critical values times the estimated standard error. Since this approach is not expected to perform well in small sample sizes, as the distribution of $\hat{\rho}$ is skewed except when $\rho = 0$, other methods have been developed. One of them is an approach based on Fisher's Z-transformation (Fisher, 1925, Ch.7), which is used to normalize and stabilize the sampling distributions of $\hat{\rho}$.

A similar transformation to that of Fisher's Z-transformation is inverse hyperbolic tangent transformation. Lachin (2004) suggested that this transformation with its variance obtained by the Delta method may perform better than that of Fisher's Z-transformation. Confidence intervals are first obtained on the transformed scale and then transformed back to obtain the limits on the original scale.

Another method also known to researchers is based on the F-distribution. Since under variable class sizes, the distribution of variance-ratio statistics is not exact, Thomas and Hultquist (1978) and Donner (1979) approximated the F-distribution by modifying the variance-ratio statistic using a common mean class size in place of a constant class size. Thomas and Hultquist (1978) made use of harmonic mean and unweighted sum of squares of class means in the variance-ratio statistics to approximate the F-distribution.

There also exist a vast literature concerning ICCs in multiple samples, especially in the case of independent samples. A further discussion of inferential procedures for ICCs in multiple samples will be given in the the next chapter.

1.3 Motivating example

An example which motivates this thesis is based on the well known Galton's 1886 data set on human stature. These data consist of records of heights for 205 families including father, mother, sons, and daughters. For further details about these data, we refer to Hanley (2004a).

These data played a profound role in statistical research. First, Galton used these data to calculate the correlation coefficient of the bivariate Gaussian distribution. Second, Pearson used these data to quantify multiple and partial correlations. Recently, these data were analyzed by Naik and Helu (2007), as an illustration for testing the equality of two independent ICCs. For that purpose, the data were divided into two groups, the first one consisting of data for daughters only from the first 102 families, the second one consisting of data for sons from the remaining 103 families. Then two sex-specific ICCs were compared by means of hypothesis testing. In this thesis, these data are used to compare the degree of resemblance with respect to height of brothers to that of sisters. In comparing the two sex-specific ICCs, the proposed confidence interval approach is implemented taking into account the dependency among siblings and also the distributional properties of ICCs.

1.4 The objective of the thesis

The objective of this thesis is to construct confidence interval for a difference of two correlated ICCs. The strategy is to apply the procedure proposed by Zou and Donner (2008), given that confidence intervals for a single ICC are well established. The proposed method has been referred to as the MOVER, method of variance estimates recovery (Zou, 2008), because the idea is to recover variance estimates needed for the confidence interval for a difference between two correlated ICCs from those for the separate ICCs. Thus, the proposed method takes its validity from those limits obtained for a single ICC. The MOVER takes into account the asymmetric sampling distribution of ICC estimators in contrast to the standard simple asymptotic method, which lacks this ability. The performance of the proposed method is evaluated by a Monte Carlo simulation with respect to the empirical coverage, balance of tail errors, and confidence interval width.

1.5 Organization of the thesis

The thesis is composed of six chapters. Chapter 2 provides a review of the literature on inferential procedures for ICC. Construction of intervals for a difference of two ICCs are in Chapter 3. Chapter 4 presents the simulation study. In Chapter 5, Galton's data are used to illustrate the proposed method. The thesis ends with conclusions.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

The intraclass correlation coefficient has a long standing history of applications in diverse disciplines including epidemiologic, psychological, and biomedical research. Due to its wide spectrum of applications across different disciplines, an extensive literature on its inference is available, with a review focusing on inference based on a one-way random effects model provided by Donner (1986).

The purpose of this chapter is to provide a review of statistical inference for ICC with an emphasis on recent procedures. Specifically, Section 2.2 presents a summary of point estimators and in Section 2.3, various methods for confidence interval estimation for a single ICC are discussed. Section 2.4 summarizes statistical inference in multiple samples.

2.2 Point estimators

The concept of intraclass correlation coefficient, ρ, can be traced back to Pearson (1901), where ρ was estimated by calculating product-moment correlation over all possible pairs of observations within a class. In the calculation process each pair of observations within a class is counted twice. This estimator is known as the pairwise estimator of ρ and is given by

$$\hat{\rho}_p = \frac{\sum_{i=1}^{k} \sum_{\substack{j=1 \\ j'=1 \\ j\neq j'}}^{n_i} (Y_{ij} - \bar{Y}_p)(Y_{ij'} - \bar{Y}_p)}{\sum_{i=i}^{k}(n_i - 1) \sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_p)^2}, \tag{2.1}$$

where Y_{ij} denotes a random variable representing a measurement on the j^{th} member of the i^{th} class, n_i is the size of the i^{th} class, k is the number of classes, $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$, and $\bar{Y}_p = \sum_{i=1}^{k} n_i(n_i - 1)\bar{Y}_i / \sum_{i=1}^{k} n_i(n_i - 1)$. In the case of variable class sizes, this estimator can be used but it tends to give more weight to classes of larger sizes than to classes of smaller sizes (Fieller and Smith, 1951). For example, a class of size 10 receives 45 times as much weight in the calculation process as class of size 2.

Following Pearson, the most well known estimator of ρ is attributed to Fisher (1925, Ch.7) who placed it in the context of a balanced one-way random effects model. An extension of the ANOVA estimator to an unbalanced one-way random effects model is due to Fieller and Smith (1951). An unbalanced one-way random effects model assumes that observation of the j^{th} member of the i^{th} class ($j = 1, 2, 3 \cdots n_i; i = 1, 2, \cdots, k$), where n_i is the class size, k represents number of classes, can be described as

$$Y_{ij} = \mu + a_i + e_{ij}, \tag{2.2}$$

where μ is the true mean of all observations, a_i are the class effects which are normally distributed with mean 0 and variance σ_A^2, e_{ij} are the residual errors which are normally distributed with mean 0 and variance σ_e^2. The a_i and e_{ij} are independent. The variance of Y_{ij} is given by $\text{var}(Y_{ij}) = \sigma_A^2 + \sigma_e^2$. The analysis of variance corresponding to this model is shown in Table 2.1, where $N = \Sigma_i n_i$ is the total number of observations across classes, $\bar{Y}_{i.} = \Sigma_j Y_{ij}/n_i$ is the mean of the observations in the i^{th} class ($i = 1, \cdots, k$), and $\bar{Y}_{..} = \Sigma_{i,j} Y_{ij}/N$ is the overall mean of all observation over all classes. Then, ANOVA estimator of ρ is given by

$$\hat{\rho}_A = \frac{\text{MSA} - \text{MSE}}{\text{MSA} + (n_0 - 1)\text{MSE}} = \frac{(F-1)}{(F + n_0 - 1)}, \tag{2.3}$$

where MSA and MSE are the respective mean sum of squares between and within classes, and

$$n_0 = (N - \Sigma_{i=1}^k n_i^2/N)/(k-1). \tag{2.4}$$

Although ρ is defined as a non-negative parameter under model (2.2), its ANOVA estimator may be negative. This may happen whenever MSE > MSA.

Table 2.1: Analysis of variance for an unbalanced one-way random effects model

Source of Variation	Degrees of freedom	Sum of squares	Mean square	Expected mean square
Among classes	$k-1$	$\text{SSA} = \sum n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$	MSA = SSA/(k-1)	$\sigma_e^2 + \frac{N-\sum n_i^2/N}{k-1}\sigma_A^2$
Within classes	$N-k$	$\text{SSE} = \sum\sum(Y_{ij} - \bar{Y}_{i.})^2$	MSE = SSE/(N-k)	σ_e^2
Total	$N-1$	$\text{SST} = \sum\sum(Y_{ij} - \bar{Y}_{..})^2$		

Smith (1956) derived the large-sample variance formula of $\hat{\rho}_A$ by first obtaining variances of variance components of $\hat{\rho}_A$ under the normality assumption, and then implementing them in conjunction with the Delta method to find the variance of $\hat{\rho}_A$. The formula as a function of ρ is given as

$$\begin{aligned}\text{var}(\hat{\rho}_A) =& \frac{2(1-\rho)^2}{n_0^2}\{\frac{[1+\rho(n_0-1)]^2}{N-k} \\ &+ \frac{(k-1)(1-\rho)[1+\rho(2n_0-1)]+\rho^2[\sum n_i^2 - 2N^{-1}\sum n_i^3 + N^{-2}(\sum n_i^2)^2]}{(k-1)^2}\}.\end{aligned} \quad (2.5)$$

A simpler expression for a large-sample variance of $\hat{\rho}_A$ (Swiger *et al.*, 1964) intended for situations when there is small variation in class sizes is given by

$$\text{var}(\hat{\rho}_A) = \frac{2(N-1)(1-\rho)^2[1+(n_0-1)\rho]^2}{n_0^2(N-k)(k-1)}.$$

In the case of constant class sizes, the Smith's (1956) large sample variance reduces to the variance due to Swiger *et al.* (1964).

A weighted ANOVA estimator, where a weight is associated with each component of the between sum of squares, and its associated asymptotic variance was derived by Smith (1956). There are 3 commonly used weights: class size weights, uniform weights, and pairwise weights. Implementing a uniform weighting scheme leads to assigning identical weights to all classes. On the other hand, in a pairwise weighting scheme each class is weighted by the possible number of pairs of observations that can be constructed within a class. The ANOVA estimator defined in equation (2.3) is a special case of the generalized ANOVA estimator with a class weighting scheme. Similarly, Karlin *et al.* (1981) considered generalization of the pairwise intraclass estimator referred as generalized product moment estimator. The variance of the later was provided by Eliasziw and Donner (1991). The pairwise estimator defined in equation (2.1) is a special case of the generalized pairwise estimators with pairwise weights.

The performance of the weighted estimators depends on the choice of weights. For example, it is known that the ANOVA estimator with uniform weights is efficient for small values of ρ but inefficient for large values of ρ. On the other hand, the ANOVA estimator using the pairwise weighting scheme is more efficient at larger values of ρ than at the lower values (Keen, 1993). As noted by Keen (1993), both the generalized ANOVA and the generalized product-moment estimator utilizing the same weighting scheme perform similarly.

Since no single estimator in closed-form is efficient on the whole range of values of ρ, several authors suggested estimating ρ by combining two weighted estimators. For instance, Srivistava (1993) proposed estimating ρ by algebraically combining two weighted ANOVA estimators with an appropriate choice of weights. In particular, combination of the ANOVA estimator utilizing the pairwise weighting scheme and the ANOVA estimator using the uniform weighting scheme has been shown to be nearly fully efficient on the whole range of the parameter values (Keen, 1996). Keen (1993) extended Srivistava's method by finding the minimum-variance linear combination of a pair of the weighted estimators. Keen (1996) also provided a unified mathematical expression that can be used to calculate the large-sample standard deviation of all previously mentioned non-iterative estimators.

Although based on an iterative procedure, under the common correlation model with an assumption that observations within each class follow a multivariate normal distribution, the maximum likelihood estimator (MLE) of ρ was derived by Donner and Koval (1980a). The large sample variance of the latter is given by Donner and Koval (1980b).

Many biological and psychological characteristics exhibit small to moderate ($<$ 0.5) value of ρ (Donner and Koval, 1980a). Thus, if prior knowledge about the value of ρ is available, the ANOVA estimator utilizing the class size weighting scheme may be used, because its performance in an intermediate range of values of ρ has been shown to be good (Keen, 1996). However, when prior knowledge of ρ is not available, the estimator formed by algebraic combination of the ANOVA estimator with the uniform and the pairwise weights may be implemented as it has been shown by (Keen, 1993, 1996) to be nearly fully efficient over the whole range of the parameter values. It is worth noting that similar results were obtained for the minimum-variance ANOVA combination of the pairwise and the uniform weights. Thus, when prior knowledge of ρ is not available, the assembled estimators may serve as a good alternative to MLE.

2.3 Confidence interval estimation

Confidence interval estimation has been regarded as a preferable way of presenting study results because it encompasses hypothesis testing (Altman, 2005). There are several methods for constructing confidence intervals for ρ in studies with variable class sizes. The exact method given by Wald (1940), due to its complexity in solving two non-linear equations, is not often used in practice. Therefore a number of other methods, although based on the large sample theory, have been proposed in the literature. The following are the description of five such approximate approaches.

2.3.1 Simple asymptotic method

This method assumes that the sampling distribution of $\hat{\rho}$ is approximately normal. In other words, asymptotically the estimated variance of $\hat{\rho}$ is independent of the underlying parameter. This is because Lukacs (1942) has shown that independence between sample mean and sample variance is a necessary and sufficient condition for normality. Thus, assuming adequately large sample size, the skewed sampling distribution of $\hat{\rho}$, based on the Central Limit Theorem, approaches a normal curve so that the simple asymptotic method can be implemented i.e.,

$$\hat{\rho} \sim N(\rho, \mathrm{var}(\hat{\rho})),$$

then

$$\frac{\hat{\rho} - \rho}{\sqrt{\mathrm{var}(\hat{\rho})}} \sim N(0, 1).$$

Let $z_{\alpha/2}$ be the upper $\alpha/2$ quantile of the standard normal distribution, then

$$\Pr\left(-z_{\alpha/2} \leq \frac{\hat{\rho} - \rho}{\sqrt{\mathrm{var}(\hat{\rho})}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Thus, a $100(1-\alpha)\%$ confidence interval for ρ is given by

$$(l, u) = \left(\hat{\rho} - z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{\rho})},\ \hat{\rho} + z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{\rho})}\right). \tag{2.6}$$

Confidence limits based on the ANOVA estimator and the Smith's (1956) large sample variance are obtained simply by substituting the ANOVA estimator for $\hat{\rho}$ and the Smith's large-sample variance estimate for $\widehat{\text{var}}(\hat{\rho})$.

2.3.2 Method based on Fisher's Z-transformation

To normalize the sampling distribution of $\hat{\rho}_A$, Fisher (1925, Ch.7) proposed the variance stabilizing transformation known in the literature as Fisher's Z-transformation. By employing this transformation the resulting distribution of $\hat{\rho}_A$ on the transformed scale is free of the underling parameter value and approaches normality more rapidly. However, this transformation is only useful for constant class size. To accommodate variable class sizes, Fisher's transformation has been modified and it was shown to be accurate for moderately large sample (Weinberg and Patel, 1981). The transformation is given by

$$Z_F = \frac{1}{2}\ln\left[\frac{1+(n_0-1)\hat{\rho}_A}{1-\hat{\rho}_A}\right], \tag{2.7}$$

which follows as asymptotically normal distribution with mean

$$\frac{1}{2}\ln\left[\frac{1+(n_0-1)\rho}{1-\rho}\right]$$

and variance

$$\text{var}(Z_F) = \frac{1}{2}\left[(k-1)^{-1} + (N-k)^{-1}\right],$$

where $N = \Sigma n_i$, $i = 1, \cdots, k$. An approximate $100(1-\alpha)\%$ confidence interval on the transformed scale is given by

$$(Z_l, Z_u) = \left(Z_F - z_{\alpha/2}\sqrt{\text{var}(Z_F)},\ Z_F + z_{\alpha/2}\sqrt{\text{var}(Z_F)}\right).$$

Transforming the confidence interval back to the original scale, the limits are given by

$$(l, u) = \left(\mathrm{I}\{Z_F - z_{\alpha/2}\sqrt{\mathrm{var}(Z_F)}\},\ \mathrm{I}\{Z_F + z_{\alpha/2}\sqrt{\mathrm{var}(Z_F)}\}\right), \tag{2.8}$$

where

$$\mathrm{I}(Z_F) = [\exp(2\mathrm{Z_F}) - 1] \,/\, [\exp(2Z_F) + n_0 - 1] \,.$$

2.3.3 Method based on inverse hyperbolic tangent transformation

Another transformation known in the literature is the inverse hyperbolic tangent transformation given as

$$Z(\hat{\rho}_A) = \mathrm{arctanh}(\hat{\rho}_A) = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_A}{1-\hat{\rho}_A}\right).$$

Fisher's Z-transformation reduces to the inverse tanh hyperbolic transformation when $n_0 = 2$. Lachin (2004) suggested that confidence intervals for ρ employing the inverse hyperbolic tangent transformation with the variance obtained by the Delta method may perform better than that based on Fisher's Z-transformation.

By the Delta method, the variance of $Z(\hat{\rho}_A)$ is

$$\begin{aligned} \mathrm{var}(Z) &= (Z')^2 \mathrm{var}(\hat{\rho}_A) \\ &= \left[\frac{1}{(1-\hat{\rho}_A)(1+\hat{\rho}_A)}\right]^2 \mathrm{var}(\hat{\rho}_A), \end{aligned} \tag{2.9}$$

where Z' is the first derivative of $Z(\hat{\rho}_A)$ with respect to ρ_A. The $\mathrm{var}(\hat{\rho}_A)$ term in equation (2.8) can be estimated using Smith's large sample variance estimator given by the equation (2.5). Then, a $100(1-\alpha)$ confidence interval on the transformed scale for $Z(\hat{\rho}_A)$ is obtained as follows

$$(Z_l, Z_u) = \left(Z - z_{\alpha/2}\sqrt{\widehat{\text{var}}(Z)},\ Z + z_{\alpha/2}\sqrt{\widehat{\text{var}}(Z)}\right).$$

Denote an inverse of this transformation by

$$\text{I}(Z) = [\exp(2Z) - 1] \,/\, [\exp(2Z) + 1].$$

Then, an approximate $100(1-\alpha)\%$ confidence interval for ρ is given by

$$(l, u) = \left(\text{I}\{Z - z_{\alpha/2}\sqrt{\text{var}(Z)}\}, \text{I}\{Z + z_{\alpha/2}\sqrt{\text{var}(Z)}\}\right). \tag{2.10}$$

2.3.4 Method based on the modification to the exact confidence limits

Confidence intervals based on the F-distribution are available in the case of constant class size, $n_i = n$, for all $i = 1, \ldots, k$ (Haggard, 1958; Searle, 1971). Since the limits are obtained using the results that MSA/MSE is distributed as F-distribution they are usually referred as exact limits. In order to accommodate the variable class sizes, Thomas and Hultquist (1978) and Donner (1979) modified the exact formulae for confidence limits for ρ by replacing the constant class size n by n_0. The modified limits are given by

$$(l, u) = \left(\frac{F/F_U - 1}{n_0 + F/F_U - 1}, \frac{F/F_L - 1}{n_0 + F/F_L - 1}\right), \tag{2.11}$$

where

$$F = \text{MSA}/\text{MSE},$$

$$F_L = F_{(\alpha/2, k-1, \sum_{i=1}^{k} n_i - k)},$$

$$doF_U = F_{(1-\alpha/2, k-1, \sum_{i=1}^{k} n_i - k)}.$$

$F_{(\alpha/2,k-1,\sum_{i=1}^{k} n_i-k)}$ and $F_{(1-\alpha/2,k-1,\sum_{i=1}^{k} n_i-k)}$ are the lower and upper quantiles of the F-distribution with degrees of freedom $k-1$ and $\sum_{i=1}^{k} n_i - k$, respectively. These limits are approximate since the distribution of variance-ratio statistics is not exact, unless ρ=0 (Donner and Wells, 1986). Therefore, as ICC increases the approximation to the central F-distribution may decline.

2.3.5 Method based on F-distribution

Thomas and Hultquist (1978) suggested that the above results can be improved by replacing n_0 with the harmonic mean class size and MSA with unweighted sum of squares. Specifically, with the harmonic mean class size defined by $\hat{n} = k/\sum_{i=1}^{k}(1/n_i)$, the confidence interval for a single intraclass correlation coefficient is given by

$$(l, u) = \left(\frac{F^*/F_U - 1}{\hat{n} + F^*/F_U - 1}, \frac{F^*/F_L - 1}{\hat{n} + F^*/F_L - 1}\right), \tag{2.12}$$

where

$$F^* = \hat{n}[\sum_{i=1}^{k} \overline{Y}_{i.}^2 - \frac{1}{k}(\sum_{i=1}^{k} \overline{Y}_{i.})^2]/[(k-1)\text{MSE}],$$

$$F_L = F_{(\alpha/2,k-1,\Sigma_{i=1}^{k} n_i-k)},$$

$$F_U = F_{(1-\alpha/2,k-1,\Sigma_{i=1}^{k} n_i-k)}.$$

In the context of familial resemblance studies under variable sibship sizes for normally distributed data, Donner and Wells (1986) and Donner (1987) evaluated a number of methods for interval estimation of ρ. They concluded that a simple asymptotic method incorporating Smith's 1956 large sample variance provided consistently good coverage for all values of ρ. They also showed that application of Fisher's Z-transformation provided good coverage probabilities for values of $\rho \leq 0.3$ and adequate coverage for values as high as $\rho < 0.7$. Further, evaluation of method based on the F-distribution showed good coverage over all values of ρ but the width of confidence intervals were wider than that of simple asymptotic method. In conclusion, they recommended that the simple asymptotic method based on the Smith's (1956) large sample variance be implemented in practice. In their evaluation, confidence intervals based on the inverse hyperbolic tangent transformation were not considered.

2.4 Inference for intraclass correlation coefficients in multiple samples

Hypothesis testing procedures for a single ICC based on the normal distribution with variable class sizes and its extension to multiple independent samples are well established. The focus has been on estimation of the common ICC and constructing tests of homogeneity of several ICCs. For instance, under a mixed ANOVA model, Donner (1985) derived an estimator of common ICC and provided a test for homogeneity of two ICCs based on the extension to Fisher's variance stabilizing transformation. Young and Bhandary (1998) proposed three asymptotic tests including: likelihood ratio test and two based on the large sample Z-test for the equality of ICCs for two independent multivariate samples. Bhandary and Alam (2000) extended this approach to several multivariate samples. It is worth noting that a unified test for the equality of ICCs in multiple sample cases has been constructed (Bhandary and Fujiwara, 2006).

When the assumption of homogeneity of variances across samples does not hold, Mian and Shoukri (1997) considered estimating the common ICC by ANOVA and the maximum likelihood method and proposed several tests for the equality of ICCs. In the same context a number of other asymptotic tests have been proposed by Naik and Helu (2007).

In contrast to hypothesis testing, literature on confidence interval estimation is limited. Nonetheless, Donner (1985) considered constructing confidence intervals for the common ICC based on the modification to the Fisher's Z-transformation. Recently, Bhandary and Fujiwara (2008) discussed confidence intervals for a linear contrast of ICCs. Two flaws are apparent in their presentation. First, their method is only valid if the between class variances are equal in the two populations being compared. This is a rather restrictive assumption. Second, they specified the lower limit for the difference between two ICCs as the difference between the lower limit for one ICC and the upper limit for another ICC, and similarly for the upper limit. It can be recognized that this method is equivalent to one using overlapping to perform hypothesis testing. Shenker and Gentleman (2001) discussed the disadvantages of using overlap between confidence intervals as a method for significance testing. Under assumption of normality, independence, and consistency of two estimators, they have shown that the overlap method is more conservative and less powerful in comparison to the traditional method.

The methods described above aimed at ICCs obtained from independent samples. In the case of constant class size, Donner and Zou (2002) proposed a procedure for testing the equality of two dependent ICCs implementing the modified form of Fisher's Z-transformation due to Konishi and Gupta (1987). Donner and Zou (2002) showed by their simulation study that the proposed method has greater power than a procedure based on the traditional Fisher's Z-transformation, and it is comparable with that of the likelihood ratio test. Since no explicit formula exists for the likelihood ratio statistics for testing two dependent ICCs and thus for calculating the power of the test, iterative procedures have been used. For macros that allow for such tests and power calculations are provided by Giraudeau *et al.* (2005). The likelihood ratio test for testing two dependent ICCs with variable class sizes has been provided by Donner *et al.* (1984). A confidence interval for a difference of two dependent ICCs was developed by Ramasundarahettige *et al.* (2009) applying an approach proposed by Zou and Donner (2008). Ramasundarahettige *et al.* (2009) showed by their simulation study that the proposed method performs superior to the conventional method of point estimate plus/minus a critical value multiplied by the estimated standard error.

Chapter 3

CONFIDENCE INTERVAL ESTIMATION FOR A DIFFERENCE BETWEEN TWO INTRACLASS CORRELATION COEFFICIENTS

3.1 Introduction

Confidence interval for a difference between two correlated ICCs with constant class size has been recently developed in the context of reliability studies (Ramasundarahettige *et al.*, 2009). The confidence interval approach relied on method of variance estimates recovery (MOVER) as outlined in Zou and Donner (2008). The MOVER is most useful when confidence limits for single ICCs are available. In this chapter, the MOVER is used to construct confidence interval for a difference between two correlated ICCs with an extension to variable class sizes.

This chapter starts with providing the definition and interpretation of a confidence interval given in Section 3.2. The MOVER is presented in Section 3.3. The chapter ends with the construction of the proposed confidence interval for a difference between two correlated ICCs with variable class sizes presented in Section 3.4.

3.2 *Definition and interpretation of a confidence interval*

Denote a population parameter by θ. For θ, a two-sided $100(1-\alpha)\%$ confidence interval, (l, u), provides a range of plausible parameter values that cannot be rejected at the $\alpha\%$ level. Then, the upper limit, u, is a value that, under random samples, may be expected to exceed θ's true parameter value $100(1-\alpha/2)\%$ of the time. Similarly, the lower limit, l, is a value that, under random samples, may fall below θ's true parameter value $100(1-\alpha/2)\%$ of the time. In other words, the lower limit l satisfies

$$\Pr(\theta \leq l) = \alpha/2$$

and the upper limit u satisfies

$$\Pr(\theta \geq u) = \alpha/2.$$

Therefore, a $100(1-\alpha)\%$ confidence interval satisfies

$$\Pr(l < \theta < u) = (1-\alpha).$$

A two-sided $100(1-\alpha)\%$ confidence interval is constructed from the observed data in such a way that under infinite replications of the study, $100(1-\alpha)\%$ of these confidence intervals will contain the true parameter value.

Confidence intervals answer a more general question than hypothesis testing, providing not only a yes/no answer but also the magnitude and direction of the effect in question. To this end, it may be more useful to consider inferences consisting not only in terms of point estimation and hypothesis testing, but rather in terms of quantifying the uncertainty in the estimation process by means of confidence interval construction.

3.3 The method of variance estimates recovery (MOVER)

This section describes a general approach for constructing a confidence interval about a difference between two effect measures as derived by Zou and Donner (2008). This approach has been previously applied in many special cases. For instance, it has been applied to construct confidence intervals for the lognormal mean (Zou and Donner, 2008), a difference between correlation coefficients (Zou, 2007), and measures of additive interaction (Zou, 2008). Recently, it has been applied to comparison of two correlated intraclass correlation coefficients under constant class size (Ramasundarahettige *et al.*, 2009).

The fundamental idea of the method of variance estimates recovery is to recover variance estimates from already reliable limits for each separate effect measure, and then use these estimated variances to construct limits about the difference. The method is formally referred to as the MOVER, method of variance estimates recovery (Zou, 2008).

Let θ_i be a respective parameter of interest for population i, $i = 1, 2$, with a point estimate and variance estimate given as $\hat{\theta}_i$ and $\widehat{\text{var}}(\hat{\theta}_i)$, respectively.

A traditional approach for an approximate confidence interval for $\theta_1 - \theta_2$ is based on the Central Limit Theorem. The resulting $100(1-\alpha)\%$ confidence limits, (L, U), are then given by

$$(L, U) = \hat{\theta}_1 - \hat{\theta}_2 \pm z_{\alpha/2}\sqrt{\text{var}(\hat{\theta}_1 - \hat{\theta}_2)}, \quad (3.3.1)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Under the assumption that $\hat{\theta}_1$ and $\hat{\theta}_1$ are independent,

$$\text{var}(\hat{\theta}_1 - \hat{\theta}_2) = \text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2),$$

otherwise

$$\text{var}(\hat{\theta}_1 - \hat{\theta}_2) = \text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2) - 2\text{cov}(\hat{\theta}_1, \hat{\theta}_2).$$

In a simple asymptotic method, $\text{var}(\hat{\theta}_1 - \hat{\theta}_2)$ is estimated assuming $\theta_1 - \theta_2 \doteq \hat{\theta}_1 - \hat{\theta}_2$. This further implies that $\widehat{\text{var}}(\hat{\theta}_1 - \hat{\theta}_2)$ is the same at the lower (L) and the upper limit (U) of $\theta_1 - \theta_2$ and the obtained $100(1-\alpha)\%$ confidence interval is symmetric. The resulting confidence interval based on the simple asymptotic method is given as follows:

$$(L, U) = \hat{\theta}_1 - \hat{\theta}_2 \pm z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{\theta}_1 - \hat{\theta}_2)}.$$

Validity of the simple asymptotic procedure is assured if $\hat{\theta}_i$ follows a normal distribution with constant variance. When these assumptions are not satisfied, the imposed symmetry on the confidence intervals is considered as the most severe error one can make in confidence estimation (Efron and Tibshirani, 1993, p. 180).

Recognizing the limitation of the traditional method, a general approach for constructing a confidence interval for a difference between two effect measures has been given by Zou and Donner (2008), which requires no specific sampling distribution for $\hat{\theta}_i$, but only reliable confidence intervals for θ_i.

Suppose that a two-sided $100(1-\alpha)\%$ confidence interval (l_i, u_i) for θ_i is available. Using the Central Limit Theorem,

$$l_i = \hat{\theta}_i - z_{\alpha/2}\sqrt{\text{var}(\hat{\theta}_i)}.$$

Without imposing symmetry about $\hat{\theta}_i$, one can recover variances estimates for $\hat{\theta}_i$ at the lower and upper limits of θ_i. Thus, under the assumption that $\theta_i \approx l_i$, the estimated variance of $\hat{\theta}_i$ is given by

$$\widehat{\text{var}}_l(\hat{\theta}_i) = \frac{(\hat{\theta}_i - l_i)^2}{z^2_{\alpha/2}}.$$

Similarly,

$$u_i = \hat{\theta}_i + z_{\alpha/2}\sqrt{\text{var}(\hat{\theta}_i)}$$

and thus under the assumption of $\theta_i \approx u_i$, the estimated variance of $\hat{\theta}_i$ is given by

$$\widehat{\text{var}}_u(\hat{\theta}_i) = \frac{(u_i - \hat{\theta}_i)^2}{z^2_{\alpha/2}}.$$

Denote the $100(1-\alpha)\%$ lower and upper limit for $\theta_1 - \theta_2$ as L and U, respectively. Then, $l_1 - u_2$ is the value near L and $u_1 - l_2$ is the value near U. To retrieve the variance estimates of $\hat{\theta}_1 - \hat{\theta}_2$ at the neighborhood of L, $\text{var}(\hat{\theta}_1)$ and $\text{var}(\hat{\theta}_2)$ are estimated under $\theta_1 \approx l_1$ and $\theta_2 \approx u_2$, respectively (Zou and Donner, 2008). Then the lower limit L is obtained by substituting the retrieved variances into the asymptotic formula (3.3.1) for L

$$\begin{aligned} L &= \hat{\theta}_1 - \hat{\theta}_2 - z_{\alpha/2}\sqrt{\text{var}_l(\hat{\theta}_1 - \hat{\theta}_2)} \\ &= \hat{\theta}_1 - \hat{\theta}_2 - z_{\alpha/2}\sqrt{\frac{(\hat{\theta}_1 - l_1)^2}{z^2_{\alpha/2}} + \frac{(u_2 - \hat{\theta}_2)^2}{z^2_{\alpha/2}}} \qquad (3.3.2) \\ &= \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2}. \end{aligned}$$

Following the same logic, we obtain the estimated variance of $\hat{\theta}_1 - \hat{\theta}_2$ near U so that $\text{var}(\hat{\theta}_1)$ and $\text{var}(\hat{\theta}_2)$ are estimated under $\theta_1 \approx u_2$ and $\theta_2 \approx l_2$, respectively. The upper limit for $\theta_1 - \theta_2$ is obtained by substituting the retrieved variances into the asymptotic formula (3.3.1) for U

$$\begin{aligned} U &= \hat{\theta}_1 - \hat{\theta}_2 + z_{\alpha/2}\sqrt{\text{var}_u(\hat{\theta}_1 - \hat{\theta}_2)} \\ &= \hat{\theta}_1 - \hat{\theta}_2 + z_{\alpha/2}\sqrt{\frac{(u_1 - \hat{\theta}_1)^2}{z^2_{\alpha/2}} + \frac{(\hat{\theta}_2 - l_2)^2}{z^2_{\alpha/2}}} \\ &= \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2}. \end{aligned} \tag{3.3.3}$$

A confidence interval for a difference between two effect measures, $\theta_1 - \theta_2$, obtained by the MOVER can be reduce to a confidence interval obtained by the traditional method when the confidence intervals for θ_i are symmetric, as shown below:

Now, the lower and upper confidence limits for θ_i are given by

$$\begin{aligned} l_i &= \hat{\theta}_i - z_{\alpha/2}\sqrt{\text{var}(\hat{\theta}_i)} \\ u_i &= \hat{\theta}_i + z_{\alpha/2}\sqrt{\text{var}(\hat{\theta}_i)} \end{aligned}$$

Then,

$$\begin{aligned} (\hat{\theta}_i - l_i)^2 &= z^2_{\alpha/2}\text{var}(\hat{\theta}_i) \\ (u_i - \hat{\theta}_i)^2 &= z^2_{\alpha/2}\text{var}(\hat{\theta}_i). \end{aligned}$$

If l_i and u_i are symmetric around $\hat{\theta}_i$, then the estimated variances for θ_i at the lower and upper limit of θ_i are the same. Hence, for θ_1

$$(\hat{\theta}_1 - l_1)^2 = (u_1 - \hat{\theta}_1)^2 = z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_1) \tag{3.3.4}$$

Similarly, for θ_2

$$(\hat{\theta}_2 - l_2)^2 = (u_2 - \hat{\theta}_2)^2 = z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_2). \tag{3.3.5}$$

When equations (3.3.4) and (3.3.5) are substituted into equation (3.3.2) and (3.3.3), the lower and the upper limit for $\theta_1 - \theta_2$ reduce to limits obtained by the simple asymptotic method shown below

$$\begin{aligned} L &= \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2} \\ &= \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_1) + z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_2)} \\ &= \hat{\theta}_1 - \hat{\theta}_2 - z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{\theta}_1) + \widehat{\text{var}}(\hat{\theta}_2)}, \end{aligned}$$

$$\begin{aligned} U &= \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2} \\ &= \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_1) + z^2_{\alpha/2}\widehat{\text{var}}(\hat{\theta}_2)} \\ &= \hat{\theta}_1 - \hat{\theta}_2 + z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{\theta}_1) + \widehat{\text{var}}(\hat{\theta}_2)}. \end{aligned}$$

3.4 Confidence interval for a difference between two intraclass correlated coefficients

3.4.1 Notation and terminology

Suppose we have data collected on two dependent groups: group 1 and group 2, and in each of these groups observations on a single variable Y are arranged in k classes. Let $\mathbf{Y}_\text{i} = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i}, Y_{i,n_i+1}, Y_{i,n_i+2}, \ldots, Y_{i,n_i+m_i})$, where $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ represent the observations from the i^{th} class in group 1 and $Y_{i,n_i+1}, Y_{i,n_i+2}, \ldots, Y_{i,n_i+m_i}$ represent the observations from the i^{th} class in group 2; $i = 1, \ldots, k$, n_i and m_i are the sizes of the i^{th} class in group 1 and group 2, respectively. The following model is assumed:

$$\mathbf{Y_i} = (Y_{i1}, Y_{i2} \dots, Y_{in_i}, Y_{i,n_i+1}, Y_{i,n_i+2}, \dots, Y_{i,n_i+m_i}) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

where $\boldsymbol{\mu}_i = (\mu_1, \mu_1, \dots, \mu_1, \mu_2, \mu_2, \dots, \mu_2)$, and

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} [(1-\rho_1)\mathbf{I}_{n_i} + \rho_1 \mathbf{J}_{n_i}]\sigma_1^2 & \rho_{12}\sigma_1\sigma_2\mathbf{J}_{n_i \times m_i} \\ \rho_{12}\sigma_1\sigma_2\mathbf{J}_{n_i \times m_i} & [(1-\rho_2)\mathbf{I}_{m_i} + \rho_2 \mathbf{J}_{m_i}]\sigma_2^2 \end{pmatrix} \tag{3.4.1}$$

In expression (3.4.1), $\mathbf{I}_{n_i}$ and $\mathbf{I}_{m_i}$ are $n_i \times n_i$ and $m_i \times m_i$ identity matrices, $\mathbf{J}_{n_i}$ and $\mathbf{J}_{m_i}$ are $n_i \times n_i$ and $m_i \times m_i$ matrices with all the elements equal to 1 and, $\mathbf{J}_{n_i \times m_i}$ is a $n_i \times m_i$ matrix for which all elements equal 1. Under this model, it is assumed that members from the k classes in group 1 have a common mean μ_1 and variance σ_1^2. Similarly, members from k classes in group 2 have a common mean μ_2 and variance σ_2^2. The intraclass correlations for group 1 and group 2 are denoted by ρ_1 and ρ_2, respectively. Furthermore, the interclass correlation, ρ_{bs}, is used to quantify the degree of resemblance among members belonging to two different groups (with the mean and variance of each being estimated separately) with respect to some quantitative characteristic. The model assumes also that σ_1^2, σ_2^2, ρ_1, ρ_2, ρ_{12} are constant across all classes.

Due to the complexity of obtaining maximum likelihood estimators of the three correlations, the ANOVA estimators are used to estimate ρ_1 and ρ_2. The corresponding estimates of ρ_1 and ρ_2 are given by

$$\hat{\rho}_t = \frac{\text{MSA}_t - \text{MSE}_t}{\text{MSA}_t + (n_{t0} - 1)\text{MSE}_t}, \quad t = 1, 2.$$

The pairwise estimator is used, for simplicity of calculation, to estimate the interclass correlation ρ_{12}. The pairwise estimator of ρ_{12} is obtained by computing the ordinary Pearson product-moment correlation over the set of $\Sigma_{i=1}^{k} n_i m_i$ pairs of observations $(Y_{ij}, Y_{i,n_i+j'})$ in the sample (Rosner, 1982), given by

$$\hat{\rho}_{12} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{j'=n_i+1}^{n_i+m_i}\left(y_{ij}-\bar{\bar{y}}_1\right)\left(y_{ij'}-\bar{\bar{y}}_2\right)}{\left[\left\{\sum_{i=1}^{k} n_i \sum_{j=1}^{n_i}(y_{ij}-\bar{\bar{y}}_1)^2\right\}\cdot\left\{\sum_{i=1}^{k} m_i \sum_{j'=n_i+1}^{n_i+m_i}(y_{ij'}-\bar{\bar{y}}_2)^2\right\}\right]^{1/2}}, \tag{3.4.2}$$

where

$$\bar{\bar{y}}_1 = \frac{\sum_{i=1}^{k} n_i m_i \bar{y}_{i1}}{\sum_{i=1}^{k} n_i m_i},\ \bar{\bar{y}}_2 = \frac{\sum_{i=1}^{k} n_i m_i \bar{y}_{i2}}{\sum_{i=1}^{k} n_i m_i},\ \bar{y}_{i1} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i},\ \bar{y}_{i2} = \frac{\sum_{j'=n_i+1}^{n_i+m_i} y_{ij'}}{m_i}.$$

3.4.2 Confidence interval for a difference between two correlated ICCs using the MOVER

Let a $100(1-\alpha)\%$ confidence interval for ρ_1 and ρ_2 be (l_1, u_2) and (l_2, u_2), respectively. Then, under the assumption that $\hat{\rho}_1$ and $\hat{\rho}_2$ are independent a $100(1-\alpha)\%$ confidence interval for $\rho_1 - \rho_2$ is developed. Applying MOVER to the problem of estimating limits for two independent intraclass correlations results in

$$\begin{aligned} L &= (\hat{\rho}_1 - \hat{\rho}_2) - z_{\alpha/2}\sqrt{\frac{(\hat{\rho}_1 - l_1)^2}{z_{\alpha/2}^2} + \frac{(u_2 - \hat{\rho}_2)^2}{z_{\alpha/2}^2}} \\ &= (\hat{\rho}_1 - \hat{\rho}_2) - \sqrt{(\hat{\rho}_1 - l_1)^2 + (u_2 - \hat{\rho}_2)^2}. \end{aligned} \tag{3.4.3}$$

Similarly, the upper limit for $\rho_1 - \rho_2$ is given by

$$\begin{aligned} U &= (\hat{\rho}_1 - \hat{\rho}_2) + z_{\alpha/2}\sqrt{\frac{(u_1 - \hat{\rho}_1)^2}{z_{\alpha/2}^2} + \frac{(\hat{\rho}_2 - l_2)^2}{z_{\alpha/2}^2}} \\ &= (\hat{\rho}_1 - \hat{\rho}_2) + \sqrt{(u_1 - \hat{\rho}_1)^2 + (\hat{\rho}_2 - l_2)^2}. \end{aligned} \tag{3.4.4}$$

In the case when two $\hat{\rho}_1, \hat{\rho}_2$ are dependent,

$$\text{var}(\hat{\rho}_1 - \hat{\rho}_2) = \text{var}(\hat{\rho}_1) + \text{var}(\hat{\rho}_2) - 2\text{cov}(\hat{\rho}_1, \hat{\rho}_2)$$

The covariance term by the definition is given as

$$\text{corr}(\hat{\rho}_1, \hat{\rho}_2) = \frac{\text{cov}(\hat{\rho}_1, \hat{\rho}_2)}{\sqrt{\text{var}(\hat{\rho}_1) \times \text{var}(\hat{\rho}_2)}},$$

by which

$$\text{cov}(\hat{\rho}_1, \hat{\rho}_2) = \text{corr}(\hat{\rho}_1, \hat{\rho}_2) \times \sqrt{\text{var}(\hat{\rho}_1) \times \text{var}(\hat{\rho}_2)}.$$

When class sizes are constant, $n_i = n$ and $m_i = m$, Elston (1975) derived an asymptotic formula for $\text{var}(\hat{\rho}_t)$ given by

$$\text{var}(\hat{\rho}_t) = \frac{2(s-1)}{ks}(1-\rho_t)^2 \left(\frac{1}{s-1} + \rho_t\right)^2,$$

$t = 1, 2$ and $s = n, m$. Also, he derived an asymptotic formula for $\text{cov}(\hat{\rho}_1, \hat{\rho}_2)$ given by

$$\text{cov}(\hat{\rho}_1, \hat{\rho}_2) = \frac{2\rho_{12}^2}{k}(1-\rho_1)(1-\rho_2)$$

Then,

$$\begin{aligned}
\text{corr}(\hat{\rho}_1, \hat{\rho}_2) &= \frac{\text{cov}(\hat{\rho}_1, \hat{\rho}_2)}{\sqrt{\text{var}(\hat{\rho}_1) \times \text{var}(\hat{\rho}_2)}} \\
&= \frac{\frac{2\rho_{12}^2}{k}(1-\rho_1)(1-\rho_2)}{\sqrt{\frac{2(n-1)}{kn}(1-\rho_1)^2\left(\frac{1}{n-1}+\rho_1\right)^2} \times \sqrt{\frac{2(m-1)}{km}(1-\rho_2)^2\left(\frac{1}{m-1}+\rho_2\right)^2}} \\
&= \frac{\rho_{12}^2\left[nm(n-1)(m-1)\right]^{1/2}}{[1+(n-1)\rho_1][1+(m-1)\rho_2]}.
\end{aligned}$$

An extension of the correlation formula to variable class sizes is considered by replacing n and m by $\hat{n}$ and $\hat{m}$, respectively, where $\hat{n}$ and $\hat{m}$ correspond to harmonic means of class sizes in group 1 and group 2 . The resulting modified formula is given by

$$\mathrm{corr}(\hat{\rho}_1, \hat{\rho}_2) = \frac{\rho_{12}^2 \left[\hat{n}\hat{m}(\hat{n}-1)(\hat{m}-1)\right]^{1/2}}{[1+(\hat{n}-1)\rho_1][1+(\hat{m}-1)\rho_2]}.$$

The estimator of $\mathrm{corr}(\rho_1, \hat{\rho}_2)$ is obtained by replacing ρ_{12}, ρ_1 and ρ_2 by their respective estimators, i.e.,

$$\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2) = \frac{\hat{\rho}_{12}^2 \left[\hat{n}\hat{m}(\hat{n}-1)(\hat{m}-1)\right]^{1/2}}{[1+(\hat{n}-1)\hat{\rho}_1][1+(\hat{m}-1)\hat{\rho}_2]}. \tag{3.4.5}$$

A $100(1-\alpha)\%$ confidence limits for the difference of two correlated ICCs are now constructed by taking into account the correlation between $\hat{\rho}_1$ and $\hat{\rho}_2$. Thus, the confidence interval for a difference of two independent ICCs given in equation (3.4.3) and equation (3.4.4) can be extended to the case of two correlated ICCs, and the following confidence interval limits are obtained

$$L = (\hat{\rho}_1 - \hat{\rho}_2) - z_{\alpha/2}\sqrt{\frac{(\hat{\rho}_1 - l_1)^2}{z^2_{\alpha/2}} + \frac{(u_2 - \hat{\rho}_2)^2}{z^2_{\alpha/2}} - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)\sqrt{\frac{(\hat{\rho}_1 - l_1)^2}{z^2_{\alpha/2}}}\sqrt{\frac{(u_2 - \hat{\rho}_2)^2}{z^2_{\alpha/2}}}}$$

$$= (\hat{\rho}_1 - \hat{\rho}_2) - \sqrt{(\hat{\rho}_1 - l_1)^2 + (u_2 - \hat{\rho}_2)^2 - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)(\hat{\rho}_1 - l_1)(u_2 - \hat{\rho}_2)}$$

and

$$U = (\hat{\rho}_1 - \hat{\rho}_2) + z_{\alpha/2}\sqrt{\frac{(u_1 - \hat{\rho}_1)^2}{z^2_{\alpha/2}} + \frac{(\hat{\rho}_2 - l_2)^2}{z^2_{\alpha/2}} - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)\sqrt{\frac{(u_1 - \hat{\rho}_1)^2}{z^2_{\alpha/2}}}\sqrt{\frac{(\hat{\rho}_2 - l_2)^2}{z^2_{\alpha/2}}}}$$

$$= (\hat{\rho}_1 - \hat{\rho}_2) + \sqrt{(u_1 - \hat{\rho}_1)^2 + (\hat{\rho}_2 - l_2)^2 - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)(u_1 - \hat{\rho}_1)(\hat{\rho}_2 - l_2)}.$$

Thus, 95% lower and upper limit for $\rho_1 - \rho_2$ are given by

$$L = (\hat{\rho}_1 - \hat{\rho}_2) - \sqrt{(\hat{\rho}_1 - l_1)^2 + (u_2 - \hat{\rho}_2)^2 - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)(\hat{\rho}_1 - l_1)(u_2 - \hat{\rho}_2)} \tag{3.4.6}$$

$$U = (\hat{\rho}_1 - \hat{\rho}_2) + \sqrt{(u_1 - \hat{\rho}_1)^2 + (\hat{\rho}_2 - l_2)^2 - 2\widehat{\mathrm{corr}}(\hat{\rho}_1, \hat{\rho}_2)(u_1 - \hat{\rho}_1)(\hat{\rho}_2 - l_2)}. \tag{3.4.7}$$

Chapter 4

SIMULATION STUDY

4.1 Introduction

The derivation in Chapter 3 using the MOVER is based on large sample theory. Therefore, a Monte Carlo simulation study is undertaken to evaluate the performance of the methods in terms of empirical coverage, balance of tail errors, and interval width. This chapter provides the description of the simulation design in Section 4.2. Simulation results are provided in Section 4.3, followed by a discussion in Section 4.4.

4.2 Simulation design

4.2.1 Parameter selection

We designed the simulation study in the context of family studies. The parameters for the simulation study included: the number of families k, values of ρ_b, ρ_s and ρ_{bs} corresponding to brother-brother correlation, sister-sister correlation, and brother-sister correlation. The total number of families was chosen to be $k = 50, 100, 200$ to reflect small, medium, and large sample sizes. The variable family sizes were generated using the negative binomial distribution truncated below one, given by

$$Pr(n) = \frac{(m+n-1)!Q^{-1}(P/Q)^n}{(m-1)!)m!(1-Q^{-m})} n = 1, 2, \ldots, Q = 1 + P.$$

Brass (1958) has shown that this distribution fits the observed distribution of sibship sizes for most of the human populations provided appropriate choice of parameters, m and P. In order to avoid generating unreasonably large class sizes we truncated the extreme right tails of class sizes at the probability of .999. Values of $m = 2.84$ and $P = 0.93$ were chosen corresponding to the mean sibship size of 3.12 and the variance of family size 4.52, reported by Brass (1958) for the United States in 1950. To classify each sibling as a male or a female, a random number generator based on the binomial distribution was used. The probability of a male or a female was taken as 0.5. In order to evaluate MOVER, a wide range of parameter values were considered, $\rho_b = 0.1$ 0.3, 0.5, 0.7, 0.9; $\rho_s = 0.1$, 0.3, 0.5, 0.7, 0.9; $\rho_{bs} = 0.1$ to $\min(\rho_b,\rho_s)$ with an increment of 0.2 such that the variance-covariance matrix 3.4.1 is positive definite. Without loss of generality, $\mu_b = \mu_s = 0$ and $\sigma_b^2 = \sigma_s^2 = 1$, where μ_b, μ_s and σ_b^2, σ_s^2 are the common mean and common variance of the observations obtained on brothers and sisters, respectively. In the simulation study 129 parameter combinations were considered. For each parameter combination, observations for brothers and sisters were generated from a multivariate normal distribution with the correlated structure defined by equation 3.4.1 in Section 3.1. The number of runs was set to 10000 such that the empirical coverage is expected (with 95% confidence interval) to vary between 94.6% to 95.4% for a nominal level of 95%.

4.2.2 Confidence interval methods compared

The proposed method uses limits for a single ICC to set limits for the difference of two correlated ICCs. Four different confidence interval procedures were used for single ICCs. Namely, the simple asymptotic approach with Smith's large sample variance, Fisher's Z-transformation method, application of the inverse hyperbolic tangent transformation, and approach based on the F-distribution due to Thomas and Hultquist (1978). These are described in Section 2.2.

4.2.3 Evaluation criteria

4.2.3.1 Coverage

The coverage of a confidence interval is defined as the proportion of times, under repeated sampling, the obtained confidence interval contains the true parameter value. The usual confidence level for an interval is taken as 95%. Hence, a confidence interval procedure whose empirical coverage approaches that value is said to perform well. When the empirical coverage is short of the nominal level, the confidence interval is said to be too liberal. On the other hand, when the empirical coverage is above the desired interval, the confidence interval is considered to be conservative.

4.2.3.2 Tail errors

The coverage alone is not enough to assess the performance of a confidence interval. As pointed out by Efron and Tibshirani (1993), balance of tail errors is also important and should be evaluated. When a confidence interval misses the parameter value from the left, it refers to as missing from the left (ML). Similarly, when a confidence interval misses the parameter value from the right it refers to as missing from the right (MR). Consequently, the evaluation of total tail errors as well as the balance between the left tail and the right tail are considered. Tail errors are estimated by calculating the frequencies of the 10000 intervals lying completely to the left of the parameter value (ML), and similarly, for those lying completely to the right of the parameter values (MR). The tail imbalance is measured by

$$\text{Relative imbalance }(\%) = \frac{|\text{MR} - \text{ML}|}{\text{MR} + \text{ML}} \times 100.$$

The smaller the relative imbalance the less unbalanced the tail errors are.

4.2.3.3 Interval width

The performance of the methods considered in the simulation study was also analyzed with respect to interval width. A narrower confidence interval can be considered to have a greater accuracy, so that a procedure yielding a narrower confidence interval is preferable. The width of a confidence interval was calculated by subtracting the upper limit from the lower limit.

4.3 Simulation results

The results of the study for a 95% confidence interval for a difference of two correlated ICCs under variable class sizes are given in Tables 4.1-4.3 corresponding to small, medium and large sample size, respectively. Each table presents the comparative performance of MOVER incorporating four different methods with respect to empirical coverage, tail errors, and interval width. Moreover, the performance of the methods under consideration is graphically illustrated in Figure 4.1-4.3. Each respective figure presents the results split over small, medium, and large sample sizes with respect to the evaluative criteria. Specifically, Figure 4.1 presents box plots of empirical coverage percentages where on each box plot a horizontal line of 95% was drawn to reflect the nominal level reference. Also, two additional horizontal lines were drawn on each box plots to locate reference lines of the expected empirical coverage of 94.6% and 95.4%. Figure 4.2 presents the imbalance of tail errors across all four methods measured by the relative imbalance. The smaller the relative imbalance the less unbalanced the tail errors are. At last, the performance of the four methods with regards to a confidence interval width is presented in Figure 4.3.

The results in Table 4.1 indicate that, when the number of classes is $k = 50$, the simple asymptotic method provides good empirical coverage when both ρ_b and $\rho_s \leq 0.5$. For the rest of the parameter combinations, the method exhibits erratic behavior. For instance, when $\rho_b = 0.9$ and $\rho_s = 0.5$ this method shows undercoverage as low as 93.19% and for $\rho_s = 0.9$ and $\rho_s = 0.9$ overcoverage as high as 99.19%. As sample size increases, Tables 4.2-4.3 show that the performance of the simple asymptotic is less erratic. However, the method still provides empirical coverage in excess of the nominal level of 95%, especially when values of ρ_b and ρ_s are high and of the same value ($\rho_b = \rho_s = 0.9$, $\rho_b = \rho_s = 0.7$ and $\rho_{bs} \geq 0.5$). Furthermore, the method provides unbalanced tail errors in comparison to the other methods, except when both ρ_b, $\rho_s \leq 0.5$, across all sample sizes. It is worth mentioning that when $k \geq 100$, even though this method has a good coverage for $\rho_b = 0.9$ and $\rho_s = 0.5, 0.7$, the tail errors are severely unbalanced. In other words, the total error is concentrated in one tail, either the right or the left tail. The confidence interval width obtained by simple asymptotic method is comparable to that of other methods (except for F-distribution based method) across all sample sizes.

The results in Tables 4.1-4.3 show that the method based on Fisher's Z-transformation provides empirical coverage percentages close to the nominal level of 95% at all levels of interclass correlations given that both ρ_b and $\rho_s \leq 0.5$. When ρ_b and and $\rho_s > 0.5$, the empirical coverage percentages are close to the nominal level provided that $\rho_{bs} \geq 0.5$. In other cases, the method has a tendency to fall below the nominal level resulting in undercoverage. With regards to the tail errors, this method does not show a high degree of imbalance on the tails and for most of the parameter combinations the tails are symmetrically distributed between the left and the right tails. Also, this method is comparable to the other methods under consideration in terms of interval width providing slightly narrower interval width over the entire parameter space. As sample size increases the coverage of this method improves although it still provides slight undercoverage over the entire parameter spectrum.

The results in Tables 4.1-4.3 also show that the inverse hyperbolic tangent transformation method provides consistently good empirical coverage percentages for all values of ρ_b and ρ_s provided ρ_{bs} is ≤ 0.3. When ρ_{bs} increases to 0.5, although obtained empirical coverage percentages are good, they have a tendency to be slightly farther away from the nominal level of 95%. When both ρ_b and ρ_s are 0.7 or 0.9 and high values of ρ_{bs} (≥ 0.7) are considered, this method is inclined to slight overcoverage. Tail errors obtained by this method are balanced in most cases, showing a minor degree of imbalance in tail errors when both ρ_b and $\rho_s \leq 0.5$ and $\rho_{bs} \leq 0.1$. Interval width is comparable to that of other methods (except for method based on F-distribution) across all sample sizes.

The performance of the method based on the F-distribution displays a similar pattern across all sample sizes (Tables 4.1-4.3). In particular, this method tends to provide overcoverage when both ρ_b and ρ_s are ≤ 0.5. Also, overcoverage is evident when both ρ_b and ρ_s are ≥ 0.7 and $\rho_{bs} \geq 0.5$. Otherwise, the coverage percentages are acceptable yet more conservative in comparison to the other three methods discussed previously. In terms of tail error imbalance, the tails are symmetric over the whole parameter space. In comparison to all other methods discussed above, the method based on the F-distribution exhibits the widest confidence interval at almost every parameter value combination. Nonetheless, when both ρ_b and ρ_s are ≥ 0.7, the discrepancy in terms of interval width between the four methods appears to vanish.

4.4 Discussion

Simulation results have shown that the simple asymptotic method, the Fisher's Z-transformation method, and the application of the inverse hyperbolic tangent transformation give comparable results when values of ρ are small to moderate (≤ 0.5). All methods provide good empirical coverage and balanced tail errors and are competitive with respect to interval width. overall the inverse hyperbolic tangent transformation method provides consistently good empirical coverage over the whole parameter space even with samples as small as $k = 50$ (Figure 4.1). Also, this method does not provide any major tail imbalance and is competitive in terms of interval width with the other methods (Figure 4.2-4.3). The simple asymptotic method, on the other hand, provides erratic coverage in small sample size ($k = 50$) and even for large sample sizes ($k = 100$, $k = 200$), obtained tails error are unbalanced. The poor performance of the simple asymptotic method may occur because it ignores the asymmetric sampling distribution of $\hat{\rho}$ while imposing symmetry on the confidence interval. The Fisher's Z-transformation method overall has a tendency to provide coverage short of the nominal level of 95% whereas the empirical coverage percentages for the F-distribution method are in excess of the nominal level. Also, confidence intervals obtained by F-distribution tend to be wide.

Recently, Ramasundarahettige *et al.* (2009) applied MOVER to construct a confidence interval for a difference of two correlated ICCs with constant class size. They compared a number of methods including; the simple asymptotic method; a method based on the F-distribution; a method based on the Z-transformation, application of inverse hyperbolic tangent transformation; and one based on the modified Fisher's Z-transformation. Simulation results show that major differences in the performance of aforementioned methods were seen with small sample sizes (k=15 and k= 50). In contrast to the results obtained in this thesis, they concluded that the procedure based on the F-distribution for single ICCs is preferable as it provides consistently good coverage even with samples as small as $k = 15$. Also, the method provides balanced tail errors and is competitive in terms of interval width. Not surprisingly, consistent results were obtained regarding the simple asymptotic method, where the simulation shows erratic behavior especially in small sample sizes. In other words, for some parameter combinations the coverage was overly conservative and for others overly liberal. In contrast to the results obtained in this thesis, they showed the procedure based on the Fisher's Z-transformation provides coverage in excess of 95%, whereas the method based on the inverse hyperbolic tangent transformation tends to fall below the nominal level.

Table 4.1: Performance of procedures for constructing two-sided 95% confidence interval (CI) for a difference between two correlated intraclass correlation coefficients with respect to empirical coverage based on 10000 runs and sample size, k=50

$k = 50$	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)†% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.1 0.1 0.0	94.19(2.87,2.94)0.91	94.63(2.67,2.70)0.87	95.16(2.41,2.43)0.88	96.27(1.78,1.95)1.17
0.1	93.96(2.97,3.07)0.90	94.32(2.82,2.86)0.87	95.09(2.41,2.50)0.87	96.16(1.94,1.90)1.17
0.3 0.1 0.0	94.25(2.71,3.04)0.87	94.34(2.49,3.17)0.84	94.98(2.05,2.97)0.85	96.13(1.78,2.09)1.10
0.1	94.56(2.59,2.85)0.87	94.81(2.30,2.89)0.84	95.43(1.90,2.67)0.85	96.41(1.54,2.05)1.10
0.3 0.0	94.50(2.75,2.75)0.84	94.43(2.77,2.80)0.81	95.15(2.37,2.48)0.82	96.30(1.81,1.89)1.02
0.1	94.10(2.98,2.92)0.84	94.21(2.97,2.82)0.81	94.79(2.67,2.54)0.82	96.25(1.88,1.87)1.02
0.3	94.86(2.58,2.56)0.83	94.90(2.50,2.60)0.80	95.45(2.25,2.30)0.81	96.35(1.76,1.89)1.01
0.5 0.1 0.0	94.82(2.53,2.65)0.81	94.68(2.28,3.04)0.78	95.41(1.76,2.83)0.80	96.10(1.61,2.29)1.01
0.1	94.32(2.74,2.94)0.81	94.29(2.43,3.28)0.78	94.78(2.12,3.10)0.79	95.73(1.91,2.36)1.01
0.3 0.0	94.45(2.50,3.05)0.78	94.06(2.62,3.32)0.75	94.70(2.23,3.07)0.77	95.65(1.81,2.54)0.92
0.1	94.68(2.62,2.70)0.78	94.36(2.58,3.06)0.75	95.08(2.27,2.65)0.77	96.18(1.85,1.97)0.92
0.3	95.65(2.03,2.32)0.77	95.46(1.96,2.58)0.74	96.06(1.72,2.22)0.76	96.91(1.36,1.73)0.92
0.5 0.0	94.92(2.67,2.41)0.71	93.69(3.29,3.02)0.69	94.69(2.78,2.53)0.71	95.92(2.18,1.90)0.82
0.1	94.93(2.33,2.74)0.71	93.92(2.94,3.14)0.69	94.82(2.45,2.73)0.71	95.95(1.87,2.18)0.82
0.3	95.36(2.46,2.18)0.70	94.41(2.98,2.61)0.68	95.11(2.64,2.25)0.70	96.39(1.96,1.65)0.81
0.5	96.38(1.87,1.75)0.69	95.43(2.33,2.24)0.67	96.16(1.94,1.90)0.69	96.75(1.70,1.55)0.80

Continued on next page

Table 4.1 – Continued from previous page

$k = 50$	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.7 0.3 0.0	93.98(2.66,3.36)0.69	93.59(2.72,3.69)0.67	94.49(2.47,3.04)0.69	95.19(2.03,2.78)0.81
0.1	93.94(2.39,3.67)0.69	93.47(2.45,4.08)0.67	94.56(2.16,3.28)0.68	95.23(1.90,2.87)0.81
0.3	94.84(2.08,3.08)0.69	94.77(2.14,3.09)0.67	95.42(1.90,2.68)0.68	95.74(1.73,2.53)0.81
0.5 0.0	94.64(2.19,3.17)0.62	93.32(2.98,3.70)0.60	94.46(2.58,2.96)0.62	95.35(2.02,2.63)0.70
0.1	94.79(1.96,3.25)0.61	93.52(2.80,3.68)0.60	94.50(2.36,3.14)0.62	95.43(1.84,2.73)0.69
0.3	95.32(1.89,2.79)0.61	94.02(2.61,3.37)0.59	95.10(2.25,2.65)0.62	95.67(1.93,2.40)0.69
0.5	95.53(1.51,2.96)0.60	94.52(2.05,3.43)0.59	95.48(1.78,2.74)0.61	96.06(1.46,2.48)0.68
0.7 0.0	96.04(1.88,2.08)0.50	93.39(3.22,3.39)0.49	94.87(2.46,2.67)0.52	95.35(2.25,2.40)0.55
0.1	95.95(2.19,1.86)0.50	93.46(3.49,3.05)0.49	94.74(2.78,2.48)0.52	95.48(2.41,2.11)0.55
0.3	96.28(1.77,1.95)0.50	93.64(3.04,3.32)0.49	95.09(2.33,2.58)0.52	95.72(2.01,2.27)0.55
0.5	**97.26**(1.53,1.21)0.49	94.83(2.74,2.43)0.48	96.23(1.97,1.80)0.51	96.47(1.82,1.71)0.54
0.7	**97.90**(1.09,1.01)0.48	96.01(2.08,1.91)0.48	96.88(1.60,1.52)0.50	**97.17**(1.46,1.37)0.54
0.9 0.5 0.0	93.19(1.82,4.99)0.52	93.56(2.71,3.73)0.50	94.55(2.49,2.96)0.52	94.67(2.39,2.94)0.58
0.1	93.29(1.80,4.91)0.52	93.74(2.68,3.58)0.50	94.75(2.44,2.81)0.52	94.57(2.30,3.13)0.58
0.3	93.49(1.70,4.81)0.52	93.86(2.66,3.48)0.50	94.83(2.40,2.77)0.52	94.99(2.18,2.83)0.58
0.5	93.82(1.60,4.58)0.51	94.24(2.54,3.22)0.50	95.11(2.32,2.57)0.52	95.41(1.94,2.65)0.58
0.7 0.0	94.78(0.84,4.38)0.38	93.86(2.67,3.47)0.37	95.12(2.13,2.75)0.39	95.48(1.76,2.76)0.41
0.1	94.05(1.06,4.89)0.38	93.10(3.04,3.86)0.37	94.75(2.52,2.73)0.39	95.06(2.14,2.80)0.41
0.3	94.65(0.82,4.53)0.38	93.90(2.59,3.51)0.37	95.14(2.14,2.72)0.39	95.24(1.89,2.87)0.41
0.5	94.40(0.98,4.62)0.38	93.51(2.94,3.55)0.37	94.93(2.35,2.72)0.39	95.49(1.94,2.57)0.41
0.7	95.30(0.61,4.09)0.37	94.98(2.05,2.97)0.36	96.21(1.60,2.19)0.38	96.25(1.36,2.39)0.40
0.9 0.0	**97.80**(0.96,1.24)0.20	**92.94**(3.37,3.69)0.20	94.90(2.39,2.71)0.22	95.38(2.02,2.60)0.21
0.1	**98.01**(0.96,1.03)0.20	93.18(3.44,3.38)0.20	95.27(2.43,2.30)0.22	95.37(2.38,2.25)0.21
0.3	**97.82**(1.10,1.08)0.20	**92.82**(3.40,3.78)0.20	94.87(2.51,2.62)0.22	95.50(2.16,2.34)0.21
0.5	**98.35**(0.77,0.88)0.19	94.11(2.84,3.05)0.20	95.77(1.91,2.32)0.21	96.00(1.80,2.20)0.21
0.7	**98.53**(0.77,0.70)0.19	94.88(2.63,2.49)0.19	96.25(1.87,1.88)0.21	96.48(1.90,1.62)0.20
0.9	**99.19**(0.37,0.44)0.19	96.29(1.63,2.08)0.19	**97.34**(1.15,1.51)0.21	**97.59**(1.05,1.36)0.20

†ML: the interval lies below the parameter; MR: the interval lies above the parameter

Note: Data generated with the mean family size of 3.12 and with the family size variance of 4.52

Table 4.2: Performance of procedures for constructing two-sided 95% confidence interval (CI) for a difference between two correlated intraclass correlation coefficients with respect to empirical coverage based on 10000 runs and sample size, k=100

$k = 100$	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)†% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.1 0.1 0.0	94.65(2.77,2.58)0.63	94.85(2.70,2.45)0.62	95.13(2.56,2.31)0.62	96.41(1.91,1.68)0.82
0.1	94.74(2.43,2.83)0.63	94.95(2.31,2.74)0.62	95.25(2.19,2.56)0.62	96.43(1.61,1.96)0.82
0.3 0.1 0.0	94.70(2.67,2.63)0.61	94.73(2.49,2.78)0.60	95.09(2.24,2.67)0.60	96.26(1.70,2.04)0.76
0.1	94.98(2.52,2.50)0.61	94.93(2.41,2.66)0.60	95.40(2.06,2.54)0.60	96.18(1.77,2.05)0.76
0.0	94.93(2.57,2.50)0.59	94.73(2.70,2.57)0.57	95.19(2.40,2.41)0.58	96.26(1.87,1.87)0.70
0.3 0.1	94.70(2.72,2.58)0.59	94.41(2.88,2.71)0.57	94.99(2.54,2.47)0.58	96.08(2.00,1.92)0.70
0.3	95.55(2.43,2.02)0.58	95.30(2.55,2.15)0.57	95.72(2.35,1.93)0.58	96.51(1.85,1.64)0.70
0.5 0.1 0.0	94.65(2.76,2.59)0.57	94.48(2.56,2.96)0.55	94.96(2.29,2.75)0.56	95.87(1.80,2.33)0.70
0.1	94.61(2.68,2.71)0.57	94.44(2.41,3.15)0.55	94.91(2.16,2.93)0.56	95.55(2.01,2.44)0.70
0.3 0.0	94.65(2.59,2.76)0.54	94.15(2.67,3.18)0.53	94.82(2.43,2.75)0.54	95.92(1.91,2.17)0.63
0.1	95.19(2.28,2.53)0.54	94.72(2.45,2.83)0.53	95.28(2.12,2.60)0.54	96.12(1.75,2.13)0.63
0.3	95.45(2.12,2.43)0.54	94.99(2.23,2.78)0.52	95.57(1.97,2.46)0.54	96.32(1.60,2.08)0.63
0.5 0.0	95.13(2.35,2.52)0.50	94.13(2.88,2.99)0.48	95.10(2.38,2.52)0.50	96.20(1.86,1.94)0.55
0.1	94.91(2.67,2.42)0.50	93.79(3.23,2.98)0.48	94.89(2.65,2.46)0.50	95.90(2.12,1.98)0.55
0.3	95.55(2.13,2.32)0.49	94.50(2.60,2.90)0.47	95.44(2.18,2.38)0.49	96.28(1.91,1.81)0.55
0.5	96.50(1.64,1.86)0.49	95.62(2.03,2.35)0.47	96.43(1.68,1.89)0.49	96.90(1.44,1.66)0.54

Continued on next page

Table 4.2 – Continued from previous page

$k = 100$	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.7 0.3 0.0	95.23(2.15,2.62)0.48	94.68(2.36,2.96)0.47	95.34(2.15,2.51)0.48	95.75(1.99,2.26)0.56
0.1	94.81(2.06,3.13)0.48	94.34(2.25,3.41)0.47	95.07(2.04,2.89)0.48	95.52(1.83,2.65)0.56
0.3	94.71(2.33,2.96)0.48	94.13(2.58,3.29)0.46	94.87(2.32,2.81)0.48	95.55(2.07,2.38)0.55
0.5 0.0	95.02(2.17,2.81)0.43	93.93(2.76,3.31)0.41	94.98(2.41,2.61)0.43	95.41(1.95,2.64)0.47
0.1	95.25(1.98,2.77)0.43	93.89(2.81,3.30)0.41	95.11(2.34,2.55)0.43	95.89(1.85,2.26)0.47
0.3	95.35(1.90,2.75)0.43	94.05(2.74,3.21)0.41	95.20(2.18,2.62)0.43	95.66(1.99,2.35)0.47
0.5	96.07(1.64,2.29)0.42	95.13(2.22,2.65)0.41	96.07(1.85,2.08)0.42	96.41(1.64,1.95)0.46
0.7 0.0	95.71(2.06,2.23)0.35	93.41(3.10,3.49)0.34	94.97(2.43,2.60)0.36	95.63(2.11,2.26)0.37
0.1	95.13(2.45,2.42)0.35	**92.94**(3.57,3.49)0.34	94.40(2.82,2.78)0.36	95.05(2.51,2.44)0.37
0.3	95.94(2.05,2.01)0.35	94.04(2.91,3.05)0.33	95.41(2.32,2.27)0.35	95.93(2.02,2.05)0.36
0.5	96.62(1.69,1.69)0.34	94.77(2.76,2.47)0.33	96.12(2.00,1.88)0.35	96.20(2.06,1.74)0.36
0.7	**97.45**(1.32,1.23)0.34	95.79(2.17,2.04)0.32	96.98(1.57,1.45)0.34	**97.07**(1.45,1.48)0.36
0.9 0.5 0.0	94.08(1.83,4.09)0.36	93.65(3.05,3.30)0.35	94.74(2.70,2.56)0.36	95.06(2.31,2.63)0.40
0.1	94.27(1.83,3.90)0.36	94.17(2.72,3.11)0.35	95.06(2.37,2.57)0.36	95.06(2.22,2.72)0.40
0.3	94.29(1.71,4.00)0.36	93.98(2.69,3.33)0.35	94.95(2.42,2.63)0.36	95.25(1.99,2.76)0.39
0.5	94.60(1.58,3.82)0.36	94.35(2.55,3.10)0.35	95.46(2.14,2.40)0.36	95.19(2.10,2.71)0.39
0.7 0.0	94.64(1.50,3.86)0.26	93.58(2.93,3.49)0.25	95.15(2.25,2.60)0.27	95.05(2.29,2.66)0.27
0.1	94.35(1.23,4.42)0.26	93.19(2.93,3.88)0.25	94.82(2.31,2.87)0.27	95.10(1.98,2.92)0.27
0.3	94.64(1.24,4.12)0.26	93.74(2.78,3.48)0.25	95.24(2.23,2.53)0.27	95.44(1.93,2.63)0.27
0.5	95.41(0.99,3.60)0.26	94.26(2.55,3.19)0.25	95.90(1.81,2.29)0.26	95.88(1.72,2.40)0.27
0.7	95.83(0.88,3.29)0.26	94.75(2.38,2.87)0.25	95.86(1.96,2.18)0.26	96.04(1.73,2.23)0.27
0.9 0.0	96.50(1.79,1.71)0.13	93.08(3.53,3.39)0.13	95.03(2.55,2.42)0.14	95.16(2.53,2.31)0.13
0.1	96.55(1.80,1.65)0.14	**92.98**(3.35,3.67)0.13	94.97(2.53,2.50)0.14	95.14(2.39,2.47)0.14
0.3	96.87(1.50,1.63)0.13	93.60(3.28,3.12)0.13	95.42(2.23,2.35)0.14	95.56(2.23,2.21)0.13
0.5	**97.06**(1.46,1.48)0.13	93.97(3.04,2.99)0.13	95.72(2.11,2.17)0.14	95.87(2.04,2.09)0.13
0.7	**97.52**(1.13,1.35)0.13	94.73(2.54,2.73)0.13	96.34(1.74,1.92)0.14	96.67(1.66,1.67)0.13
0.9	**98.49**(0.81,0.70)0.13	96.39(1.86,1.75)0.12	**97.59**(1.31,1.10)0.13	**97.75**(1.21,1.04)0.13

†ML: the interval lies below the parameter; MR: the interval lies above the parameter

Note: Data generated with the mean family size of 3.12 and with the family size variance of 4.52

Table 4.3: Performance of procedures for constructing two-sided 95% confidence interval (CI) for a difference between two correlated intraclass correlation coefficients with respect to empirical coverage based on 10000 runs and sample size, k=200

$k = 200$	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)†% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.1 0.1 0.0	95.15(2.30,2.55)0.44	95.22(2.24,2.54)0.44	95.33(2.23,2.44)0.44	96.27(1.79,1.94)0.57
0.1	94.98(2.39,2.63)0.44	95.05(2.36,2.59)0.44	95.18(2.32,2.50)0.44	96.27(1.67,2.06)0.57
0.3 0.1 0.0	95.06(2.51,2.43)0.43	94.92(2.42,2.66)0.42	95.21(2.26,2.53)0.43	96.15(1.97,1.88)0.53
0.1	94.91(2.54,2.55)0.43	94.79(2.46,2.75)0.42	95.16(2.24,2.60)0.43	96.14(1.75,2.11)0.53
0.3 0.0	95.12(2.37,2.51)0.41	94.75(2.57,2.68)0.40	95.27(2.29,2.44)0.41	96.03(1.97,2.00)0.49
0.1	95.16(2.43,2.41)0.41	94.78(2.60,2.62)0.40	95.27(2.40,2.33)0.41	96.32(1.88,1.80)0.49
0.3	95.41(2.37,2.22)0.41	95.15(2.48,2.37)0.40	95.57(2.29,2.14)0.41	96.65(1.75,1.60)0.49
0.5 0.1 0.0	95.02(2.58,2.40)0.40	94.69(2.48,2.83)0.39	95.28(2.12,2.60)0.40	95.93(1.89,2.18)0.49
0.1	94.98(2.53,2.49)0.40	94.63(2.42,2.95)0.39	95.11(2.20,2.69)0.40	95.94(1.78,2.28)0.49
0.3 0.0	94.80(2.63,2.57)0.38	94.20(2.76,3.04)0.37	94.86(2.52,2.62)0.38	95.86(1.96,2.18)0.44
0.1	94.76(2.74,2.50)0.38	94.16(2.92,2.92)0.37	94.82(2.62,2.56)0.38	96.05(2.02,1.93)0.44
0.3	95.49(2.25,2.26)0.38	95.04(2.45,2.51)0.37	95.57(2.16,2.27)0.38	96.22(1.78,2.00)0.44
0.5 0.0	95.07(2.47,2.46)0.35	94.09(2.98,2.93)0.34	95.03(2.49,2.48)0.35	95.86(2.01,2.13)0.38
0.1	94.89(2.55,2.56)0.35	93.79(3.09,3.12)0.34	94.82(2.58,2.60)0.35	95.65(2.12,2.23)0.38
0.3	95.39(2.54,2.07)0.35	94.57(2.91,2.52)0.33	95.35(2.57,2.08)0.35	96.10(2.20,1.70)0.38
0.5	96.46(1.82,1.72)0.34	95.59(2.23,2.18)0.33	96.45(1.80,1.75)0.34	96.96(1.64,1.40)0.38

Continued on next page

Table 4.3 – Continued from previous page

k = 200	Simple asymptotic	Fisher's Z	Inverse tanh	F-distribution
ρ_b ρ_s ρ_{bs}	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width	Cover(ML,MR)% CI Width
0.7 0.3 0.0	94.50(2.42,3.08)0.34	93.86(2.73,3.41)0.33	94.68(2.35,2.97)0.34	95.24(2.11,2.65)0.39
0.1	95.38(2.08,2.54)0.34	94.56(2.45,2.99)0.33	95.48(2.10,2.42)0.34	95.69(2.06,2.25)0.39
0.3	95.11(2.08,2.81)0.34	94.70(2.29,3.01)0.33	95.32(2.05,2.63)0.34	95.81(1.78,2.41)0.39
0.5 0.0	94.83(2.30,2.87)0.30	93.73(2.95,3.32)0.29	94.79(2.52,2.69)0.30	95.55(1.97,2.48)0.32
0.1	95.08(1.80,3.12)0.30	93.80(2.52,3.68)0.29	94.87(2.09,3.04)0.30	95.58(1.89,2.53)0.32
0.3	95.80(2.00,2.20)0.30	94.78(2.58,2.64)0.29	95.83(2.16,2.01)0.30	96.23(2.01,1.76)0.32
0.5	96.09(1.73,2.18)0.30	94.83(2.46,2.71)0.28	95.99(1.96,2.05)0.30	96.14(1.68,2.18)0.32
0.7 0.0	95.21(2.50,2.29)0.24	93.38(3.36,3.26)0.23	94.80(2.71,2.49)0.25	95.44(2.34,2.22)0.25
0.1	95.23(2.31,2.46)0.24	93.69(3.10,3.21)0.23	94.93(2.44,2.63)0.25	95.66(2.15,2.19)0.25
0.3	95.77(2.18,2.05)0.24	94.08(3.09,2.83)0.23	95.49(2.32,2.19)0.24	95.88(2.14,1.98)0.25
0.5	96.43(1.86,1.71)0.24	94.87(2.62,2.51)0.23	96.04(2.08,1.88)0.24	96.69(1.67,1.64)0.25
0.7	**97.40**(1.28,1.32)0.23	96.04(1.97,1.99)0.22	**97.13**(1.40,1.47)0.24	**97.44**(1.32,1.24)0.24
0.9 0.5 0.0	94.63(1.86,3.51)0.26	93.86(2.89,3.25)0.25	94.90(2.49,2.61)0.26	95.11(2.29,2.60)0.28
0.1	94.76(1.91,3.33)0.26	93.95(2.95,3.10)0.24	95.00(2.57,2.43)0.26	94.86(2.38,2.76)0.28
0.3	94.81(2.05,3.14)0.26	94.12(3.00,2.88)0.24	95.01(2.67,2.32)0.26	94.92(2.61,2.47)0.27
0.5	95.11(1.76,3.13)0.25	94.65(2.51,2.84)0.24	95.52(2.26,2.22)0.25	95.50(2.02,2.48)0.27
0.7 0.0	94.90(1.49,3.61)0.18	93.46(3.10,3.44)0.17	95.15(2.40,2.45)0.19	95.18(2.29,2.53)0.19
0.1	94.87(1.66,3.47)0.18	93.41(3.21,3.38)0.17	94.78(2.57,2.65)0.19	95.11(2.33,2.56)0.19
0.3	95.17(1.44,3.39)0.18	93.63(3.07,3.30)0.17	95.23(2.33,2.44)0.19	95.16(2.38,2.46)0.19
0.5	95.56(1.33,3.11)0.18	94.22(2.79,2.99)0.17	95.70(2.14,2.16)0.18	95.97(1.90,2.13)0.18
0.7	95.98(1.14,2.88)0.18	94.80(2.50,2.70)0.17	96.07(1.85,2.08)0.18	96.21(1.86,1.93)0.18
0.9 0.0	95.20(2.29,2.51)0.09	92.51(3.61,3.88)0.09	94.45(2.64,2.91)0.10	94.87(2.52,2.61)0.09
0.1	95.80(1.99,2.21)0.09	92.93(3.52,3.55)0.09	94.89(2.52,2.59)0.10	95.14(2.48,2.38)0.09
0.3	96.33(1.87,1.80)0.09	93.44(3.43,3.13)0.09	95.51(2.31,2.18)0.10	95.53(2.31,2.16)0.09
0.5	96.75(1.55,1.70)0.09	93.94(3.05,3.01)0.09	95.82(2.02,2.16)0.09	96.18(1.90,1.92)0.09
0.7	**97.45**(1.36,1.19)0.09	95.61(2.28,2.11)0.09	96.88(1.65,1.47)0.09	**97.05**(1.59,1.36)0.09
0.9	**98.25**(0.85,0.90)0.09	96.20(1.69,2.11)0.08	**97.86**(1.08,1.06)0.09	**97.92**(0.96,1.12)0.09

†ML: the interval lies below the parameter; MR: the interval lies above the parameter

Note: Data generated with the mean family size of 3.12 and with the family size variance of 4.52

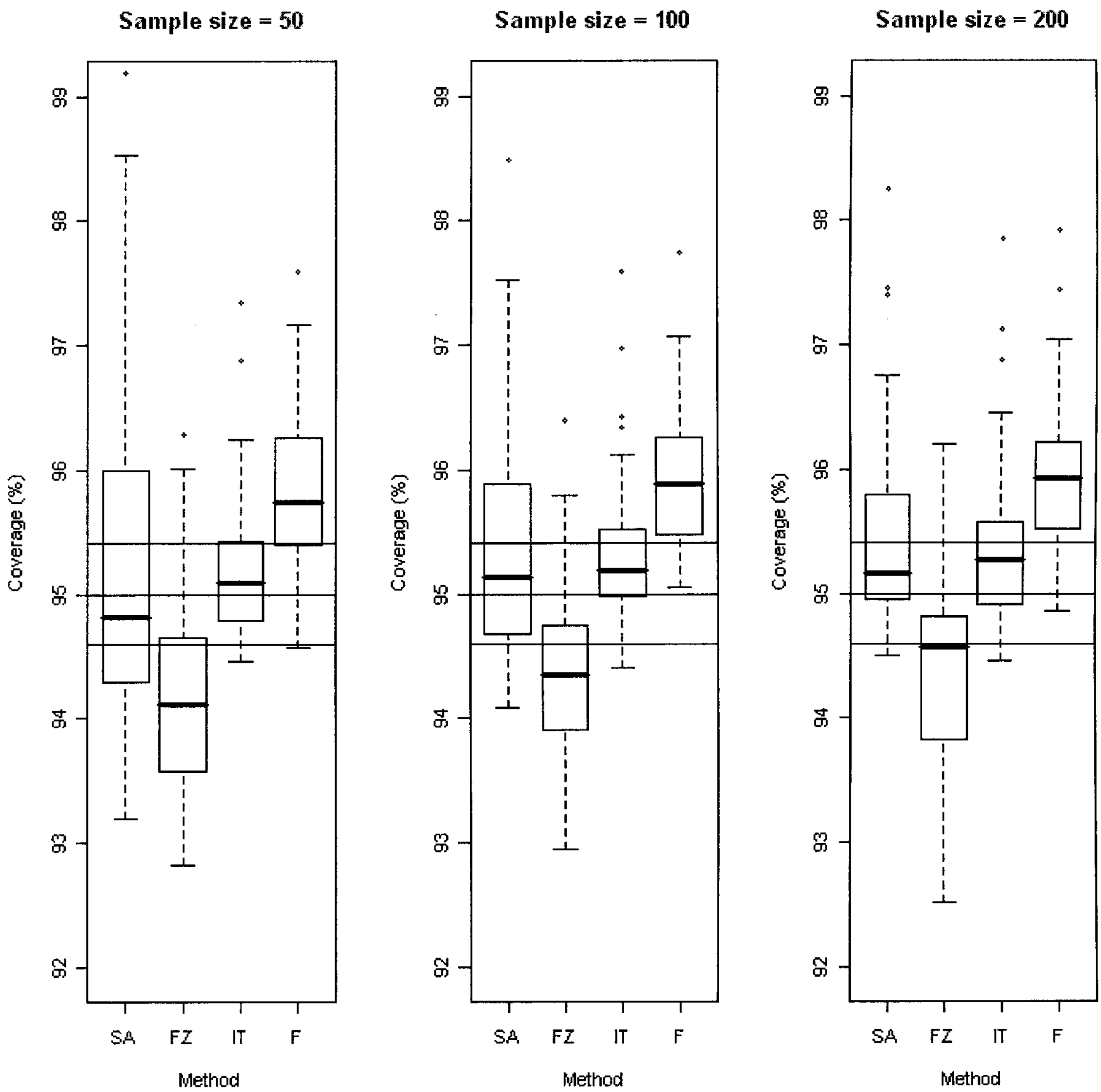

Figure 4.1: Performance of MOVER with respect to empirical coverage for two-sided 95% confidence interval for a difference between two correlated ICCs using four different methods: simple asymptotic method (SA), method based on the Fisher's Z-transformation (FZ), method based on the inverse hyperbolic tangent transformation (IT), method based on the F-distribution (F). There are 43 data points per sample size.

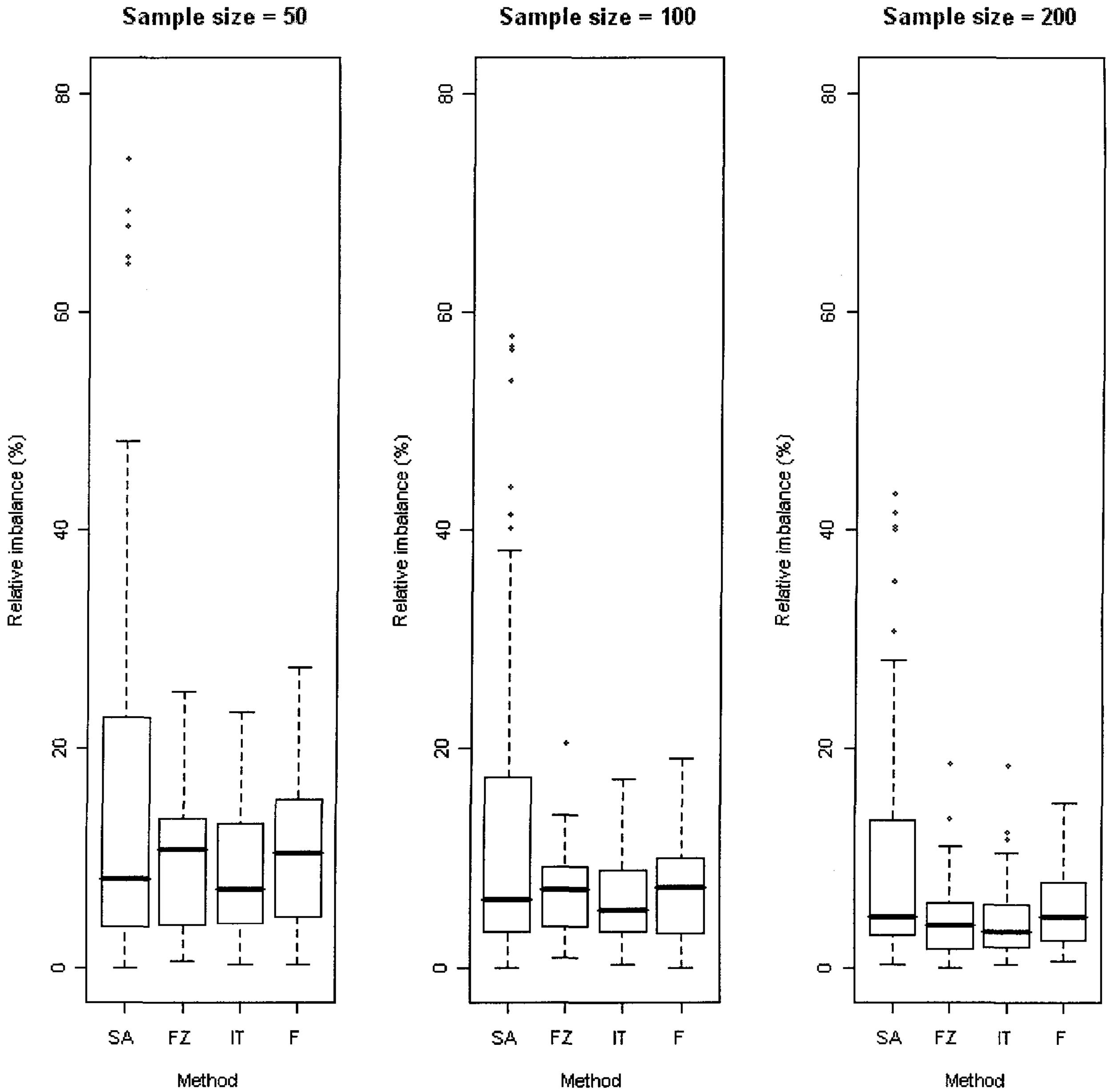

Figure 4.2: Performance of MOVER with respect to tail errors for two-sided 95% confidence interval for a difference between two correlated ICCs using four different methods: simple asymptotic method, method based on the Fisher's Z-transformation (FZ), method based on the inverse hyperbolic tangent transformation (IT), method based on the F-distribution (F). There are 43 data points per sample size.

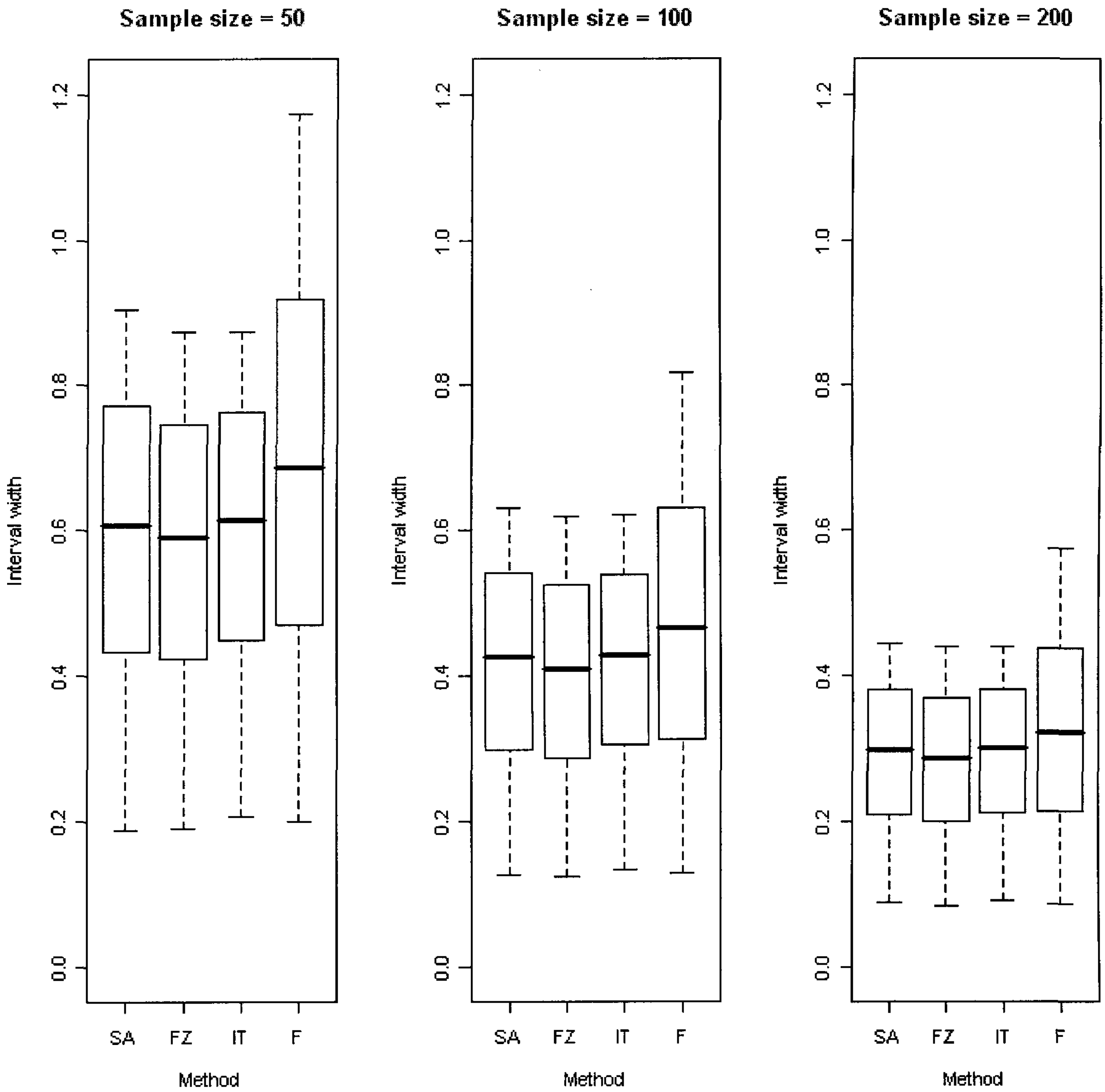

Figure 4.3: Performance of MOVER with respect to interval width for two-sided 95% confidence interval for a difference between two correlated ICCs using four different methods: simple asymptotic method (SA), method based on the Fisher's Z-transformation (FZ), method based on the inverse hyperbolic tangent transformation (IT), method based on the F-distribution (F). There are 43 data points per sample size.

Chapter 5

EXAMPLE - THE GALTON DATA ON SIBLINGS HEIGHTS

5.1 The data source

We now use Galton's 1886 data set to illustrate the procedures. Hanley (2004*a*) detailed the historical background of the data and how he brought the data to light from Galton's original notebook. The data set is available at Hanley's web site (Hanley, 2004*b*) with the appropriate corrections. Recently, the data were analyzed by Naik and Helu (2007) to illustrate testing two independent ICCs. Interest focused on comparing the ICC between boys in a family with that between girls in another family. In other words, the first population consisted of families for which only boys were considered, the second consisted of families for which only girls were considered. Such division lead to loss of essential information, since the data set is only partially analyzed. Moreover, in the comparison of sex-specific ICCs, it would be more informative to compare the coefficients of resemblance obtained from the same families by taking into account the dependency of observations among siblings of the opposite sex. Such comparisons lead to the comparison of two correlated ICCs.

The data on human stature consist of 205 families where the number of children ranges from 1 to 15. The entire data set contains 963 children of which 486 are sons and 476 are daughters. Among the 963 children, some of them are described verbally as: "tallish", "middle", "deformed", etc., others have numerical values recorded as "about x.0 inches". Thus, among the 963 children, only 934 from the 205 families have numeric values.The total number of families having at least one son and one daughter is 150. Also, there are 179 families with at lest one son and 176 families with at least one daughter. Based on the 934 observations, the average family size and the sample variance of family size are 4.6 and 7.4, respectively. The summary statistics for heights of brothers and sisters are provided in Table 5.1. Table 5.1 shows that when estimating ICC, sisters and brothers should be treated as separate groups, as the sample means and samples variances in these two groups are different.

The brother-brother correlation ($\hat{\rho}_b$) and the sister-sister correlation ($\hat{\rho}_s$) with respect to stature are estimated by the ANOVA estimator given by the equation (2.3) in Section 2.2. Thus, for brothers, based on 179 families, $\text{MSE}_b = 4.242$, $\text{MSA}_b = 11.367$, $n_{b0} = 2.682$ and

$$\begin{aligned}\hat{\rho}_b &= \frac{\text{MSA}_b - \text{MSE}_b}{\text{MSA}_b + (n_{b0} - 1)\text{MSE}_b} \\ &= \frac{11.367 - 4.242}{11.367 + (2.682 - 1)4.242} \\ &= 0.385.\end{aligned}$$

Similarly, for sisters, based on 176 families, $\text{MSE}_s = 3.222$, $\text{MSA}_s = 9.238$, $n_{s0} = 2.567$, and $\hat{\rho}_s = 0.421$.

Table 5.1: Summary of Galton's 1886 data set on siblings heights

	Number of children of the same sex	Average height† (inches)	Sample variance of height† (inches)	$\hat{\rho}$
Brothers	481	69.233	6.898	0.385
Sisters	453	64.103	5.565	0.421

†Average height and the sample variance were calculated based on the ANOVA table

5.2 *Confidence intervals for sex-specific intraclass correlation coefficients*

To compare two sex-specific ICCs, the confidence interval for $\rho_b - \rho_s$ is computed using MOVER incorporating four different methods for sex-specific ICCs including: the simple asymptotic method, the method based on Fisher's Z-transformation, the inverse hyperbolic tangent transformation method, and the method based on the F-distribution. The confidence interval estimates of ρ_b, ρ_s and $\rho_b - \rho_s$ using the four methods are summarized in Table 5.2.

A 95% confidence interval for each ρ_b and ρ_s is obtained using the simple asymptotic method as described in Section 2.3. For brothers, $\rho_b = 0.385$, $\widehat{\text{var}}(\hat{\rho}_b) = 0.003$

$$\begin{aligned} l_b &= \hat{\rho}_b - z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{\rho}_b)} \\ &= 0.385 - 1.96\sqrt{0.003} \\ &= 0.278, \\ u_b &= \hat{\rho}_b + z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{\rho}_b)} \\ &= 0.385 + 1.96\sqrt{0.003} \\ &= 0.492 \end{aligned}$$

and similarly for sisters, $\rho_b = 0.421$, $\widehat{\text{var}}(\hat{\rho}) = 0.003$

$$(l_s, u_s) = (0.314,\ 0.528).$$

A 95% confidence interval for each ρ_b and ρ_s obtained using method based on the Fisher's Z-transformation as explained in Section 2.3. For brothers, $n_{0b} = 2.682$,

$\hat{\rho}_b = 0.385$, $N_b = 481$, $k_b = 179$. Then,

$$\begin{aligned} Z_{F_b} &= \frac{1}{2}\ln\left[\frac{1+(n_{0b}-1)\hat{\rho}_b)}{1-\hat{\rho}_b}\right] \\ &= \frac{1}{2}\ln\left[\frac{1+(2.682-1)0.385}{1-0.385}\right] \\ &= 0.493, \\ \text{var}(Z_{F_b}) &= \frac{1}{2}[(k_b-1)^{-1}+(N_b-k_b)^{-1}] \\ &= \frac{1}{2}[(179-1)^{-1}+(481-179)^{-1}], \\ &= 0.004. \end{aligned}$$

A 95% confidence interval for ρ_b on the transformed scale is

$$(Z_l, Z_u) = \left(Z_{F_b} - z_{\alpha/2}\sqrt{\text{var}(Z_{F_b})},\ Z_{F_b} + z_{\alpha/2}\sqrt{\text{var}(Z_{F_b})}\right)$$

$$(Z_l, Z_u) = (0.369, 0.617).$$

Then, transforming the limits back to the original scale result in

$$\begin{aligned} l_b &= \frac{\exp(2Z_l)-1}{\exp(2Z_l)+(n_{bo}-1)} \\ &= \frac{\exp(2\times 0.369)-1}{\exp(2\times 0.369)+(2.682-1)} \\ &= 0.289, \end{aligned}$$

$$\begin{aligned} u_b &= \frac{\exp(2\mathrm{Z_u})-1}{\exp(2Z_u)+(n_{bo}-1)} \\ &= \frac{\exp(2\times 0.617)-1}{\exp(2\times 0.617)+(2.682-1)} \\ &= 0.476. \end{aligned}$$

Similarly for sisters,

$$
\begin{aligned}
Z_{F_s} &= 0.527 \\
\text{var}(Z_{F_s}) &= 0.005, \\
(Z_l, Z_u) &= (0.388,\ 0.666), \\
(l_s, u_s) &= (0.314,\ 0.521).
\end{aligned}
$$

A 95% confidence interval for each ρ_b and ρ_s is obtained using the inverse hyperbolic tangent transformation method as described in Section 2.3. For brothers, $\hat{\rho}_b = 0.385$, $\widehat{\text{var}}(\hat{\rho}_b) = 0.003$

$$
\begin{aligned}
Z_b &= \frac{1}{2}\ln\left[\frac{1+\hat{\rho}_b}{1-\hat{\rho}_b}\right] \\
&= \frac{1}{2}\ln\left[\frac{1+0.385}{1-0.385}\right] \\
&= 0.406, \\
\widehat{\text{var}}(Z_b) &= \frac{\widehat{\text{var}}(\hat{\rho}_b)}{[(1+\hat{\rho}_b)(1-\hat{\rho}_b)]^2} \\
&= \frac{0.003}{[(1+0.385)(1-0.385)]^2} \\
&= 0.004.
\end{aligned}
$$

Then, a 95% confidence interval for ρ_b on the transformed scale is

$$
(Z_l, Z_u) = \left(Z_b - z_{\alpha/2}\sqrt{\widehat{\text{var}}(Z_b)},\ Z_b + z_{\alpha/2}\sqrt{\widehat{\text{var}}(Z_b)}\right)
$$

$$
(Z_l, Z_u) = (0.282,\ 0.530).
$$

Then, transforming the limits to the original scale result in

$$\begin{aligned} l_b &= \frac{\exp(2Z_l) - 1}{\exp(2Z_l) + 1} \\ &= \frac{\exp(2 \times 0.282) - 1}{\exp(2 \times 0.282) + 1} \\ &= 0.275, \\ u_b &= \frac{\exp(2Z_u) - 1}{\exp(2Z_u) + 1} \\ &= \frac{\exp(2 \times 0.530) - 1}{\exp(2 \times 0.530) + 1} \\ &= 0.485. \end{aligned}$$

Similarly for sisters,

$$\begin{aligned} Z_s &= 0.449, \\ \widehat{\text{var}}(Z_s) &= 0.004, \\ (Z_l, Z_u) &= (0.325,\ 0.573) \\ (l_s, u_s) &= (0.314,\ 0.518). \end{aligned}$$

A 95% confidence interval for each ρ_b and ρ_s based on the F -distribution method due to Thomas and Hultquist (1978) as described in Section 2.3. For brothers,

$$\begin{aligned} F_{L_b} &= F_{(\alpha/2, 179-1, \Sigma_{i=1}^{k_b} b_i - k_b)} \\ &= F_{(0.05/2, 179-1, 481-179)} \\ &= 0.766, \end{aligned}$$

$$F_{U_b} = F_{(1-\alpha/2,179-1,\Sigma_{i=1}^{k_b} b_i - k_b)}$$
$$= F_{(1-0.05/2,179-1,481-179)}$$
$$= 1.294,$$
$$F_b^* = 2.254,$$
$$\hat{n}_b = 1.968.$$

When the values for F_b^*, F_{L_b}, F_{U_b}, $\hat{n}_b$ are substituted into the formula (2.12) in Section 2.3, a 95% confidence interval for ρ_b is produced

$$l_b = \frac{F_b^*/F_{U_b} - 1}{\hat{n}_b + F_b^*/F_{U_b} - 1}$$
$$= \frac{2.254/1.294 - 1}{1.968 + 2.254/1.294 - 1}$$
$$= 0.274,$$
$$u_b = \frac{F_b^*/F_{L_b} - 1}{\hat{n}_b + F_b^*/F_{L_b} - 1}$$
$$= \frac{2.254/0.766 - 1}{1.968 + 2.254/0.766 - 1}$$
$$= 0.497.$$

Similarly, for sisters

$$F_{L_s} = 0.762,$$
$$F_{U_s} = 1.303,$$
$$F_s^* = 2.284,$$
$$\hat{n}_s = 1.753,$$
$$(l_s,\ u_s) = (0.300,\ 0.533).$$

5.3 Confidence intervals for a difference between two sex-specific intraclass correlation coefficients

The 95% confidence limits for $\rho_b - \rho_s$ are obtained using the equations (3.4.6) and equation (3.4.7) derived in Section 3.4 are given as

$$L = (\hat{\rho}_b - \hat{\rho}_s) - \sqrt{(\hat{\rho}_b - l_b)^2 + (u_s - \hat{\rho}_s)^2 - 2\widehat{\text{corr}}(\hat{\rho}_b, \hat{\rho}_s)(\hat{\rho}_b - l_b)(u_s - \hat{\rho}_s)}, \quad (5.3.1)$$

$$U = (\hat{\rho}_b - \hat{\rho}_s) + \sqrt{(u_b - \hat{\rho}_b)^2 + (\hat{\rho}_s - l_s)^2 - 2\widehat{\text{corr}}(\hat{\rho}_b, \hat{\rho}_s)(u_b - \hat{\rho}_b)(\hat{\rho}_s - l_s)}. \quad (5.3.2)$$

The estimated interclass correlation, $\hat{\rho}_{bs}$, is 0.264 (based on 150 families who have at least one son and one daughter). Now, the $\text{corr}(\hat{\rho}_b, \hat{\rho}_s)$ term in equation (5.3.1) and equation (5.3.2) can be estimated using the formula (3.4.5) in Section 3.4

$$\begin{aligned}
\widehat{\text{corr}}(\hat{\rho}_b, \hat{\rho}_s) &= \frac{\rho_{bs}^2 \left[\hat{n}_b \hat{n}_s (\hat{n}_b - 1)(\hat{n}_s - 1)\right]^{1/2}}{[1 + (\hat{n}_b - 1)\rho_b][1 + (\hat{n}_s - 1)\rho_s]} \\
&= \frac{0.264^2 \left[1.968 \times 1.753(1.968 - 1)(1.753 - 1)\right]^{1/2}}{[1 + (1.968 - 1)0.385][1 + (1.753 - 1)0.421]} \\
&= 0.061.
\end{aligned}$$

A 95% confidence interval for $\rho_b - \rho_s$ using the MOVER and incorporating confidence intervals for ρ_b and ρ_s obtained by the simple asymptotic method. The 95% confidence interval for separate ρ_b and ρ_s using the simple asymptotic method are as given in Section 5.2,

$$(l_b, u_b) = (0.278,\ 0.492),$$

$$(l_s, u_s) = (0.314,\ 0.528).$$

By substituting $\hat{\rho}_b$, $\hat{\rho}_s$, l_b, u_b, l_s, u_s into the formulae (5.3.1) and (5.3.2), we obtained 95% confidence limits for $\rho_b - \rho_s$ as shown below

$$\begin{aligned}
L &= (\hat{\rho}_b - \hat{\rho}_s) - \sqrt{(\hat{\rho}_b - l_b)^2 + (u_s - \hat{\rho}_s)^2 - 2\widehat{\text{corr}}(\hat{\rho}_b, \hat{\rho}_s)(\hat{\rho}_b - l_b)(u_s - \hat{\rho}_s)} \\
&= -0.036 - \sqrt{(0.107)^2 + (0.107)^2 - 2 \times 0.061(0.107)(0.107)} \\
&= -0.183, \\
U &= (\hat{\rho}_b - \hat{\rho}_s) + \sqrt{(u_b - \hat{\rho}_b)^2 + (\hat{\rho}_s - l_s)^2 - 2\widehat{\text{corr}}(\hat{\rho}_b, \hat{\rho}_s)(u_b - \hat{\rho}_b)(\hat{\rho}_s - l_s)} \\
&= -0.036 + \sqrt{(0.107)^2 + (0.107)^2 - 2 \times 0.061(0.107)(0.107)} \\
&= 0.111.
\end{aligned}$$

A second 95% confidence interval for $\rho_b - \rho_s$ using the MOVER incorporating confidence intervals of ρ_b and ρ_s is based on Fisher's Z-transformation. The 95% confidence intervals for ρ_b and ρ_s using Fisher's Z-transformation are as given in Section 5.2,

$$(l_b, u_b) = (0.289,\ 0.476),$$

$$(l_s, u_s) = (0.314,\ 0.521)$$

and the corresponding 95% confidence limits for $\rho_b - \rho_s$ are

$$(L, U) = (-0.170,\ 0.100).$$

For a 95% confidence interval for $\rho_b - \rho_s$ using MOVER incorporating the inverse hyperbolic tangent transformation, the 95% confidence interval for ρ_b and ρ_s based on inverse hyperbolic tangent transformation are as given in Section 5.2,

$$(l_b, u_b) = (0.275,\ 0.485),$$

$$(l_s, u_s) = (0.314,\ 0.518)$$

and the corresponding 95% confidence limits for $\rho_b - \rho_s$ are

$$(L, U) = (-0.178,\ 0.101).$$

Lastly, for a 95% confidence interval for $\rho_b - \rho_s$ using the MOVER based on the F- distribution method, the 95% confidence intervals for ρ_b and ρ_s are as given in Section 5.2:

$$(l_b, u_b) = (0.274,\ 0.497),$$

$$(l_s, u_s) = (0.300,\ 0.533)$$

and 95% confidence limits for $\rho_b - \rho_s$ are

$$(L, U) = (-0.188,\ 0.124).$$

Table 5.2: 95% two sided confidence intervals (CIs) for $\rho_b - \rho_s$ using four different methods of confidence interval estimation for sex-specific ICCs

Method	95% CI about ρ_b	95% CI about ρ_s	95% CI about $(\rho_b - \rho_s)$
	$\hat{\rho}_b = 0.385$	$\hat{\rho}_s = 0.421$	$\hat{\rho}_b - \hat{\rho}_s = -0.036$
Simple asymptotic	(0.278, 0.492)	(0.314, 0.528)	(−0.183, 0.111)
Fisher's Z	(0.289, 0.476)	(0.314, 0.521)	(−0.170, 0.100)
Inverse hyperbolic tangent	(0.275, 0.485)	(0.314, 0.518)	(−0.178, 0.101)
F-distribution	(0.274, 0.497)	(0.300, 0.533)	(−0.188, 0.124)

5.4 Summary

Galton's 1886 data consist of 963 children, 934 of which have numeric values (29 children were described verbally and were omitted from the analysis). Among the 934 children, there are 481 brothers and 453 sisters and 205 families. The estimated brother-brother correlation and sister-sister correlation (and their variances) are given respectively by 0.385(0.003) for brothers and 0.421(0.003) for sisters. The interclass correlation coefficient is 0.264. Furthermore, 95% confidence intervals for ρ_b and ρ_s were obtained using four different methods as they were used in the process of estimating a 95% confidence interval for $\rho_b - \rho_s$. The estimated difference between two sex-specific intraclass correlations, $\hat{\rho}_b - \hat{\rho}_s$ is -0.036. Based on the results provided in Table 5.2, there is no significant difference between the sex-specific intraclass correlations coefficients at the $\alpha = 5\%$ level regardless of the method used, as the resulting confidence intervals for $\rho_b - \rho_s$ all contained zero. Similar results were obtained by Naik and Helu (2007), where no significant difference was found between independent sex-specific intraclass correlations by means of significant testing. Even though no significant difference was found between the two sex-specific intraclass correlation coefficients, the proposed confidence intervals still showed substantial variability. Relying on hypothesis testing alone, this information is discarded and the direction and magnitude of the underlying difference ignored.

Chapter 6

CONCLUSIONS

This thesis has focused on confidence interval estimation for a difference between two correlated ICCs assuming an unbalanced one-way random effects model. Even though an hypothesis testing procedure for the equality of two correlated ICCs has been developed (Donner *et al.*, 1984), confidence intervals are considered more informative as they indicate statistical significance as well as the magnitude of the effect for practical importance.

Limited methodology exist for constructing a confidence interval for a difference of two correlated ICCs when class sizes are variable. This thesis aimed at filling that gap. The approach taken is to recover variance estimates needed for the difference from confidence limits for single ICCs. The advantage of this approach is that is can take into account asymmetric properties of $\hat{\rho}$, and in contrast to the traditional method, does not enforce symmetry on the interval. The method has been referred to as the MOVER, method of variance estimates recovery (Zou, 2008). Another advantage of the MOVER is that it can be used when the raw data are not available, given the summary in terms of point estimators of ICCs and their respective confidence limits.

The good performance of the MOVER as applied to a confidence interval for a difference of two correlated ICCs in the case of constant class size has been demonstrated by Ramasundarahettige *et al.* (2009). The MOVER with limits for single ICCs based on F-distribution was recommended as its performance even in small sample size was shown to be good.

The major finding of this thesis is that the MOVER method implementing limits for single ICCs obtained by application of inverse hyperbolic tangent transformation performs consistently well even in small sample sizes. The traditional method did not provide satisfactory results in small sample sizes where the assumption of normality may not hold. These results are consistent with those obtained by Ramasundarahettige *et al.* (2009) where they showed that, although the performance of the simple asymptotic method in terms of empirical coverage was good as sample increased, its performance in small sample sizes was not acceptable. The poor performance of the simple asymptotic method in small sample size occurs because this method has difficulty adjusting for the asymmetric sampling distribution of $\hat{\rho}$, which is a function of the parameter.

It must be emphasized that conclusions obtained in this thesis will only be valid under several assumptions. First, the distribution of class sizes was obtained from families typically seen in United States in 1950. In the cases of greater imbalance, the conclusions of this thesis may not apply and further investigation is required for evaluating the proposed method. Second, the MOVER takes its validity from that of confidence limits for single ICCs. In this thesis, confidence intervals for single ICCs relied on the assumption of approximate normality; therefore in situations when this assumption may not hold, other procedures should be sought. One possible solution is to first transform the data using a power transformation (Box and Cox, 1964) and then obtain the desired confidence interval using the proposed method. However, due to the transformation the confidence intervals on the transformed scale may not be directly interpretable. Another possibility is to use the bootstrap. This procedure has been widely used with non-normal data. For example, Ukoumunne *et al.* (2003) implemented a non-parametric bootstrap confidence interval for ICC with constant class size in the context of cluster randomization trials. They showed that the application of bootstrap-t method to the Fisher's Z-transformation of ρ performs good even in small sample sizes. Evaluation of the bootstrap for the difference of two correlated ICCs, however, is beyond the scope of this thesis.

The proposed method is not only limited to ICCs obtained under unbalanced one-way random effects model but it can be extended to models where one is able to adjust for other source of variability in the data. For instance, Stanish and Taylor (1983) considered estimation of ICC under the covariance model where one is able to adjust for a potential confounding variable. Other models can also be considered including mixed effects models and random effects models. Confidence intervals for ICCs under these models are given in Harville and Fenech (1985).

BIBLIOGRAPHY

Altman, D. G. (2005). Why we need confidence intervals? *World Journal of Surgery* **29**, 554–556.

Bhandary, M. and Alam, M. K. (2000). Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics: Theory and Methods* **29**, 755–768.

Bhandary, M. and Fujiwara, K. (2006). A small sample test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics: Simulation and Computation* **35**, 765–778.

Bhandary, M. and Fujiwara, K. (2008). Confidence interval estimation for a linear contrast in intraclass correlation coefficients under unequal family sizes for several populations. *Journal of Statistical Computation and Simulation* **78**, 609–621.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J.R. Stat. Soc. B* **26**, 211–252.

Brass, W. (1958). Models of birth distribution in human populations. *Bulletin of International Statistical Institute* **36**, 165–179.

Donner, A. (1979). The use of correlation and regression in the analysis of family resemblance. *American Journal of Epidemiology* **110**, 335–342.

Donner, A. (1985). The analysis of intraclass correlation in multiple samples. *Annals of Human Genetics* **49**, 75–82.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review* **54**, 67–82.

Donner, A. (1987). Correction. *Biometrics* **43**, 1035–1035.

Donner, A. and Eliasziw, M. (1991). Methodology for inferences concerning familial correlations: Areview. *Journal of Clinical Epidemiology* **44**, 449–455.

Donner, A. and Koval, J. J. (1980*a*). The estimation of intraclass correlation in the analysis of family data. *Biometrics* **36**, 19–25.

Donner, A. and Koval, J. J. (1980*b*). The large sample variance of an intraclass correlation. *Biometrika* **67**, 719–722.

Donner, A., Koval, J. J. and Bull, S. (1984). Testing the effect of sex differences on sib-sib correlations. *Biometrics* **40**, 349–356.

Donner, A. and Wells, G. (1986). A comparison of confidence interval methods for the intraclass correlation coefficient. *Biometrics* **42**, 401–412.

Donner, A. and Zou, G. (2002). Testing the equality of dependent intraclass correlation coefficients. *The Statistician* **51** (3), 367–379.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

Eliasziw, M. and Donner, A. (1991). A generalized non-iterative approach to the analysis of family data. *Annals of Human Genetics* **55**, 77–90.

Elston, R. C. (1975). On the correlation between correlations. *Biometrika* **62**, 133–140.

Fieller, E. C. and Smith, C. A. B. (1951). Note on the analysis of variance and intraclass correlation. *Annals of Eugenics* **16**, 97–105.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

Giraudeau, B., Porcher, R. and Mary, J. (2005). Power calculation for the likelihood ratio-test when comparing two dependent intraclass correlation coefficients. *Computer Methods and Programs in Biomedicine* **77**, 165–173.

Haggard, E. A. (1958). *Intraclass Correlation and the Analysis of Variance.* Dryden Press, Inc, New York.

Hanley, J. A. (2004*a*). Transmuting women into men: Galton's family data on human stature. *American Statistician* **58**, 237–243.

Hanley, J. A. (2004*b*). http://www.epi.mcgill.ca/hanley/galton. .

Harville, D. A. and Fenech, A. P. (1985). Confidence intervals for a variance of ratio, or heritability, in an unbalanced mixed linear model. *Biometrics* **41**, 137–154.

Karlin, S., Cameron, E. C. and Williams, D. T. (1981). Siblings and parentoffspring correlation estimation with variance family size. *Proceedings of the National Academy of Science* **78**, 2664–2668.

Keen, K. J. (1993). Estimating the correlation among siblings. *Annals of Human Genetics* **57**, 297–305.

Keen, K. J. (1996). Limiting the effects of single-member families in the estimation of the intraclass correlation. *Biometrics* **52**, 823–832.

Konishi, S. and Gupta, A. K. (1987). Inferences about interclass and intraclass correlations from familial data. *In Advances in the Statistical Sciences* **4**, 225–233.

Lachin, J. M. (2004). The role of measurement reliability in clinical trials. *Clinical Trials* **1**, 553–566.

Lukacs, E. (1942). A characterization of the normal distribution. *The Annals of Mathematical Statistics* **13**, 91–93.

Mian, I. U. H. and Shoukri, M. M. (1997). Statistical analysis of intraclass correlations from multiple samples with applications to arterial blood pressure data. *Statistics in Medicine* **16**, 1497–1514.

Naik, D. N. and Helu, A. (2007). On testing equality of intraclass correlations under unequal family sizes. *Computational Statistics and Data Analysis* **51**, 6498–6510.

Pearson, K. (1901). Mathematical Contribution to the Theory of Evolution. IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and that of the Race. Part I. Homotyposis in the Vegetable Kingdom. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **197**, 285–379.

Ramasundarahettige, C. F., Donner, A. and Zou, G. Y. (2009). Confidence interval construction for a difference between two dependent intraclass correlation coefficients. *Statistics in Medicine* **28**, 1041–1053.

Rosner, B. (1982). On the estimation and testing of intraclass correlations: The general case of multiple replicates for each variable. *American Journal of Epidemiology* **116**, 722–730.

Searle, S. R. (1971). *Linear Models.* Wiley, New York.

Shenker, N. and Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician* **55**, 182–186.

Smith, C. A. B. (1956). On the estimation of intraclass correlations. *Annals of Human Genetics* **21**, 363–373.

Srivistava, M. S. (1993). Estimation of the intraclass correlation coefficient. *Annals of Human Genetics* **57**, 159–165.

Stanish, W. M. and Taylor, N. (1983). Estimator of the intraclass correlation coefficient for the analysis of covariance model. *American Statistician* **37**, 221–224.

Swiger, L. A., Harvey, L. R., Everson, D. O. and Gregory, K. E. (1964). The variance of intraclass correlation involving groups with one observation. *Biometrics* **20**, 818–826.

Thomas, J. D. and Hultquist, R. A. (1978). Interval estimation for the unbalanced case of the one-way random effects model. *Annals of Statistics* **6**, 582–587.

Ukoumunne, O. C., Davison, A. C., Gulliford, M. C. and Chinn, S. (2003). Nonparametric bootstrap confidence intervals for the intraclass correlation coefficient. *Statistics in Medicine* **22**, 3805–3821.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. *Annals of Mathematical Statistics* **11**, 96–100.

Weinberg, R. and Patel, Y. C. (1981). Simulated intraclass correlation coefficient and their z-transforms. *Journal of Statistical Computing and Simulation* **13**, 13–26.

Young, D. J. and Bhandary, M. (1998). Test for equality of intraclass correlation coefficients under unequal family sizes. *Biometrics* **54** (4), 1363–1373.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods* **12** (4), 399–413.

Zou, G. Y. (2008). On estimation of additive interaction by use of the four-by-four table and beyond. *American Journal of Epidemiology* **168**, 212–224.

Zou, G. Y. and Donner, A. (2008). Construction of confidence limits about effect measures: A general approach. *Statistics in Medicine* **27**, 1693–1702.