Thailand Statistician
2013; 11(2): 97-110
http://statassoc.or.th
Contributed paper

Multivariate meta-analysis on correlation coefficients

Arvind Shah $[a]$, **Jianxin Lin** $[a]$, **Thomas Mathew** $[b]$, **Bimal Sinha**$^*[b]$, **Gaurav Sharma**$^*[b]$, **Dihua Xu**$^*[b]$

$[a]$ Medical Communications Medical Support Group Clinical Biostatistics,
Merck Research Laboratories, Rahway, NJ 07065, USA

$[b]$ Department of Mathematics and Statistics,
University of Maryland, Baltimore 20250, USA

* corresponding author; e-mail: sinha@umbc.edu

Received: 28 November 2012
Accepted: 11 April 2013

Abstract

In this paper a large sample test is derived for testing the homogeneity of correlation matrices based on Fisher's z-transformation, and it is demonstrated that the test maintains the type I error rate satisfactorily. Towards this, the asymptotic joint distribution of the sample correlations is derived when the samples come from a multivariate population that could be non-normal. Assuming that the homogeneity hypothesis holds, methodology is provided to perform a meta-analysis of the common correlation matrix. An application to the correlations among the three cholesterol related variables: low-density lipoprotein (LDL), non-high density lipoprotein (NHDL) and Apolipoprotein B (APOB), in an investigation of the efficacy of cholesterol lowering drug, Ezetimibe, in combination with statins in patients with hypercholesterolemia, is provided.

Keywords: Correlation matrix, hypercholesterolemia, meta-analysis, multivariate CLT, Type I error.

1. Introduction

In this paper we address the problem of meta-analysis involving a set of correlations from a multivariate population. The basic premise is that we have k multivariate populations each of dimension p involving the same set of variables $X_1, \cdots, X_p$ across the populations, representing p characteristics of some physical entities. We want to test if there is homogeneity among the set of $q = p(p-1)/2$ population pairwise correlations across the k populations, and if so, how we would infer about the common set of correlations. Our solution is asymptotic in nature under the assumption that we have samples of a reasonably large size from all the k populations so that we can use the multivariate central limit theorem (MCLT) (Rao [1]). Trivially, for $p = 2$, the problem boils down to testing the equality of ordinary pairwise correlations from several bivariate populations for which a large sample test based on sample correlations, or based on their Fisher's variance stabilized versions, exists in the literature under the normality assumption (Rao [1], Hartung, Knapp and Sinha [2]). We should mention about two related papers in this context. Rayner et al. [3] discussed the likelihood ratio test of equality of covariance matrices (see also Anderson [4]), and Schott [5] addressed the same problem as ours using Wald statistic based on sample correlations. Our approach consists of first deriving the asymptotic multivariate normal distribution of the sample correlations, and their Fisher's z-transformed variables. Since the parent populations are not assumed to be normal, such an asymptotic distribution has a covariance matrix that involves the third and fourth moments of the parent distribution, which are assumed to exist. Based on such an asymptotic distribution, asymptotic tests can then be developed in an obvious manner, for testing the equality of the set of correlations across the k multivariate populations. In the case of a sample from a bivariate Type A Edgeworth distribution, Gayen [6] had obtained the asymptotic distribution of the sample correlation. Later, Hawkins [7] derived the asymptotic distribution of the sample correlation based on samples from an arbitrary bivariate population; this is obviously a special case of our general result. Hawkins [7] actually used asymptotic results on U-statistics (Hoeffding [8]) in order to obtain the asymptotic distribution of the sample correlation. Our derivation on the other hand is more direct. We provide brief details for dimensions three and four; the general case is rather obvious from the results for these special cases.

The organization of the paper is as follows. In Section 2, after some preliminary general discussion, we consider the first non-trivial case of dimension $p = 3$ involving a vector of three pairwise population correlations $(\rho_{12}, \rho_{13}, \rho_{23})$, and provide details of our asymptotics. We then provide the results for $p = 4$ and also briefly indicate how they can be used for a general p. A clinical application for $p = 3$ is given in Section 3. Some concluding remarks appear in Section 4.

Here is our general setup. Consider a $p \times 1$ vector $\mathbf{X} = (X_1, \cdots, X_p)'$ with correlation matrix $\mathbf{R}$ (the population correlation matrix), having the obvious representation

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & & \rho_{1p} \\ & 1 & \rho_{23} & \cdots & & \rho_{2p} \\ & & 1 & \cdots & & \rho_{3p} \\ & & & \ddots & & \vdots \\ & & & & 1 & \rho_{p-1p} \\ & & & & & 1 \end{pmatrix}.$$

The inference problem that we eventually want to address is in the context of statistical meta-analysis of correlation matrices. That is, given k population correlation matrices $\mathbf{R}_1, \cdots, \mathbf{R}_k$, we are interested in testing the homogeneity of the correlation matrices, i.e., testing the null hypothesis

$$H_0 : \mathbf{R}_1 = \cdots = \mathbf{R}_k,$$

based on samples from each of the k populations. We assume that we have a large sample of size n_i from the ith population, in order to develop a large sample test.

2. Asymptotic Distributions

In order to develop a large sample test, we shall first derive the asymptotic joint distribution of the sample correlations based on a sample of size n from a single multivariate population, where normality is not assumed. Since we are dealing with correlations, without loss of generality, we assume that the population mean vector is zero, and the population variances are all equal to one; i.e., the population covariance matrix is a correlation matrix. Let ρ_{ij}, $i < j$

(i, j = 1, 2, ... p), denote the correlation between the ith and jth variables in the population. For the sample $\mathbf{x}_l = (x_{1l}, x_{2l}, ..., x_{pl})'$, l = 1, 2, ..., n, we note that the sample correlations r_{ij}'s, $i < j$ (i, j = 1, 2, ... p), are functions of $\sum_{l=1}^{n} x_{il}^2/n$ (i = 1, 2, ... p), and $\sum_{l=1}^{n} x_{il}x_{jl}/n$, ($i < j$, i, j = 1, 2, ... p). We shall first derive the asymptotic multivariate normal distribution of the quantities $\sum_{l=1}^{n} x_{il}^2/n$ and $\sum_{l=1}^{n} x_{il}x_{jl}/n$, and then obtain the asymptotic multivariate normal distribution of the r_{ij}'s, $i < j$ (i, j = 1, 2, ... p). We shall illustrate the derivation for dimensions $p = 3$ and $p = 4$. The general case will then be obvious.

2.1 The Case $p = 3$

Let $(X_1, X_2, X_3)'$ follow a trivariate population. Let ρ_{ij} denote the correlation between X_i and X_j, and write $\mu_{rst} = E(X_1^r X_t^s X_3^t)$. Based on a random sample $\mathbf{x}_l = (x_{1l}, x_{2l}, x_{31})'$, l = 1, 2, ..., n, define the 6×1 vector

$$\begin{aligned} \mathbf{W} &= \left(\sum_{l=1}^{n} x_{1l}^2/n, \sum_{l=1}^{n} x_{2l}^2/n, \sum_{l=1}^{n} x_{3l}^2/n, \sum_{l=1}^{n} x_{1l}x_{2l}/n, \sum_{l=1}^{n} x_{1l}x_{3l}/n, \sum_{l=1}^{n} x_{2l}x_{3l}/n \right)' \\ &= (W_1, W_2, ..., W_6)'. \end{aligned} \tag{1}$$

The mean vector of $\mathbf{W}$, say $\tilde{\boldsymbol{\mu}}$, is given by

$$\tilde{\boldsymbol{\mu}} = (1, 1, 1, \rho_{12}, \rho_{13}, \rho_{23})'.$$

By the multivariate central limit theorem, the asymptotic joint distribution of $\sqrt{n}(\mathbf{W} - \tilde{\boldsymbol{\mu}})$ is multivariate normal with mean vector zero, and covariance matrix, say $\boldsymbol{\Sigma}$, given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mu_{400} - 1 & \mu_{220} - 1 & \mu_{202} - 1 & \mu_{310} - \rho_{12} & \mu_{301} - \rho_{13} & \mu_{211} - \rho_{23} \\ \mu_{220} - 1 & \mu_{040} - 1 & \mu_{022} - 1 & \mu_{130} - \rho_{12} & \mu_{121} - \rho_{13} & \mu_{031} - \rho_{23} \\ \mu_{202} - 1 & \mu_{022} - 1 & \mu_{004} - 1 & \mu_{112} - \rho_{12} & \mu_{103} - \rho_{13} & \mu_{013} - \rho_{23} \\ \mu_{310} - \rho_{12} & \mu_{130} - \rho_{12} & \mu_{112} - \rho_{12} & \mu_{220} - \rho_{12}^2 & \mu_{211} - \rho_{12}\rho_{13} & \mu_{121} - \rho_{12}\rho_{23} \\ \mu_{301} - \rho_{13} & \mu_{121} - \rho_{13} & \mu_{103} - \rho_{13} & \mu_{211} - \rho_{12}\rho_{23} & \mu_{202} - \rho_{13}\rho_{23} & \mu_{112} - \rho_{13}\rho_{23} \\ \mu_{211} - \rho_{23} & \mu_{031} - \rho_{23} & \mu_{013} - \rho_{23} & \mu_{121} - \rho_{12}\rho_{23} & \mu_{112} - \rho_{13}\rho_{23} & \mu_{022} - \rho_{23}^2 \end{pmatrix},$$

where we recall the notation $\mu_{rst} = E(X_1^r X_t^s X_3^t)$. The above structure of $\boldsymbol{\Sigma}$ follows directly from a close inspection of the elements of $\mathbf{W}$. From the definition of $\mathbf{W}$ given in (1), we see that the three pairwise sample correlations, say r_{12}, r_{13} and r_{23}, are given by

$$r_{12} = \frac{W_4}{\sqrt{W_1 W_2}}, \; r_{13} = \frac{W_5}{\sqrt{W_1 W_3}}, \; r_{23} = \frac{W_6}{\sqrt{W_2 W_3}}.$$

Applying the delta method, we conclude that $\sqrt{n}(r_{12}-\rho_{12}, r_{13}-\rho_{13}, r_{23}-\rho_{23})' \sim N(\mathbf{0}, \mathbf{\Sigma}_r)$, with

$$\mathbf{\Sigma}_r = \begin{pmatrix} \sigma_{11r} & \sigma_{12r} & \sigma_{13r} \\ & \sigma_{22r} & \sigma_{23r} \\ & & \sigma_{33r} \end{pmatrix},$$

where,

$$\begin{aligned}
\sigma_{11r} &= \frac{\rho_{12}^2}{4}(\mu_{400}+\mu_{040}+2\mu_{220})+\mu_{220}-\rho_{12}(\mu_{310}+\mu_{130}) \\
\sigma_{22r} &= \frac{\rho_{12}^2}{4}(\mu_{400}+\mu_{004}+2\mu_{202})+\mu_{202}-\rho_{13}(\mu_{301}+\mu_{103} \\
\sigma_{33r} &= \frac{\rho_{23}^2}{4}(\mu_{040}+\mu_{004}+2\mu_{022})+\mu_{022}-\rho_{23}(\mu_{031}+\mu_{013}) \\
\sigma_{12r} &= \mu_{211}-\frac{\rho_{12}}{2}(\mu_{301}+\mu_{121})-\frac{\rho_{13}}{2}(\mu_{310}+\mu_{112})+\frac{\rho_{12}\rho_{23}}{4}(\mu_{400}+\mu_{202}+\mu_{220}+\mu_{022}) \\
\sigma_{13r} &= \mu_{121}-\frac{\rho_{12}}{2}(\mu_{031}+\mu_{211})-\frac{\rho_{23}}{2}(\mu_{130}+\mu_{112})+\frac{\rho_{12}\rho_{23}}{4}(\mu_{400}+\mu_{022}+\mu_{220}+\mu_{202}) \\
\sigma_{23r} &= \mu_{111}-\frac{\rho_{13}}{2}(\mu_{013}+\mu_{211})-\frac{\rho_{23}}{2}(\mu_{103}+\mu_{121})+\frac{\rho_{13}\rho_{23}}{4}(\mu_{004}+\mu_{022}+\mu_{202}+\mu_{220}).
\end{aligned}$$

Now consider the Fisher's z-transformed variables (see Morrison [9], p. 101) $Z_{12} = \frac{1}{2}\ln[\frac{1+r_{12}}{1-r_{12}}]$, $Z_{13} = \frac{1}{2}\ln[\frac{1+r_{13}}{1-r_{13}}]$, and $Z_{23} = \frac{1}{2}\ln[\frac{1+r_{23}}{1-r_{23}}]$. Then $(Z_{12}, Z_{13}, Z_{23})'$ has an asymptotic multivariate normal distribution with mean vector $(\xi_{12}, \xi_{13}, \xi_{23})'$, where $\xi_{12} = \frac{1}{2}\ln[\frac{1+\rho_{12}}{1-\rho_{12}}]$ etc., and the asymptotic variances and covariances among $(Z_{12}, Z_{13}, Z_{23})'$ are given by

$$\begin{aligned}
\text{Var}(Z_{12}) &= (1-\rho_{12}^2)^{-2}\text{Var}(r_{12}) \\
\text{Var}(Z_{13}) &= (1-\rho_{13}^2)^{-2}\text{Var}(r_{13}) \\
\text{Var}(Z_{23}) &= (1-\rho_{23}^2)^{-2}\text{Var}(r_{23}) \\
\text{Cov}(Z_{12}, Z_{13}) &= (1-\rho_{12}^2)^{-1}(1-\rho_{13}^2)^{-1}\text{Cov}(r_{12}, r_{13}) \\
\text{Cov}(Z_{12}, Z_{23}) &= (1-\rho_{12}^2)^{-1}(1-\rho_{23}^2)^{-1}\text{Cov}(r_{12}, r_{23}) \\
\text{Cov}(Z_{13}, Z_{23}) &= (1-\rho_{13}^2)^{-1}(1-\rho_{23}^2)^{-1}\text{Cov}(r_{13}, r_{23}),
\end{aligned}$$

where $\text{Var}(r_{12}) = \sigma_{11r}$ etc. appearing above are the asymptotic variances and covariances. For $p = 2$, our expression for σ_{11r} exactly coincides, as expected, with that derived in Hawkins [7] based on Hoeffding's [8] projection method.

If the underlying population is multivariate normal, then the above expressions can be simplified, and the asymptotic covariance matrix of $(Z_{12}, Z_{13}, Z_{23})'$, say $\mathbf{\Sigma}_z$, is given by

$$\mathbf{\Sigma}_z = \frac{1}{n}\begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ & 1 & \alpha_3 \\ & & 1 \end{pmatrix}, \tag{2}$$

where

$$\alpha_1 = \frac{2\rho_{23}(1-\rho_{12}^2-\rho_{13}^2)+\rho_{12}\rho_{13}(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2-1)}{2(1-\rho_{12}^2)(1-\rho_{13}^2)},$$

$$\alpha_2 = \frac{2\rho_{13}(1-\rho_{12}^2-\rho_{23}^2)+\rho_{12}\rho_{23}(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2-1)}{2(1-\rho_{12}^2)(1-\rho_{23}^2)},$$

$$\alpha_3 = \frac{2\rho_{12}(1-\rho_{13}^2-\rho_{23}^2)+\rho_{13}\rho_{23}(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2-1)}{2(1-\rho_{13}^2)(1-\rho_{23}^2)}.$$

2.2 The Case $p = 4$

Let $(X_1, X_2, X_3, X_4)'$ follow a four-variate population. Let ρ_{ij} denote the correlation between X_i and X_j, and write $\mu_{rstu} = E(X_1^r X_2^s X_3^t X_4^u)$. Based on a random sample $\mathbf{x}_l = (x_{1l}, x_{2l}, x_{31}, x_{4l})'$, l = 1, 2, ..., n, define the 10×1 vector

$$\begin{aligned} \mathbf{W} &= \left(\sum_{l=1}^{n} x_{1l}^2/n, ..., \sum_{l=1}^{n} x_{4l}^2/n, \sum_{l=1}^{n} x_{1l}x_{2l}/n, ..., \sum_{l=1}^{n} x_{3l}x_{4l}/n\right)' \\ &= (W_1, W_2, ..., W_{10})'. \end{aligned} \tag{3}$$

The mean vector of $\mathbf{W}$, say $\tilde{\boldsymbol{\mu}}$, is given by

$$\tilde{\boldsymbol{\mu}} = (1, 1, 1, 1, \rho_{12}, \rho_{13}, ..., \rho_{34})'.$$

By the multivariate central limit theorem, the asymptotic joint distribution of $\sqrt{n}(\mathbf{W} - \tilde{\boldsymbol{\mu}})$ is multivariate normal with mean vector zero. The corresponding 10×10 asymptotic covariance matrix, say $\mathbf{\Sigma}$, can be worked out as a function of μ_{rstu}'s and the ρ_{ij}'s. For details, we refer to our technical report Shah et. al. [10]. The asymptotic joint distribution of the the sample correlations r_{ij}'s can be derived by the delta method. Upon direct computations, it turns out that the vector $\sqrt{n}(r_{12} - \rho_{12}, ...r_{34} - \rho_{34})$ is asymptotically multivariate normal with mean vector zero and covariance matrix, say $\mathbf{\Sigma}_r$, having elements σ_{ijr} given by

$$\begin{aligned}
\sigma_{11r} &= \frac{\rho_{12}^2}{4}(\mu_{4000}+\mu_{0400}+2\mu_{2200})+\mu_{2200}-\rho_{12}(\mu_{3100}+\mu_{1300}) \\
\sigma_{12r} &= \mu_{2110}-\frac{\rho_{12}}{2}(\mu_{3010}+\mu_{1210})-\frac{\rho_{13}}{2}(\mu_{3100}+\mu_{1120}) \\
&\quad +\frac{\rho_{12}\rho_{13}}{4}(\mu_{4000}+\mu_{2020}+\mu_{2200}+\mu_{0220}) \\
\sigma_{16r} &= \mu_{1111}-\frac{\rho_{12}}{2}(\mu_{2011}+\mu_{0211})-\frac{\rho_{34}}{2}(\mu_{1120}+\mu_{1102}) \\
&\quad +\frac{\rho_{12}\rho_{34}}{4}(\mu_{2020}+\mu_{2002}+\mu_{0220}+\mu_{0202}).
\end{aligned}$$

The expressions for σ_{22r}, ..., σ_{66r} are similar to that of σ_{11r}. The expressions for σ_{13r}, σ_{14r}, σ_{15r}, σ_{23r}, σ_{24r}, σ_{26r}, σ_{35r}, σ_{36r}, σ_{45r}, σ_{46r} and σ_{56r} are similar to that of σ_{16r}. Finally, the expressions for σ_{25r} and σ_{34r} are similar to that of σ_{16r}. Obviously, the asymptotic multivariate normal distribution of the Fisher's z-transformed variables can be worked out in a straightforward manner.

2.3 Meta-Analysis of Correlations: an Asymptotic Test

For testing the equality of the correlation matrices across k populations, we need to test the equality of the vector of correlations, or equivalently, that of the Fisher's z-transformed quantities. Let $\mathbf{z}^{(j)}$ denote the vector consisting of the Fishers z-transformed sample correlations based on a sample of size n_j from the jth population. Let $\mathbf{\Sigma}_z^{(j)}$ denote the covariance matrix of the asymptotic distribution of $\mathbf{z}^{(j)}$ and $\hat{\mathbf{\Sigma}}_z^{(j)}$ denote its estimate obtained by replacing the population mixed moments and pairwise correlations with the corresponding sample mixed moments and sample correlations. Define

$$T(\mathbf{z})=\sum_{j=1}^{k}(\mathbf{z}^{(j)}-\bar{\mathbf{z}})'\hat{\mathbf{\Sigma}}^{(j)^{-1}}(\mathbf{z}^{(j)}-\bar{\mathbf{z}}), \tag{4}$$

where $\bar{\mathbf{z}} = \big[\sum_{j=1}^{k}\hat{\mathbf{\Sigma}}^{(j)^{-1}}\big]^{-1}\big[\sum_{j=1}^{k}\hat{\mathbf{\Sigma}}^{(j)^{-1}}\mathbf{z}^{(j)}\big]$. We reject the hypothesis of equality of correlation matrices if $T(\mathbf{z}) > \chi^2_{p(k-1);\alpha}$, where $\chi^2_{r;\alpha}$ denotes the upper α percentile of a chisquare distribution with df = r. This test is the familiar and widely used Cochran's [11] test of homogeneity in meta-analysis.

Suppose the equality of the correlation matrices is accepted, and we want to derive a confidence set for the common set of correlations. For this, let ζ

denote the vector of Fisher's z-transformed population correlations. Note that

$$\bar{\mathbf{z}} \sim N\left[\zeta, \left(\sum_{j=1}^{k}(\hat{\Sigma}^{(j)})^{-1}\right)^{-1}\right]$$

An ellipsoidal confidence set (ECS) for ζ is then readily given by

$$ECS = \left[\zeta : (\zeta - \bar{\mathbf{z}})' \sum_{j=1}^{k}(\hat{\Sigma}^{(j)})^{-1}(\zeta - \bar{\mathbf{z}}) \leq \chi^2_{p(k-1);\alpha}\right]$$

3. A Clinical Application

In this section we apply the methods developed in the previous section to a clinical problem involving an investigation of the efficacy of the cholesterol lowering drug, Ezetimibe, in combination with statins (treatment), compared to the administration of statins alone (control), based on data from different studies. Furthermore, the patients in the different studies came from two lines of therapy: first line therapy, where the patients were not on any statins prior to entering the study, and second line therapy, where the patients were already on statins when they entered the study. Our analysis is based on baseline data set involving three cholesterol related variables: low-density lipoprotein (LDL), non-high density lipoprotein (NHDL) and Apolipoprotein B (APOB). A similar analysis can be carried out based on study end data also. We test the homogeneity of correlation matrices for overall, line 1 alone, and line 2 alone. The meta-analysis results are presented below. We also point out that although use of the typical scale factor $(n-3)$ in Fisher's z-transformation produces slightly better results, in our specific application it does not matter due to the large sample sizes. The analysis reported below assumes that we have samples from trivariate *normal* populations. Thus the asymptotic covariance matrix used is of the form (2).

3.1. Results of baseline correlation analysis

Table 1 shows the pairwise correlations at baseline from all the studies, identified by line 1 and 2. Table 2 shows the result from meta-analysis based on the statistic $T(\mathbf{z})$ in (4). It turns out that the homogeneity hypothesis about the correlation matrices is rejected for all the three cases we have considered.

Table 1. Baseline correlation analysis.

Trial	n	r_{12}	r_{13}	r_{23}	line
1	875	0.9002	0.7448	0.8553	1
2	664	0.8656	0.7001	0.7915	1
3	210	0.8307	0.8750	0.9624	2
4	700	0.9455	0.8732	0.9353	1
5	745	0.9173	0.8558	0.9355	1
6	619	0.9697	0.8771	0.9185	2
7	1528	0.8924	0.7880	0.9117	1
8	2832	0.8926	0.8268	0.9238	2
9	1790	0.8979	0.8283	0.9182	1
10	2854	0.8789	0.7714	0.9098	1
11	1167	0.9028	0.8209	0.9109	1
12	179	0.8582	0.7615	0.8686	2
13	551	0.8448	0.7379	0.8612	2
14	1069	0.8655	0.8250	0.9009	1
15	1021	0.9265	0.6860	0.7750	2
16	465	0.9167	0.8330	0.9175	2
17	296	0.9751	0.9256	0.9482	2
18	237	0.8985	0.7547	0.8350	1
19	545	0.8685	0.6842	0.7735	1
20	531	0.8819	0.7407	0.8061	1
21	625	0.8858	0.7270	0.8004	1
22	100	0.9581	0.8828	0.9132	2
23	406	0.7926	0.6527	0.8663	2
24	362	0.7867	0.6618	0.8503	2
25	170	0.7357	0.6241	0.8939	2
26	269	0.8340	0.6849	0.8826	2
27	422	0.8399	0.7093	0.8866	2
28	627	0.8802	0.8573	0.9569	2
29	593	0.8259	0.5495	0.6901	2

r_{12} is correlation between LDLC and NHDL;

r_{13} is correlation between LDLC and APOB;

r_{23} is correlation between NHDL and APOB

Table 2. Baseline correlation analysis; testing homogeneity of correlation matrices.

	$T(\mathbf{z})$	d.f.	p-value	Decision
Overall	2865.02	84	$< .0001$	Reject H_0
Line 1 alone	820.80	36	$< .0001$	Reject H_0
Line 2 alone	1994.76	45	$< .0001$	Reject H_0

3.2 Results of study end correlation analysis

A similar meta-analysis can be provided for study-end (post effect) correlations from all the studies identified by line 1 and 2, presented in Table 3. It turns out that the test of homogeneity hypothesis about the correlation matrices based on $T(\mathbf{z})$ is rejected for all the cases we have considered (see Shah et al. [10] for details).

Table 3. Study-end correlations.

Trial	n	r_{12}	r_{13}	r_{23}	line
1	849	0.9589	0.9074	0.9540	1
2	646	0.9653	0.9371	0.9687	1
3	201	0.8729	0.8332	0.8881	2
4	681	0.9497	0.9140	0.9659	1
5	745	0.9421	0.8923	0.9490	1
6	1509	0.9664	0.9366	0.9689	1
7	2832	0.9237	0.8746	0.9509	2
8	1790	0.9348	0.9048	0.9590	1
9	2854	0.9321	0.8907	0.9621	1
10	1167	0.9156	0.8621	0.9436	1
11	179	0.8736	0.8688	0.8981	2
12	551	0.8887	0.8369	0.9228	2
13	1069	0.9112	0.9005	0.9438	1
14	1021	0.9368	0.8647	0.9092	2
15	467	0.9318	0.8594	0.9329	2
16	296	0.9730	0.9173	0.9582	2
17	231	0.9537	0.8934	0.9496	1
18	535	0.9491	0.9163	0.9588	1
19	523	0.9344	0.8652	0.9187	1
20	614	0.9775	0.9567	0.9772	1
21	407	0.8975	0.8102	0.9253	2
22	363	0.9269	0.8588	0.9222	2
23	170	0.9071	0.8338	0.9450	2
24	271	0.9308	0.8663	0.9497	2
25	423	0.8915	0.8370	0.9380	2
26	636	0.9293	0.9022	0.9713	2
27	593	0.9338	0.8536	0.9223	2

r_{12} is correlation between LDLC and NHDL;
r_{13} is correlation between LDLC and APOB;
r_{23} is correlation between NHDL and APOB

For an increased scope of the meta-analysis of our data, we have also compared 3×3 pairwise correlation matrices under the following two scenarios:

- first line and second line correlation matrices at baseline (Table 4).
- baseline and study-end (post effect) correlation matrices by line 1 (Table 5) and line 2 (Table 6).

Our conclusion in terms of rejection of the homogeneity hypothesis remains the same in all the cases.

Table 4. Baseline, line 1 vs line 2; $T(\mathbf{z}) = 295.8704$ and p-Value $< .0001$.

	n	r_{12}	r_{13}	r_{23}
Baseline, line 1	13330	0.9040	0.8215	0.9051
Baseline, line 2	9122	0.9380	0.8648	0.9254

Table 5. Line 1, baseline vs study end; $T(\mathbf{z}) = 1189.955$ and p-Value $< .0001$.

	n	r_{12}	r_{13}	r_{23}
Baseline, line 1	13330	0.904	0.8215	0.9051
Study-end, line 1	12238	0.943	0.9087	0.9577

Table 6. Line 2, baseline vs study-end; $T(\mathbf{z}) = 169.354$ and p-Value $< .0001$.

	n	r_{12}	r_{13}	r_{23}
Baseline, line 2	9122	0.9380	0.8648	0.9254
Study-end, line 2	8417	0.9361	0.8859	0.9461

We conclude this section with a confirmatory simulation study of the Type I error rate of our proposed test based on $T(\mathbf{z})$ under the scenario of post effect overall correlation analysis when the null hypothesis H_0 is true. Such a simulation study is helpful due to the composite nature of the null hypothesis H_0. Under the ith study, we draw a random sample of size n_i from a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{R}$, then compute the sample correlation matrix $\hat{\mathbf{R}}_i$, and compute $T(\mathbf{z})$. Our choice of n_i's corresponds to those under the post effect data set (see Table 3), and the following choices were made for the true correlation matrix $\mathbf{R}$:

$$\mathbf{R}_1 = \begin{pmatrix} 1 & .9 & .8 \\ & 1 & .9 \\ & & 1 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 1 & .9 & .8 \\ & 1 & .8 \\ & & 1 \end{pmatrix}, \quad \mathbf{R}_3 = \begin{pmatrix} 1 & .9 & .7 \\ & 1 & .8 \\ & & 1 \end{pmatrix}, \quad \mathbf{R}_4 = \begin{pmatrix} 1 & .9 & .6 \\ & 1 & .5 \\ & & 1 \end{pmatrix},$$

$$\mathbf{R}_5 = \begin{pmatrix} 1 & .9 & .6 \\ & 1 & .4 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_6 = \begin{pmatrix} 1 & .7 & .7 \\ & 1 & .4 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_7 = \begin{pmatrix} 1 & .5 & .8 \\ & 1 & .4 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_8 = \begin{pmatrix} 1 & .6 & .5 \\ & 1 & .1 \\ & & 1 \end{pmatrix},$$

$$\mathbf{R}_9 = \begin{pmatrix} 1 & .9 & .3 \\ & 1 & .1 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_{10} = \begin{pmatrix} 1 & .7 & .2 \\ & 1 & .4 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_{11} = \begin{pmatrix} 1 & .3 & .7 \\ & 1 & .2 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_{12} = \begin{pmatrix} 1 & .3 & .5 \\ & 1 & .2 \\ & & 1 \end{pmatrix},$$

$$\mathbf{R}_{13} = \begin{pmatrix} 1 & .4 & .2 \\ & 1 & .2 \\ & & 1 \end{pmatrix}, \mathbf{R}_{14} = \begin{pmatrix} 1 & .3 & .2 \\ & 1 & .3 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_{15} = \begin{pmatrix} 1 & .1 & .2 \\ & 1 & .4 \\ & & 1 \end{pmatrix}, \ \mathbf{R}_{16} = \begin{pmatrix} 1 & .1 & .3 \\ & 1 & .1 \\ & & 1 \end{pmatrix}.$$

For a few values of the true correlation matrix (from the choices given above), the estimated Type I error rate using the proposed test are given in Table 7. It is clear that the proposed test maintains the stipulated Type I error rate satisfactorily. All throughout, we have used the R software, and assumed a 5% significance level.

Table 7. Type I error for multivariate test of homogeneity.

$\mathbf{R}$	Type I error
$\mathbf{R}_1$	0.0632
$\mathbf{R}_2$	0.0607
$\mathbf{R}_3$	0.0610
$\mathbf{R}_4$	0.0622
$\mathbf{R}_5$	0.0609
$\mathbf{R}_6$	0.0635
$\mathbf{R}_7$	0.0610
$\mathbf{R}_8$	0.0585
$\mathbf{R}_9$	0.0559
$\mathbf{R}_{10}$	0.0609
$\mathbf{R}_{11}$	0.0591
$\mathbf{R}_{12}$	0.0618
$\mathbf{R}_{13}$	0.0608
$\mathbf{R}_{14}$	0.0637
$\mathbf{R}_{15}$	0.0629
$\mathbf{R}_{16}$	0.0631

4. Conclusions

Meta-analysis involving correlation matrices has led to some new theoretical developments in terms of deriving an asymptotic test. The application to a real clinical data set is very illuminating, and hopefully the meta-analysis methods developed in this paper will be used elsewhere. We remark in passing that the method for testing homogeneity of correlation matrices based on Fisher's Z transformation maintains Type I error quite accurately even in the multivariate case, while the same is not true for the test based on sample correlations themselves even in dimension one (see Shah et al. [12]).

Acknowledgement: Our sincere thanks are due to Professor P.K. Sen for pointing out Hoeffding's [8] excellent paper and its scope in our context.

References

[1] Rao, C. R. *Linear Statistical Inference and Its Applications*, New York: Wiley, 1973.

[2] Hartung, J., Knapp, G., and Sinha, B. K. *Statistical Meta-Analysis with Applications*, New York: Wiley, 2008.

[3] Rayner, J. C. W., Manly, B. F. J. and Liddell, G. F. Hierarchic likelihood ratio tests for equality of covariance matrices, *Journal of Statistical Computation and Simulation*, 1990; 35: 91-99.

[4] Anderson. T. W. *An Introduction to Multivariate Statistical Analysis,* Third Edition, New York: Wiley, 2003.

[5] Schott, J. R. Testing for the equality of several correlation matrices, *Statistics and Probability Letters*, 1996; 27: 85-89.

[6] Gayen, A. K. The frequency distribution of the product moment correlation coefficient in random samples of any size drawn from non-normal universes, *Biometrika*, 1951; 38: 219-247.

[7] Hawkins, D. L. Using U statistics to derive the asymptotic distribution of Fisher's Z statistic, *The American Statistician*, 1989; 43: 235-237.

[8] Hoeffding, W. A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics*, 1948; 19: 293-325.

[9] Morrison, D. F. *Multivariate Statistical Methods*, New York: McGraw-Hill, 1967.

[10] Shah, A., Lin, J., Mathew, T., Sinha, B. K., Sharma, G., and Xu, D. Multivariate meta-analysis on correlation coefficients, *Technical Report*, Department of Mathematics and Statistics, University of Maryland Baltimore County, 2011.

[11] Cochran, W. G. Problems arising in the analysis of a series of similar experiments, *Journal of the Royal Statistical Society (Supplement)*, 1937; 4: 102-118.

[12] Shah, A., Lin, J., Mathew, T., Sinha, B. K., Sharma, G., and Xu, D. Meta-analysis and cumulative meta-analysis for correlations: methodology and applications in a study to assess the efficacy of ezetimibe coadministered with statins in patients with hypercholesterolemia, *Technical Report*, Department of Mathematics and Statistics, University of Maryland Baltimore County, 2010.