Comparison of Dependent Pearson and Spearman Correlation Coefficients with and without Correction for Measurement Error

Bernard Rosner1*, Wei Wang2, Heather Eliassen1,4 and Eileen Hibert4

1Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA
2Department of Medicine, Harvard Medical School, Brigham and Women’s Hospital, Boston, MA 02215, USA
3Department of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA 02215, USA
4Channing Division of Network Medicine, Harvard Medical School, Boston, MA 02215, USA

Abstract

There already exist methods for comparing dependent Pearson correlation coefficients. However, each of the variables (X, Y) has associated random error; and a related question is after correcting for random error, which variable correlates most highly with the outcome variable Z. In this paper, we present methods for comparing dependent deattenuated correlation coefficients. This is a generalization of previous work for obtaining confidence limits for a single deattenuated correlation coefficient. In addition, we extend this work to the comparison of dependent Spearman correlation coefficients. The methods are illustrated with two examples. The first example concerns the comparison of nephrotoxicity of phenacetin and aspirin intake as measured by repeat biomarkers obtained from the same subjects. The second example is a comparison of the validity of different storage conditions for measuring HbA1c from dried blood specimens as compared to the gold standard of immediate processing. Results from using these methods indicate that phenacetin intake is more highly correlated with serum creatinine levels than aspirin intake and that short-term storage is preferable to long-term storage for assessment of HbA1c levels. We have available SAS software for comparing dependent deattenuated Pearson correlation and dependent Spearman correlations with and without deattenuation.

Keywords: Deattenuated correlation; Dependent correlations; Measurement error correction; Pearson correlation; Spearman correlation

Introduction

In 1967/68 a group of 1,256 women were recruited for the Swiss Analgesic Study [1,2]. Women were age 30–49 years and lived in the Basel, Switzerland area. The purpose of the study was to investigate whether there was an association between intake of phenacetin-containing analgesics and prevalence and incidence of kidney disease. To measure phenacetin intake a urinary metabolite (N-acetyl-P-aminophenol (NAPAP)) was assayed from a urine sample in the clinic and from 2 additional urine samples collected at home on 2 separate days within 1 week of the clinic visit. Serum creatinine was used as a measure of kidney function. Women were seen at follow-up visits in 1969, 1970, 1971, 1972, 1975 and 1978. In the primary paper, mean NAPAP over the 3 replicates was categorized and related to both prevalence and incidence of kidney disease where an abnormal creatinine was defined as >1.5 mg/dl. In the present paper, we represent NAPAP and serum creatinine at the baseline clinic visit in 1968 as continuous variables and compute the correlation coefficient between these two measures. To distinguish phenacetin intake from general use of analgesics, urinary salicylates were also assessed in triplicate at baseline as an indication of intake of aspirin-containing analgesics. If we let X=NAPAP at the baseline clinic visit, Y=salicylates at the baseline clinic visit and Z=serum creatinine at the baseline clinic visit, then we wish to test the hypothesis $H_0: \rho_{xz} = \rho_{yz}$ vs $H_1: \rho_{xz} \neq \rho_{yz}$ (1) where $\rho_{xz}$ is the population correlation between X and Z and $\rho_{yz}$ is defined similarly.

There already exist methods for comparing dependent correlation coefficients for normally distributed random variables obtained from the same subjects. In an excellent review paper on this subject [3] it is noted that in previous simulation studies, several methods are clearly inappropriate, while other methods preserve type I error under a variety of conditions. Among the latter are the method of Williams [4] which is an enhancement of a procedure initially proposed by Hotelling [5].
The test statistic $Z$ is defined similarly, and we reject $H_0$ if $|Z| > z_p$, where $z_p$ is the $p$th percentile of a N(0,1) distribution.

Furthermore, Meng, Rosenthal, and Rubin [8] propose a statistic that is asymptotically equivalent to the Dunn and Clark procedure, but with a simpler expression for $\text{var}(Z_{x_z - Z_{y_z}})$ given by:

$$Z_0 = (Z_{x_z} - Z_{y_z}) \sqrt{\frac{N-3}{2(1-R_i^2)}} \sim N(0,1) \text{ under } H_0$$

where

$$h = (1 - R_i^2)/(1 - R^2_i),$$

and

$$f = \min\{1, (1-R_i^2)/(2(1-R_i^2)), 1\} \text{ and } R_i = (R_i^2 + R^2_i)/2.$$

In addition, Bilker, Brensinger, and Gur [9] propose a combined permutation-bootstrap approach to test the hypothesis in (1) based on the test statistic

$$Z_0 = (Z_{x_z} - Z_{y_z}) / \sqrt{\text{var}(Z_{x_z} - Z_{y_z})}$$

which is computationally intensive, but avoids the necessity of specifying the underlying distribution of $Z_0$. Finally, the R package [10] cocor maintained by Diedenhofen [11] provides software to implement a variety of methods for comparing dependent Pearson correlation coefficients.

However, an assumption of [4-8] is that the joint distribution of $(X, Y, Z)$ is multivariate normal. We relax this assumption and use the method of moments and the delta method to estimate $\text{var}(Z_{x_z} - Z_{y_z})$. Furthermore, in the Swiss Analgesic dataset, there is considerable intra-individual variation among replicate measures in both $X$ and $Y$ and the more important question is whether the true (underlying) mean value of $X$ and $Y$ at baseline, denoted by $\mu_x$ and $\mu_y$ are more highly correlated with $Z$. Thus, we wish to test the hypothesis

$$H_0: \text{corr}(\mu_x, Z) = \text{corr}(\mu_y, Z) \text{ vs } H_1: \text{corr}(\mu_x, Z) \neq \text{corr}(\mu_y, Z).$$

This is an extension of previous work [12] that provides confidence limits for a single deattenuated correlation. Another issue is that it is clear that neither $X$ nor $Y$ is normally distributed. Hence, the Spearman correlation coefficient may be a more appropriate measure of association than the Pearson correlation coefficient in this dataset. Hence, we extend our methodology to the comparison of dependent Spearman correlation coefficients both with and without correction for measurement error. To our knowledge, there is no previous literature on the comparison of dependent Spearman correlation coefficients.

In this paper, we first describe the methodology and propose both an asymptotic test and an exact test. Second, we present a simulation study to assess the validity of the asymptotic test in finite samples and compare it to existing procedures for comparing dependent correlation coefficients. Finally, we describe two examples based on real data illustrating the use of these methods.

Comparison of Dependent Deattenuated Pearson Correlations

Hypothesis testing

We consider a classical measurement error model for $X_i$ and $Y_i$ of the form

$$x_i = \alpha_i + \mu_{x_i} + e_{x_i}, \quad i = 1, \ldots, n, j = 1, \ldots, k$$

$$y_i = \alpha_i + \mu_{x_i} + e_{y_i}, \quad i = 1, \ldots, n, j = 1, \ldots, k$$

where

$$\text{var}(\mu_{x_i}) = \sigma_x^2, \text{var}(e_{x_i}) = \sigma_e^2, \text{var}(\mu_{y_i}) = \sigma_y^2, \text{var}(e_{y_i}) = \sigma_e^2, (\mu_{x_i}, e_{x_i}) \text{ and } (\mu_{y_i}, e_{y_i})$$

are respectively independent, $\text{corr}(e_{x_i}, e_{y_i}) = 0$ if replicate measures of $X$ and $Y$ are obtained from the same subjects.

We define the true correlation between $(X$ and $Z)$ and $(Y$ and $Z)$ by

$$\rho_{x_z, true} = \text{corr}(\mu_x, Z)$$

$$\rho_{y_z, true} = \text{corr}(\mu_y, Z)$$

Based on equations (2) and (3), it is straightforward to show that

$$\rho_{x_z, true} = \frac{\rho_{x_z, true}}{\sqrt{\text{ICC}_x}}$$

$$\rho_{y_z, true} = \frac{\rho_{y_z, true}}{\sqrt{\text{ICC}_y}}$$

where $\rho_{x_z} = \text{corr}(X_i, Z)$, $\rho_{y_z} = \text{corr}(Y_i, Z)$ and

$$\text{ICC}_x = \text{corr}(X_i, X_j) = \sigma_x^2, / (\sigma_x^2 + \sigma_e^2), \text{ICC}_y = \text{corr}(Y_i, Y_j) = \sigma_y^2, / (\sigma_y^2 + \sigma_e^2)$$

are the intra class correlations obtained from repeated measures of $X$ and $Y$ based on (2). We wish to test the hypothesis

$$H_0: \rho_{x_z, true} = \rho_{y_z, true} \text{ vs } H_1: \rho_{x_z, true} \neq \rho_{y_z, true}.$$

Note that it is possible that $\rho_{x_z} = \rho_{y_z}$, while $\rho_{x_z, true} > \rho_{y_z, true}$ if the reproducibility of $X$ is worse than that of $Y$ (i.e., ICC$_x$ < ICC$_y$). Similarly, it is possible that $\rho_{x_z} < \rho_{y_z}$, while $\rho_{x_z, true} = \rho_{y_z, true}$.

It is usually more efficient to base inferences on correlations based on Fisher’s $z$ statistic. Thus, we will test the hypothesis

$$H_0: Z_{x_z, true} = Z_{y_z, true} \text{ vs } H_1: Z_{x_z, true} \neq Z_{y_z, true}$$

where

$$z_{x_z, true} = \frac{1}{2} \ln[[1 - \rho_{x_z, true}]/(1 - \rho_{x_z, true})] \text{ and } Z_{y_z, true}$$

is defined similarly.

We will use the delta method to estimate

$$\text{var}(\hat{Z}_{x_z, true} - \hat{Z}_{y_z, true})$$

and obtain the test statistic

$$V_{x_z, true} = \frac{\hat{Z}_{x_z, true} - \hat{Z}_{y_z, true}}{\sqrt{\text{var}(\hat{Z}_{x_z, true})}} \sim N(0,1) \text{ for large } N.$$
Similarly, we can also test the hypothesis $H_0: \rho_{xz} = \rho_{yz} vs H_1: \rho_{xz} \neq \rho_{yz}$ based on the large sample test statistic

$$V_o = \frac{\tilde{\Delta}}{\sqrt{\text{var}(\tilde{\Delta})}} \sim N(0,1) \text{ for large } N$$

(7)

where $\tilde{\Delta} = \tilde{z}_{xz} - \tilde{z}_{yz}$ and based on the delta method

$$\text{var}(\tilde{\Delta}) = \text{var}(\tilde{z}_{xz}) = \text{var}(\tilde{z}_{yz}) - 2 \text{cov}(\tilde{z}_{xz}, \tilde{z}_{yz})$$

(8)

and

$$\text{var}[\text{ln}(R_v)], \text{var}[\text{ln}(R_s)], \text{and cov}[\text{ln}(R_v), \text{ln}(R_s)]$$

are given in Appendices A, B, and C, respectively of the supplementary materials.

**Interval estimation**

It is also of interest to obtain confidence limits for $\hat{\Delta}_{xz,yz} = \rho_{xz} - \rho_{yz}$. It is not possible to translate confidence limits for $\Delta_{xz,yz}$ to corresponding confidence limits for $\hat{\Delta}_{xz,yz}$. Instead, we will use $\Delta_{xz,yz} = R_{xz,yz} \sqrt{\text{ICC}} - R_{xz,yz} \sqrt{\text{ICC}}$ as a point estimate of $\Delta_{xz,yz}$. To obtain $\text{var}(\hat{\Delta}_{xz,yz})$ we use the formula:

$$\text{var}(\Delta_{xz,yz}) = \text{var}(R_{xz,yz}) + \text{var}(R_{xz,yz}) - 2 \text{cov}(R_{xz,yz}, R_{xz,yz})$$

(9)

Since $R_{xz,yz}$ is a ratio estimator, we first consider $\text{var}[\text{ln}(R_{xz,yz})]$ and use the delta method to obtain

$$\text{var}(R_{xz,yz}) \approx R_{xz,yz}^2 \text{var}[\text{ln}(R_{xz,yz})]$$

(10)

Furthermore, from the delta method, we have

$$\text{var}[\text{ln}(R_{xz,yz})] = \text{var}[\text{ln}(R_v)] + \frac{1}{4} \text{var}[\text{ln}(\text{ICC})]$$

(11)

where $\text{var}[\text{ln}(R_v)]$ and $\text{var}[\text{ln}(\text{ICC})]$ are obtained from Appendix A of the supplementary materials. Similarly,

$$\text{var}(R_{xz,yz}) \approx R_{xz,yz}^2 \text{var}[\text{ln}(R_v)] + \frac{1}{4} \text{var}[\text{ln}(\text{ICC})]$$

(12)

and

$$\text{var}[\text{ln}(R_v)]$$

are given in Appendix B of the supplementary materials. Finally, from the delta method we have:

$$\text{cov}(R_{xz,yz}, R_{xz,yz}) \approx R_{xz,yz} R_{xz,yz} \text{cov}[\text{ln}(R_{xz,yz}), \text{ln}(R_{xz,yz})]$$

(13)

$$\approx R_{xz,yz} R_{xz,yz} \left(\text{cov}[\text{ln}(R_v), \text{ln}(R_v)] + \frac{1}{4} \text{cov}[\text{ln}(\text{ICC}), \text{ln}(\text{ICC})]\right)$$

where $\text{cov}[\text{ln}(R_v), \text{ln}(R_v)]$ and $\text{cov}[\text{ln}(\text{ICC}), \text{ln}(\text{ICC})]$ are given in Appendix C of the supplementary materials.

Hence, if we combine (9)-(13) and assume asymptotic normality of $\Delta_{xz,yz}$, we have an approximate 100% $(1-\alpha)\text{CI}$ for $\Delta_{xz,yz}$ given by

$$\tilde{\Delta}_{xz,yz} \pm z_{1-\alpha/2} \sqrt{\text{var}(\tilde{\Delta}_{xz,yz})}$$

(14)

Similarly, we can obtain confidence limits for $\Delta = \rho_{xz} - \rho_{yz}$ given by

$$\hat{\Delta} \pm z_{1-\alpha/2} \sqrt{\text{var}(\hat{\Delta})}$$

(15)

**Small-Sample Inference**

When sample size is small, permutation methods can be used to estimate levels of significance. The permutation distribution is generated by randomly shuffling the X, Y labels and computing the empirical distribution of $\hat{\Delta}_{xz,yz}$ and $\tilde{\Delta}$ in (5) and (7), respectively. The rank of $\Delta_{xz,yz}$ and $\hat{\Delta}$ based on the observed data with reference to their permutation distributions can be used to estimate exact two-sided $p$-values given by

$$p-value = 2 \sum_{m=1}^{I} I(\hat{\Delta}_{xz,yz,perm,observed} \geq \hat{\Delta}_{xz,yz,perm}) \frac{M}{N}$$

(16)

And similarly for $\hat{\Delta}$, where $M$ is the size of the permutation distribution and $I(a) = 1$ if $a$ is true, $=0$ if $a$ is false.

**Comparison of Dependent Spearman Correlations**

**Hypothesis testing**

We can also consider the comparison of dependent Spearman correlations whereby we test the hypothesis

$$H_0: \rho_{xz} = \rho_{yz} vs H_1: \rho_{xz} \neq \rho_{yz}$$

where $\rho_{xz,ys}$ = population Spearman correlation between X and Z and $\rho_{xz,ys}$ is defined similarly.

If the data are transformed to the probit scale then from Rosner and Glynn [13],

$$\rho_{xz,ys} = \frac{6}{\pi} \text{sin}^{-1}(\rho_{xz,ys,prob}) / 2 = \frac{6}{\pi} \text{sin}^{-1}(\text{cor}(H_x,H_y) / 2)$$

(18)

$$\rho_{xz,ys} = \frac{6}{\pi} \text{sin}^{-1}(\rho_{xz,ys,prob}) / 2 = \frac{6}{2} \text{sin}^{-1}(\text{cor}(H_x,H_y) / 2)$$

(18)

where $H_x = \text{probit}(F_x(X))$ is estimated by $\Phi^{-1}\text{rank}(X) / N+1$, where $\Phi = \text{c.d.f.}$ of a $(0,0.1)$ distribution, $\rho_{xz,ys}$ = population correlation between the probit of X and the probit of Y and $\rho_{xz,ys}$ is defined similarly.

Thus, the hypothesis test in (17) is equivalent to the hypothesis test in (7) after transformation to the probit scale yielding a test statistic based on $\Delta_{xz,yz,prob} = Z_{xz,prob} - Z_{yz,prob}$ given by $V_{o,prob} = \hat{\Delta}_{xz,yz,prob} / \sqrt{\text{var}(\hat{\Delta}_{xz,yz,prob})}$.

**Interval estimation**

It is also of interest to obtain confidence limits for $\Delta_{xz,ys} = \rho_{xz} - \rho_{yz}$ which we estimate by $\Delta_{xz,ys} = R_{xz,ys} - R_{yz,ys}$. Based on the delta method and (18), we obtain:

$$\text{var}(R_{xz,ys} - R_{yz,ys}) = \text{var}(R_{xz,ys}) + \text{var}(R_{yz,ys}) - 2 \text{cov}(R_{xz,ys}, R_{yz,ys})$$

(19)

$$\approx \frac{9R_{xz,ys,prob}^2}{\pi^2} \text{var}[\text{ln}(R_{xz,ys,prob})] + \frac{9R_{yz,ys,prob}^2}{\pi^2} \text{var}[\text{ln}(R_{yz,ys,prob})]$$

where $\hat{\Delta}_{xz,ys} = R_{xz,ys} - R_{yz,ys}$ and $\tilde{\Delta}_{xz,ys} = R_{xz,ys} - R_{yz,ys}$.
Correction for measurement error

Rosner and Glynn [13] defined the deattenuated Spearman correlation by

\[ \rho_{xz,\text{true}} = \rho_{xz,\text{prob}} \left( \frac{\sigma_x \sigma_z}{\rho_{xz,\text{true}}} \right) \]

where \( \rho_{xz,\text{true}} \) is the deattenuated correlation between true scores for X and Z, respectively. Hence, the hypotheses

\[ H_0: \rho_{xz,\text{true}} = \rho_{xz,\text{true}} \text{ vs } H_1: \rho_{xz,\text{true}} \neq \rho_{xz,\text{true}} \]

are equivalent to the hypotheses

\[ H_0: \rho_{xz,\text{prob}} = \rho_{xz,\text{prob}} \text{ vs } H_1: \rho_{xz,\text{prob}} \neq \rho_{xz,\text{prob}} \]

Furthermore, the latter hypotheses can be tested using (5), where the Fisher z transformation is performed based on probit scores, yielding a test statistic based on \( \Delta_{xz,\text{prob}} = Z_{xz,\text{prob}} - Z_{xz,\text{prob}} \), given by

\[ V_{xz,\text{prob}} = \Delta_{xz,\text{prob}} / \sqrt{\text{var}(\Delta_{xz,\text{prob}})} \]

and

\[ \text{var}(\Delta_{xz,\text{prob}}) = \text{var}(Z_{xz,\text{prob}}) + \text{var}(Z_{xz,\text{prob}}) - 2\text{cov}(Z_{xz,\text{prob}}, Z_{xz,\text{prob}}) \]

and \( \text{cov}(Z_{xz,\text{prob}}, Z_{xz,\text{prob}}) \) are obtained from equations (A3), (B1) and (C1), of Appendices A, B and C, respectively, of the supplementary materials.

To obtain a 100% × (1-α)CI for \( \Delta_{xz,\text{true}} = \rho_{xz,\text{true}} - \rho_{xz,\text{true}} \), we use a similar approach as in (19) and obtain:

\[ \text{var}(\Delta_{xz,\text{true}}) = \text{var}(\Delta_{xz,\text{true}}) + \text{var}(\Delta_{xz,\text{true}}) - 2\text{cov}(\Delta_{xz,\text{true}}, \Delta_{xz,\text{true}}) \]

where

\[ \text{cov}(\Delta_{xz,\text{true}}, \Delta_{xz,\text{true}}) = \text{cov}(Z_{xz,\text{true}}, Z_{xz,\text{true}}) \]

are obtained from (11)-(13), respectively.

The corresponding large sample 100% × (1-α)CI for \( \Delta_{xz,\text{true}} \) is given by

\[ \hat{\Delta}_{xz,\text{true}} \pm z_{1-\alpha/2} \sqrt{\text{var}(\hat{\Delta}_{xz,\text{true}})} \]

where

\[ \text{var}(\hat{\Delta}_{xz,\text{true}}) = \text{var}(\Delta_{xz,\text{true}}) + \text{var}(\Delta_{xz,\text{true}}) - 2\text{cov}(\Delta_{xz,\text{true}}, \Delta_{xz,\text{true}}) \]

Simulations

We conducted simulation studies to assess the validity of the asymptotic procedure in (7) for comparing dependent correlations in finite samples and to compare this procedure with the procedures of Dunn and Clark [7] and Meng, Rosenthal and Rubin [8]. We simulated data for (X, Y, Z) from a multivariate normal distribution \( N(\mu, \Sigma) \) with

\[ \mu = (0, 0, 0) \text{ and } \Sigma = \begin{pmatrix} 1 & 0.732 & 1 \\ 0.732 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \]

We performed 4,000 simulations for each combination of \( N=100, 400 \) and

\[ \rho_{xz}, \rho_{yz} = (0.305, 0.305), (0.500, 0.500) \text{ and } (0.305, 0.500) \]

We also simulated data from a log normal distribution \((X^*, Y^*, Z^*)\), where \( X^* = \exp(X) \), \( Y^* = \exp(Y) \) and \( Z^* = \exp(Z) \). The results are given in Table 1. We see that in the case of a normal distribution that all three procedures have appropriate type I error (designs 1-4), particularly in the case of \( n=400 \). Similarly, the power of the procedures (designs 5-6) is also comparable. However, in the case of a log normal distribution, type I error is more adequately preserved by the method of moments procedure, particularly for \( n=400 \). As the sample size increases, the inconsistent estimate of the variance provided by the Dunn and Clark and Meng procedures in the setting of a non-normal distribution yields variance estimates that are too low and type I errors that are too high. Conversely, as \( n \) increases the method of moments procedure yields consistent variance estimates and actual \( p \)-values that are close to nominal levels. Alternative approaches if non-normality is suspected would be to compare dependent Spearman correlations rather than Pearson correlations.

To assess the accuracy of the procedure in (5) to compare dependent deattenuated Pearson correlations we used the same data designs as in designs 1-6. We performed each set of simulations in two ways. First, we assumed that the intra class correlation between replicate measures of \( X \) and \( Y \) was known without error and set to 0.67. Thus, \( \text{var}(\hat{\Delta}_{xz,\text{true}}) = \text{var}(\Delta_{xz,\text{true}}) + \text{var}(\Delta_{xz,\text{true}}) - 2\text{cov}(\Delta_{xz,\text{true}}, \Delta_{xz,\text{true}}) \) in equations (A3), (B1) and (C2). Second, we generated replicate values of \((X_1, X_2, Y_1, Y_2)\) from a multivariate normal distribution denoted by \( N(\mu, \Sigma) \) where

\[ \mu_{(0,0,0,0,0)} = (0.67, 0.60, 0.50, 0.67, 0.60, 0.50, 0.67, 0.50, 0.60, 0.67) \]

and estimated ICC_x and ICC_y from the replicate data. For each set of simulations we estimated the empirical type I error of \( \text{var}(\hat{\Delta}_{xz,\text{true}}) \) and the bias and variance of \( \hat{\Delta}_{xz,\text{true}} \) in (5). The results are given in Table 2.

We see in Table 2 that the median theoretical variance and the empirical variance of \( \hat{\Delta}_{xz,\text{true}} \) and \( \Delta_{xz,\text{true}} \) are very similar both for the case where the ICC is assumed known (Type 1) and where it is estimated from the simulated data (Type 2). In addition, the empirical type I error is close to 0.05 (range from 0.039 to 0.061) for all parameter
is close to 0 for all parameter combinations considered. The and when is close to 95% both for is close to 0.05 and the coverage probability both when.

*Based on equation 7

Table 1: Comparison of alternative methods for comparing dependent correlations.

| Type | ICC(ICC) | $\rho_{xz}$ | $\rho_{yz}$ | N | Method of Moments* | Dunn and Clark | Meng | Method of Moments* | Dunn and Clark | Meng |
|------|----------|-------------|-------------|---|--------------------|----------------|------|--------------------|----------------|------|
| 1    | Normal   | 0.305       | 0.305       | 100 | 0.061              | 0.054          | 0.052 | ---                | ---            | ---  |
| 2    | Normal   | 0.305       | 0.305       | 400 | 0.045              | 0.045          | 0.045 | ---                | ---            | ---  |
| 3    | Normal   | 0.5         | 0.5         | 100 | 0.063              | 0.054          | 0.052 | ---                | ---            | ---  |
| 4    | Normal   | 0.5         | 0.5         | 400 | 0.045              | 0.045          | 0.045 | ---                | ---            | ---  |
| 5    | Normal   | 0.305       | 0.5         | 100 | ---                | ---            | ---   | 0.866              | 0.842          | 0.839|
| 6    | Normal   | 0.305       | 0.5         | 400 | ---                | ---            | ---   | 0.987              | 1              | 1    |
| 7    | Lognormal| 0.305       | 0.305       | 100 | 0.102              | 0.162          | 0.16  | ---                | ---            | ---  |
| 8    | Lognormal| 0.305       | 0.305       | 400 | 0.075              | 0.197          | 0.196 | ---                | ---            | ---  |
| 9    | Lognormal| 0.5         | 0.5         | 100 | 0.116              | 0.247          | 0.244 | ---                | ---            | ---  |
| 10   | Lognormal| 0.5         | 0.5         | 400 | 0.082              | 0.302          | 0.326 | ---                | ---            | ---  |
| 11   | Lognormal| 0.305       | 0.5         | 100 | ---                | ---            | ---   | 0.495              | 0.581          | 0.577|
| 12   | Lognormal| 0.305       | 0.5         | 400 | ---                | ---            | ---   | 0.812              | 0.919          | 0.919|

Theoretical variance estimates are medians over 4000 simulated samples; Type 1 simulations assume that ICCx, ICCy are known; Type 2 simulations estimates ICCx, ICCy from the sample data; for both Type 1 and Type 2 simulations, the underlying ICCx, ICCy are provided in the 2nd column of the table. Simulation results are based on 4000 simulations.

Table 2: Simulation results for the test procedure in (5) and (6).

| Type | N | Method of Moments* | Mean | Variance | Mean | Variance | Mean | Variance |
|------|---|--------------------|------|----------|------|----------|------|----------|
| 1    |    | theoretical        | 0.391| 0.0163   | 0.391| 0.0162   | 0    | 0.0095   |
| 2    |    | theoretical        | 0.391| 0.0041   | 0.391| 0.0041   | 0    | 0.0024   |
| 3    |    | theoretical        | 0.392| 0.0212   | 0.71 | 0.0212   | 0    | 0.0146   |
| 4    |    | theoretical        | 0.71 | 0.0229   | 0.72 | 0.0233   | -0.003| 0.0161   |
| 5    |    | theoretical        | 0.71 | 0.0053   | 0.71 | 0.0055   | -0.001| 0.0036   |
| 6    |    | theoretical        | 0.391| 0.0166   | 0.391| 0.0168   | 0    | 0.01      |
| 7    |    | theoretical        | 0.391| 0.0042   | 0.391| 0.0042   | 0    | 0.0025   |
| 8    |    | theoretical        | 0.394| 0.0226   | 0.71 | 0.0222   | 0    | 0.0167   |
| 9    |    | theoretical        | 0.724| 0.027    | 0.728| 0.0269   | -0.004| 0.0201   |
| 10   |    | theoretical        | 0.71 | 0.0056   | 0.71 | 0.0056   | 0    | 0.004     |
| 11   |    | theoretical        | 0.715| 0.0058   | 0.713| 0.0057   | 0.002| 0.0044   |

1,256 women provided a urine specimen in the clinic and 2 additional specimens at home on different days which were assayed for NAPAP, a metabolite indicating recent intake of phenacetin-containing analgesics and salicylates a metabolite indicating recent intake of aspirin-containing analgesics. In addition, serum creatinine, a marker of kidney function was also measured at the baseline clinic visit. We let X, Y and Z be the NAPAP, salicylates and serum creatinine at the baseline (1968) clinic visit. We wish to compare $\text{corr}(X,Z)$ with $\text{corr}(Y,Z)$. A total of 1168 women had complete data on X, Y and Z. Descriptive statistics for the (X,Y,Z) data are provided in Table 4.

We note that the distributions of each of X, Y and Z are skewed and not normally distributed. There is moderate correlation between NAPAP and salicylates (Pearson correlation=0.423) in part because some analgesics contain both phenacetin and aspirin. Pearson correlations between NAPAP and salicylates vs. serum creatinine were 0.190 and 0.068, respectively. The ICC for NAPAP and salicylates over the 3 determinations were 0.614 and 0.404, respectively, indicating moderate variability of intake on different days.

**Examples**

**Association between analgesic intake and level of kidney function**

We use data from the Swiss Analgesic Study where in 1967/68...
Based on (7), the observed Pearson correlations were significantly different ($V_{xy} = 3.344$, $p = 0.001$) (Table 5). For comparison, we also analyzed these data using the Dunn and Clark [7] and Meng, et al. [8] procedures. Based on Dunn and Clark, we have $Z = 3.929$, $p$-value $= 8.5 \times 10^{-5}$. Based on Meng, et al, we have $Z_{xy} = 3.919$, $p$-value $= 8.9 \times 10^{-5}$. As in the simulations, the se's are inappropriately low for both of these procedures in the setting of these highly non-normal data, yielding $p$-values that are biased downward and type I errors that are too low. The deattenuated Pearson correlations were 0.243 and 0.108 for NAPAP (X) and salicylates (Y), respectively, vs. serum creatinine. Based on (5) there was a significant difference between these correlations, $V_{xy,true} = 2.640$, $p = 0.008$.

Since NAPAP and salicylates were right-skewed, we also computed observed and deattenuated Spearman correlations which are presented in Table 6.

The $V$ statistic for the observed and deattenuated Pearson correlations is based on (7) and (5), respectively.

| Variable                      | observed correlation | Z transform | deattenuated correlation | Z transform |
|-------------------------------|----------------------|-------------|--------------------------|-------------|
| NAPAP, 1968 (o.d.)            | 0.19                 | 0.192       | 0.243                    | 0.248       |
| salicylates, 1968 (mg%)       | 0.068                | 0.069       | 0.108                    | 0.108       |
| Difference between correlations | 0.122               | 0.124       | 0.135                    | 0.139       |
| (95%CI)                       | (0.051, 0.193)       | (0.035, 0.235) |                          |             |
| $p$-value                     | 0.001                | 0.001       |                          | 0.008       |

The $V$ statistic for the observed and deattenuated Spearman correlations is based on (14).

Table 5: Observed and deattenuated Pearson Correlations between serum creatinine at baseline (1968) vs. each of NAPAP and salicylates at the baseline clinic visit (1968), Swiss Analgesic Study, $n=1168$. Since NAPAP and salicylates were right-skewed, we also computed observed and deattenuated Spearman correlations which are presented in Table 6.

The observed Spearman correlations between $Z$ and (X,Y) were 0.155 and 0.081, respectively, which were significantly different ($V_{xy,prob} = 2.395$, $p = 0.017$). After correcting for deattenuation, the Spearman correlations between $Z$ and (X,Y) were 0.212 and 0.127, respectively, which only showed a trend towards statistical significance ($V_{xy,prob,true} = 1.849$, $p = 0.064$). Overall, the results from the comparison of correlation coefficients in Tables 5 and 6 were consistent with the

| Variable                  | Pearson (Spearman) correlation | Pearson (Spearman) | Type I error | Coverage probability |
|---------------------------|--------------------------------|--------------------|--------------|---------------------|
| NAPAP, 1968 (o.d.)       | 0.19                           | 0.192              | 0.243        | 0.248               |
| salicylates, 1968 (mg%)  | 0.068                          | 0.069              | 0.108        | 0.108               |
| serum creatinine (Z) (mg/dL) | 1.02                         | 0.37               | 1            |                     |

NAPAP value at the baseline (1968) clinic visit. Individual salicylate values were obtained in grouped form (1=0-19 mg%/2=20-49 mg%/3=50-99 mg%/4=100+mg%) at the baseline (1968) clinic visit, and assigned scores of 10 mg%, 35 mg%, 75 mg% and 100 mg%, respectively. Spearman correlations are estimated based on (18) N: Number of subjects with serum creatinine, NAPAP and salicylates, available at the baseline (1968) clinic visit ICC: Intraclass Correlation among replicate urine values (based on urine specimens obtained at the baseline clinic visit and 2 additional urine specimens obtained at home on different days).
recently dried blood spot assays of HbA1c have been used as biomarkers in many epidemiologic studies and simplified specimen handling would be cost-effective for these large scale community-based studies. A study was conducted at Brigham and Women’s Hospital (BWH) to test the hypothesis that dried blood spot determinations for HbA1c are valid measurements with low-intensity storage conditions [14]. Blood samples were drawn into EDTA containing tubes and submitted for duplicate HPLC analysis of HbA1c (considered as the gold standard process). Blood for spotting was also drawn to identical EDTA tubes and then dropped randomly to blood-spot cards. After having been air-dried for at least 20 minutes these dried blood specimens were placed into single-sample, airtight bags with desiccant having been air-dried for at least 20 minutes these dried blood specimens were placed into single-sample, airtight bags with desiccant. The V statistic for the observed and deattenuated probit correlations is based on (7) and (5), respectively.

Table 7: Spearman Correlations between HbA1c levels determined immediately vs. HbA1c levels determined after 2 delay periods, N=168.

| Variable | observed Spearman correlation | observed probit correlation | Z transform | deattenuated Spearman correlation | deattenuated probit correlation | Z transform | ICCa |
|----------|-------------------------------|-----------------------------|-------------|----------------------------------|---------------------------------|-------------|------|
| 2 weeks of room temperature pre-shipping plus 4 weeks of freezer (X) | 0.952 | 0.956 | 1.896 | 0.979 | 0.98 | 2.31 | 0.946 |
| 4 weeks of room temperature pre-shipping plus 12 weeks of freezer (Y) | 0.711 | 0.727 | 0.922 | 0.776 | 0.79 | 1.072 | 0.835 |
| difference | 0.241 | 0.229 | 0.974 | 0.203 | 0.19 | 1.238 | 0.952 |
| se | 0.348 | 0.827 |
| V statistic | 2.797 | 1.849 |
| p-value (large sample) | 0.005 | 0.048 |
| p-value (exact) | 0.021 | 0.064 |

The V statistics are based on (7) and (5) using the probit transformation.

Comparison of storage conditions for HbA1c measurements in plasma

We refer to storage for 2 weeks at room temperature plus 4 additional weeks in the freezer as short-term storage (X), storage for 4 weeks at room temperature plus 12 additional weeks in the freezer as long-term storage (Y) and HPLC processing as Z. The observed and deattenuated Spearman correlations between Z vs. (X,Y) respectively are given in Table 7. We see that ρ_{x,z}=0.952 while ρ_{y,z}=0.776, which are significantly different both based on large-sample methods (p=0.005) and also more appropriate exact methods (p=0.021). After deattenuation, we obtain ρ_{x,z}^{true}=0.979 and ρ_{y,z}^{true}=0.776 (2-sided p-value=0.134 based on large sample methods and 0.056 based on exact methods). Thus, the length of storage has a borderline significant effect on the validity of the HbA1c assay with short-term storage preferable. Plots of (Δ_{x,z},\Delta_{y,z}) in equation 16 over the 4,096 elements of the permutation distribution are given in Figures 1 and 2, respectively. The distributions are somewhat skewed to the right, particularly for (Δ_{x,z},\Delta_{y,z}), which indicates the necessity of using exact methods in a small-sample setting.
of two surrogate measures (X, Y) where validity is measured by the respective correlation with a third variable (Z) and both X and Y are subject to measurement error. Although the methods are asymptotic, we have shown in finite samples that the asymptotic properties of the test statistic and confidence limits are appropriate when N ≥ 100. In addition, we provide a method for confidence interval estimation of both $\Delta_p$ and $\Delta_{p,\text{true}}$.

The methods were extended to the comparison of dependent Spearman correlations which may be more appropriate when (X, Y, Z) are not multivariate normal, which provide for the comparison of both ordinary and deattenuated Spearman correlations. To our knowledge, this is the first paper in the literature to discuss comparison of dependent Spearman correlations.
In this paper, we have considered deattenuated correlations of the form $\text{corr}(\mu_x, Z)$ and $\text{corr}(\mu_y, Z)$ which account for error in the estimation of $X$ and $Y$. However, it is likely that $Z$ is also measured with error and a useful extension would be to estimate and compare $\text{corr}(\mu_x, \mu_z)$ vs $\text{corr}(\mu_y, \mu_z)$, where error in $X$, $Y$ and $Z$ are taken into account.

SAS macros for the comparison of dependent ordinary and deattenuated Pearson and Spearman correlations as discussed in this paper are available from the following website:

https://sites.google.com/a/channing.harvard.edu/bernardrosner/channing.

Acknowledgment

We would like to thank Drs. Buxton and Seeman for providing the original data on HbA1c.

The research was supported by U.S. National Institute of Health grants (U54 CA155626-03, P30 CA006516-18, R01 CA050597-18 and P30 AG017265-099002). This work was conducted with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR001102) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, of the National Institutes of Health.

References

1. Dubach UC, Levy PS, Muller A (1971) Relationships between regular analgesic intake and urogenital disorders in a working female population of Switzerland. I. Initial results (1968). Am J Epidemiol 93: 425-434.
2. Dubach UC, Rosner B, Pfister E (1983) Epidemiologic study of abuse of analgesics containing phenacetin. Renal morbidity and mortality (1968-1979). N Engl J Med 308: 357-362.
3. Steiger JH (1980) Tests for comparing elements of a correlation matrix. Psychological Bulletin 87: 245-251.
4. Williams EJ (1959) The comparison of regression variables. Journal of the Royal Statistical Society Series B (Methodological) 21: 396-399.
5. Hotelling H (1940) The selection of variates for use in prediction with comments on the general problem of nuisance parameters. Annals of Mathematical Statistics 11: 271-283.
6. Olkin I, Finn J (1990) Testing Correlated Correlations. Psychological Bulletin 108: 330-333.
7. Dunn OJ, Clark V (1969) Correlation Coefficients Measured on the Same Individuals. Journal of the American Statistical Association 64: 366-377.
8. Meng XL, Rosenthal R, Rubin DB (1992) Comparing correlated correlation coefficients. Psychological Bulletin 111: 172-175.
9. Bilker WB, Brensinger C, Gur RC (2004) A two factor ANOVA-like test for correlated correlations: CORANOVA. Multivariate Behavioral Research 39: 565-594.
10. RCoreTeam (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
11. Diedenhofen B (2014) cocor: Comparing correlations.
12. Rosner B, Willett WC (1988) Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. American journal of epidemiology 127: 377-386.
13. Rosner B, Glynn RJ (2007) Interval estimation for rank correlation coefficients based on the probit transformation with extension to measurement error correction of correlated ranked data. Stat Med 26: 633-648.
14. Buxton OM, Malarick K, Wang W, Seeman T (2009) Changes in dried blood spot Hb A1C with varied post collection conditions. Clin Chem 55: 1034-1036.