

Power Analyses for Correlations from Clustered Study Designs

X.M. Tu, J.Kowalski, P. Crits-Christoph and R. Gallop*

To appear in *Statistics in Medicine*

February 23, 2006

Abstract

Power analysis constitutes an important component of modern clinical trials and research studies. Although a variety of methods and software packages are available, almost all of them are focused on regression models, with little attention paid to correlation analysis. However, the latter is arguably a simpler and more appropriate approach for modeling concurrent events, especially in psychosocial research. In this paper, we discuss power and sample size estimation for correlation analysis arising from clustered study designs. Our approach is based on the asymptotic distribution of correlated Pearson-type estimates. Although this asymptotic distribution is easy to use in data analysis, the presence of a large number of parameters creates a major problem for power analysis due to the lack of real data to estimate them. By introducing a surrogacy-type assumption, we show that all nuisance parameters can be eliminated, making it possible to perform power analysis based only on the parameters of interest. Simulation results suggest that power and sample size estimates obtained under the proposed approach are robust to this assumption.

Key words: Asymptotic distribution, Longitudinal study, Repeated measures, Robustness, Surrogacy condition.

*Tu is Professor, Department of Biostatistics and Computational Biology, and Department of Psychiatry, University of Rochester, 601 Elmwood Ave., Rochester, NY 14642. Kowalski is Assistant Professor, Department of Oncology and Biostatistics, Johns Hopkins University. Crits-Christoph is Professor, Department of Psychiatry, University of Pennsylvania School of Medicine. Gallop is Assistant Professor, Department of Mathematics and Applied Statistics, West Chester University, West Chester, PA 19383.

1 Introduction

Power and sample size estimation constitutes an important component of modern clinical trials and research studies. As clustered study designs, particularly longitudinal trials, become more popular in biomedical and psychosocial research, the use of existing packages is becoming increasingly limited. Although the past two decades have witnessed some major developments in methodological research for clustered study designs, much of the focus has been centered around data analysis, with little attention paid to correlation analysis^{[1]–[15]}. In psychosocial research, correlation analysis is a very popular analytic tool for modeling association between different measures^{[16]–[23]}. In comparison to linear regression, correlation analysis does not distinguish between response and covariate (or predictor) variable and is particularly useful for modeling association between variables measuring concurrent events that may arise from different outcomes, scales, different raters as in inter-rater agreement analysis and multiple test outcomes as in test/retest reliability analysis. For example, in studying changes in the association between biological and psychological measures in posttraumatic stress disorder (PTSD) among trauma survivors over time^[21, 22], it is conceptually more appropriate to use correlation rather than regression to model the changes of association between the two different domains of measures over time without specifying which should be a predictor and which one should be a response. In addition, since changes occur in both the mean association and variance of each variable, it is insufficient to examine the changes in the mean association alone using a regression model. By modeling both changes simultaneously, correlation analysis has an analytic advantage over regression in this respect.

Clustered data can also arise in cross-sectional study designs. The most common instance involving such data in psychosocial research is test/retest reliability and inter-rater agreement analyses, where outcomes from multiple questions and raters form clustered data^{[23]–[25]}. As another example, consider a problem of investigating which of the popular depression measures, Hamilton rating scale for depression (HAMD)^[26] and Beck depression inventory (BDI)^[27], is more specific to the diagnosis of PTSD, in terms of their correlations with a PTSD symptom scale (PSS) in PTSD research^[22]. This problem is easily phrased within the context of correlation analysis. However, a regression analysis is problematic. Conceptually, it is difficult to interpret regression models within such a setting. Analytically, one can immediately think of at least three regression models: (1) regress PSS on HAMD; (2) regress

PSS on BDI; and (3) regress PSS on HAMD and BDI, simultaneously. Unfortunately, none of these single regressions can be used to address the issue of interest.

In this paper, we develop methods of power and sample size estimation for correlation analysis within the context of a clustered study design. In Section 2, we address the analytical challenges in developing such methods. In particular, we discuss a surrogacy-type assumption to reduce the complexity involved in constructing the power function. In Section 3, we illustrate the approach with a real study example and in addition, examine the impact of this assumption on power and sample size estimates using simulated data. In Section 4, we give our concluding remarks.

2 Power for Correlation Analysis with Clustered Study Design

We consider two approaches: one based on the asymptotic and the other on the exact distribution of the Pearson-type estimates. The former typically requires a relatively large sample size, while the latter yields valid power estimates for any sample size and thus provides a benchmark for evaluating the performance of the former under small and moderate sample sizes.

2.1 Asymptotic Distribution of Correlated Correlation Estimates

Consider a clustered study with n clusters, each of size m . As in the literature, we assume independence across clusters. Let $\{(x_{it}, y_{it}); 1 \leq i \leq n; 1 \leq t \leq m\}$ denote a set of paired variables from the i th cluster. Within our context, clustered data arise from longitudinal study designs or multivariate responses in cross-sectional studies, with clusters formed by individual subjects. We are interested in examining the changes in correlations, $\rho_t = \text{Corr}(x_{it}, y_{it})$, between x_{it} and y_{it} over t .

This general setup applies to both longitudinal and cross-sectional study designs. For example, if n denotes the number of subjects and m the number of assessments in a longitudinal study, then ρ_t represent time-dependent correlations. As another example for a clustered design in a cross-sectional study, consider the problem of assessing which of the depression measures, HAMD and BDI, is better correlated with PSS. Let u , v and w denote HAMD, BDI and PSS, respectively. Let $\rho_1 = \text{Corr}(u, w)$ and $\rho_2 = \text{Corr}(v, w)$. The problem then is simply test $H_0 : \rho_1 = \rho_2$. To couch this inference problem within the general

setup, simply let:

$$m = 2, \quad x_{i1} = u_i, \quad x_{i2} = v_i, \quad y_{i1} = y_{i2} = w_i.$$

For convenience, we refer to t as time and m as total number of assessments throughout the rest of the discussion.

Let $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_m)^\top$. In this paper, we consider the class of general linear hypotheses of the form:

$$H_0 : K\boldsymbol{\rho} = \mathbf{a} \quad \text{vs.} \quad H_a : K\boldsymbol{\rho} = \mathbf{b} \neq \mathbf{a}, \quad (1)$$

where K is some $g \times m$ full rank ($g \leq m$) matrix of known constants and \mathbf{a} (\mathbf{b}) is a $m \times 1$ vector of known constants. Note that for data analysis, H_a can be simply stated as $H_a : K\boldsymbol{\rho} \neq \mathbf{a}$. However, for power analysis, \mathbf{b} must be specified in order to compute power. The linear hypothesis in (1) is the most general class of hypotheses that has been considered for the linear and generalized linear models and applies to virtually all types of hypotheses concerning $\boldsymbol{\rho}$ arising in practice. For example, let

$$m = 2, \quad K = (1, -1), \quad \mathbf{a} = 0, \quad \mathbf{b} = 0.2.$$

Then, (1) becomes: $H_0 : \rho_1 - \rho_2 = 0$ vs. $H_a : \rho_1 - \rho_2 = 0.2$. When viewed in the context of the cross-sectional example just discussed, the null tests whether HAMD and BDI have the same correlation with PSS.

The Pearson correlation is an estimate of $\boldsymbol{\rho}$ and its asymptotic normal distribution has been used extensively for inference about $\boldsymbol{\rho}$ ^{[28]–[35]}. Let $\mathbf{r} = (r_1, r_2, \dots, r_m)^\top$, with r_t denoting the Pearson estimate of ρ_t ($1 \leq t \leq m$). Then, under a multivariate normal assumption for $\mathbf{z} = (z_1, \dots, z_m)^\top$, with $z_t = (x_t, y_t)^\top$, the asymptotic normal distribution of \mathbf{r} is given by:

$$\sqrt{n}(\mathbf{r} - \boldsymbol{\rho}) \rightarrow_d N(0, \Sigma_r = [\sigma_{st}]), \quad (2)$$

$$\sigma_{st} = \begin{cases} (1 - \rho_s^2)^2 & \text{if } s = t; \\ \frac{1}{2}\rho_s\rho_t(\rho_{xst}^2 + \rho_{yst}^2 + \rho_{xyt}^2 + \rho_{xyts}^2) + (\rho_{xst}\rho_{yst} + \rho_{xyt}\rho_{xyts}) - & \text{if } s \neq t. \\ -\rho_s(\rho_{xst}\rho_{xyt} + \rho_{yst}\rho_{xyts}) - \rho_t(\rho_{xst}\rho_{xyts} + \rho_{yst}\rho_{xyt}) & \end{cases}$$

where \rightarrow_d denotes convergence in distribution (e.g. Serfling 1980, Chap1^[36]) and

$$\rho_{xyt} = \text{Corr}(x_{is}, y_{it}), \quad \rho_{xst} = \text{Corr}(x_{is}, x_{it}), \quad \rho_{yst} = \text{Corr}(y_{is}, y_{it}), \quad 1 \leq s, t \leq m. \quad (3)$$

To use this asymptotic distribution, we must estimate the parameters in the asymptotic variance Σ_r . Let $P_g = [\rho_{gst}]$ ($g = x, y$) denote the $m \times m$ matrix defined by the within-variable correlations (WVC), ρ_{gst} , and $P_{xy} = [\rho_{xyst}]$ the $m \times m$ matrix defined by the cross-variable correlations (CVC), ρ_{xyst} . Then, Σ_r is a function of the elements of these matrices. In addition, $\boldsymbol{\rho}$ corresponds to the diagonal of P_{xy} . With real data as in data analysis, consistent estimates of P_g and P_{xy} are readily computed. When no data is available, such estimates cannot be calculated and other alternative estimates such as guesses from other similar studies must be considered. Thus, for power analysis, it is desirable to reduce the nuisance parameters as much as possible.

In most applications, elements of WVC are typically modeled as a function of $|t - s|$ and a wide class of analytic functions is available for modeling $\rho_{gst} = \rho_g (|t - s|)^{[2, 37]}$. For example, under the uniform compound symmetry assumption^[2], $\rho_{gst} = \rho_g$ for all $s \neq t$ with $-1 < \rho_g < 1$. Such an assumption may over-simplify the correlation structure, but other alternatives are available to reach a compromise between practical utility and analytic tractability, especially when partial information is available about the variables. Modeling such correlation structure has been extensively discussed in the literature and is not repeated here.

For elements of the CVC P_{xy} , the problem unfortunately is much more complicated, and in particular, is not readily addressed by existing methods such as the approach used for modeling the WVC. For example, consider a simple case with $m = 2$. If P_{xy} is modeled using a uniform compound symmetry assumption, then the null hypothesis $H_0 : \rho_1 = \rho_0, \rho_2 = \frac{1}{2}\rho_0$ would be contradictory to the assumed model, while $H_0 : \rho_1 = \rho_2 = \rho_0$ is implied by such a model ($-1 < \rho_0 < 1$). On the other hand, the unstructured model with no constraint imposed on the elements of P_{xy} would involve too many parameters to be practically useful in most applications. Thus, to use the asymptotic distribution for power analysis, it is desirable to develop an approach that not only reduces the number of parameters in P_{xy} , but also ensures this internal consistency with respect to the hypotheses for $\boldsymbol{\rho}$.

Note that Fisher's z transformation is often applied to improve the accuracy of the asymptotic distribution^[16, 17, 38, 39]. Within our context, this transformation, $T_F(r) = \frac{1}{2} \ln \frac{1+r}{1-r}$, is applied to each component of \mathbf{r} , i.e., $T_F(\mathbf{r})$ is defined as a vector:

$$T_F(\mathbf{r}) = \left(\frac{1}{2} \ln \frac{1+r_1}{1-r_1}, \dots, \frac{1}{2} \ln \frac{1+r_m}{1-r_m} \right)^\top. \quad (4)$$

By applying the Delta method^[40], we obtain the asymptotic distribution of $T_F(\mathbf{r})$:

$$\begin{aligned} \sqrt{n}[T_F(\mathbf{r}) - T_F(\boldsymbol{\rho})] &\rightarrow_d N\left(0, \Sigma_{T_F} = D_\rho \Sigma_r D_\rho^\top\right), \\ D_\rho &= \frac{\partial}{\partial r} \Big|_{r=\rho} T_F(\mathbf{r}) = \text{diag}\left(\left(1 - \rho_t^2\right)^{-1}\right), \end{aligned} \quad (5)$$

where $\text{diag}(a_t)$ denotes the diagonal matrix with a_t on the t th diagonal. For notational brevity, the asymptotic distribution (2) will be used in the following discussion. However, all calculations in Section 3 were based on the transformation in (4) and asymptotic results in (5).

2.2 Surrogacy Assumption for Simplifying Asymptotic Variance

Consider the assumption:

$$[y_t \mid x_t, x_s] = [y_t \mid x_t], \quad 1 \leq s \neq t \leq m, \quad (6)$$

where $[y \mid x]$ denotes the conditional distribution of y given x . Although similar in appearance, the above is not a Markov assumption in stochastic processes^[41], since the conditional distribution involves different variables rather than the same variable at different times. This condition is widely known as a surrogacy assumption in measurement error, or errors in variable, models literature^{[42]–[44]}. Within the current context, (6) stipulates that x_s contains no additional information that could be predictive of y_{it} once x_t is known. Thus, x_s can be viewed as a surrogate of x_t in predicting y_t .

Figure 1 goes about here.

Under (6), we can express the elements of CVC in terms of the parameters of interest, ρ_t , and the elements of the WVC matrices by assuming a linear regression model. Let

$$y_{it} = \beta_{0t} + \beta_{1t}x_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.} \left(0, \sigma_{\epsilon t}^2\right), \quad 1 \leq i \leq n, 1 \leq t \leq m, \quad (7)$$

where (μ, σ^2) denotes a distribution with mean μ and variance σ^2 . Let σ_{gt} denote the standard deviation of g_t for $g = x, y$. Then, since $\rho_t = \beta_{1t}\sigma_{xt}\sigma_{yt}^{-1}$, it follows from (6) and the law of iterated expectations that^[40]

$$\rho_{xyst} = \sigma_{xs}^{-1}\sigma_{yt}^{-1} E[(x_s - \mu_{xs})(y_t - \mu_{yt})] = \beta_{1t}\sigma_{xt}\sigma_{yt}^{-1}\rho_{xst} = \rho_t\rho_{xst}. \quad (8)$$

The above computations are diagrammatically illustrated in Figure 1; ρ_{xyt} (dotted line) is calculated as a function of ρ_t (vertical line) and ρ_{xst} (horizontal line). Thus, given P_x, P_y and $\boldsymbol{\rho}$, P_{xy} is completely determined by (8) and is consistent with the hypotheses for $\boldsymbol{\rho}$. Note that in (7), x_{it} is treated as an independent and y_{it} a dependent variable. Similar results are obtained by reversing the roles of the two variables. In practice, we only need to assume one of the models. For convenience, we assume (7) throughout the rest of the discussion.

Under (8), the asymptotic variance of the limiting normal (2) reduces to:

$$\sigma_{st} = \begin{cases} (1 - \rho_s^2)^2 & \text{if } s = t \\ \frac{1}{2}\rho_s\rho_t(\rho_{xst}^2 + \rho_t^2\rho_{xst}^2 + \rho_s^2\rho_{xst}^2 + \rho_{yst}^2) + \rho_{xst}(\rho_{yst} + \rho_s\rho_t\rho_{xst}) - & \text{if } s \neq t \\ -\rho_s\rho_{xst}(\rho_{xst} + \rho_{yst})(\rho_s + \rho_t) & \end{cases} \quad (9)$$

If P_x and P_y in addition follow a uniform compound symmetry structure, i.e., $P_g = C(\rho_g)$ ($g = x, y$), then (9) further simplifies to:

$$\sigma_{st} = \begin{cases} (1 - \rho_s^2)^2 & \text{if } s = t \\ \frac{1}{2}\rho_s\rho_t(\rho_x^2 + \rho_t^2\rho_x^2 + \rho_s^2\rho_x^2 + \rho_y^2) + \rho_x(\rho_y + \rho_s\rho_t\rho_x) - \rho_s\rho_x(\rho_x + \rho_y)(\rho_s + \rho_t) & \text{if } s \neq t \end{cases} \quad (10)$$

By imposing the surrogacy condition (6) under the linear model (7), we are able to derive (8), which greatly reduces the number of parameters in the asymptotic variance (2). Although the reverse in general is not true, i.e., not all linear regression models satisfy (6), it can be shown (Appendix A) that a large class of linear models does. We discuss some members of this class of models and their implementations within the context of power analysis in Section 3.

2.3 Power Functions based on Asymptotic and Empirical Distributions

Using the asymptotic distribution in (2), power functions for the class of hypotheses (1) are readily constructed. For example, consider testing for a positive difference in correlation between two time points s and t ($1 \leq s, t \leq m$):

$$H_0 : \rho_s = \rho_t = \rho_0, \quad \text{vs.} \quad H_a : \rho_s = \rho_0 + \delta, \quad (11)$$

where $\delta > 0$. Using the difference statistic $D = r_s - r_t$, the power function is given by:

$$\varphi_a = \Pr [D \geq c_{1-\alpha} \mid H_a] = 1 - \Phi \left(z_{1-\alpha} - \frac{\delta}{\sqrt{\sigma_\delta^2}} \right), \quad \sigma_\delta^2 = \frac{1}{n} (\sigma_{ss} + \sigma_{tt} - 2\sigma_{st}), \quad (12)$$

where Φ and z_α denote the cdf and α -percentile of the standard normal distribution. As a different example, consider the power for detecting an increase in correlation across all time points, i.e.,

$$H_0 : \rho_t = \rho_0 \quad \text{vs.} \quad H_a : \rho_t = \rho_0 + \delta > \rho_0, \quad 1 \leq t \leq m, \quad (13)$$

where $\delta > 0$. An appropriate statistic is $T = \frac{1}{m} \sum_{t=1}^m r_t$ and the associated power function is given:

$$\varphi_a = \Pr [D \geq c_{1-\alpha} \mid H_a] = 1 - \Phi \left(z_{1-\alpha} - \frac{\delta}{\sqrt{\sigma_\delta^2}} \right), \quad \sigma_\delta^2 = \frac{1}{nm^2} \mathbf{1}_m^\top \Sigma_r \mathbf{1}_m, \quad (14)$$

where $\mathbf{1}_m$ denotes a column vector of 1's.

More generally, for the class of linear hypotheses (1), it follows from the asymptotic theory (e.g. Serfling, 1980^[36]) and (2) that $K\mathbf{r} - \mathbf{a}$ has a limiting normal distribution under H_0 and H_a :

$$H_0 : \sqrt{n}(K\mathbf{r} - \mathbf{a}) \rightarrow_d N(\mathbf{0}, K\Sigma_r K^\top), \quad H_a : \sqrt{n}(K\mathbf{r} - \mathbf{b}) \rightarrow_d N(\mathbf{0}, K\Sigma_r K^\top). \quad (15)$$

Thus, under H_0 and H_a , the quadratic quantity, $Q_n^2 = n(K\mathbf{r} - \mathbf{a})^\top (K\Sigma_r K^\top)^{-1} (K\mathbf{r} - \mathbf{a})$, has approximately a central and non-central χ^2 distribution:

$$H_0 : Q_n^2 \sim \chi_g^2(0), \quad H_a : Q_n^2 \sim \chi_g^2(c), \quad (16)$$

where $c = n\boldsymbol{\delta}^\top (K\Sigma_\beta K^\top)^{-1} \boldsymbol{\delta}$ ($\boldsymbol{\delta} = \mathbf{b} - \mathbf{a}$) and $\chi_g^2(c)$ denotes a χ^2 distribution with g degrees of freedom and non-centrality parameter c . Now, let $F_{\chi_g^2(c)}$ denote the cdf of $\chi_g^2(c)$. For a given level of type I error α , let p_α denote the α th percentile of $\chi_g^2(0)$. The power function, φ , for the linear hypotheses (1) is given by:

$$\varphi(n, c, \alpha) = 1 - F_{\chi_g^2(c)}(p_{1-\alpha}), \quad \alpha = 1 - F_{\chi_g^2(0)}(p_{1-\alpha}). \quad (17)$$

As an alternative to the asymptotic results, power functions may also be constructed based on the exact sampling distribution of \mathbf{r} . For relatively small sample sizes, such an exact method provides more reliable power/sample estimates, which are useful not only for planning small trials, but also for assessing the accuracy of the asymptotic approach. To this end, let $Z_0 = \{(x_{0it}, y_{0it}); 1 \leq i \leq n, 1 \leq t \leq m\}$ and $Z_a = \{(x_{ait}, y_{ait}); 1 \leq i \leq n, 1 \leq t \leq m\}$ denote two samples of size m under H_0 and H_a , respectively. Let

$$\mathbf{r}_0 = \mathbf{r}(Z_0), \quad \mathbf{r}_a = \mathbf{r}(Z_a), \quad T_0 = T(\mathbf{r}_0), \quad T_a = T(\mathbf{r}_a),$$

which denote the sample correlations and test statistics calculated based on Z_0 and Z_a . For example, for the general linear hypotheses (1), T_0 and T_a are the quadratic statistics under H_0 and H_a , respectively.

Now, let $\mathbf{F}_{\mathbf{r}_0}(\mathbf{r})$ and $\mathbf{F}_{\mathbf{r}_a}(\mathbf{r})$ denote the cdf of \mathbf{r} , and $F_{T_0}(t)$ and $F_{T_a}(t)$ the cdf of the induced measure defined by $T(\mathbf{r}_0)$ and $T(\mathbf{r}_a)$. Then, the exact power function is given by:

$$\varphi_e = \Pr(T(\mathbf{r}) \geq c_{1-\alpha} \mid H_a) = \int I_{\{T_0(\mathbf{r}) \geq c_{1-\alpha}\}} d\mathbf{F}_{\mathbf{r}_a}(\mathbf{r}) = \int_{c_{1-\alpha}}^{\infty} dF_{T_a}(t), \quad (18)$$

where $I_{\{\cdot\}}$ denotes a set indicator and $c_{1-\alpha}$ the $(1 - \alpha)$ th percentile of the distribution $F_{T_0}(t)$, i.e.,

$$1 - \alpha = \Pr(T_0(\mathbf{r}) \geq c_{1-\alpha} \mid H_0) = \int I_{\{T_0(\mathbf{r}) \geq c_{1-\alpha}\}} d\mathbf{F}_{\mathbf{r}_0}(\mathbf{r}) = \int_{c_{1-\alpha}}^{\infty} dF_{T_0}(t). \quad (19)$$

The integral equations in (18) and (19) typically have no analytic solution, even for normally distributed (x_{it}, y_{it}) . A convenient numerical alternative is the Monte Carlo (MC) approximation^[45, 46]. To implement the MC within our context, first generate M independent data sets, $\{Z_0^j, Z_a^k; 1 \leq j, k \leq M\}$, and then set:

$$\mathbf{r}_0^j = \mathbf{r}(Z_0^j), \quad \mathbf{r}_a^k = \mathbf{r}(Z_a^k), \quad T_0^k = T(\mathbf{r}_0^k), \quad T_a^k = T(\mathbf{r}_a^k), \quad 1 \leq j, k \leq M.$$

Now, let $T_0^{(k)}$ denote the order statistic, i.e., $T_0^{(1)} \leq T_0^{(2)} \leq \dots \leq T_0^{(M)}$ and $\tau = [(1 - \alpha)M] + 1$, where $[\cdot]$ denotes the largest integer function. Then,

$$1 - \alpha = \lim_{M \rightarrow \infty} \frac{\tau}{M}, \quad c_{1-\alpha} = \lim_{M \rightarrow \infty} T_0^{(\tau)} = \lim_{\tau \rightarrow \infty} T_0^{(\tau)}, \quad \varphi_e = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M I(T_a^k \geq T_0^{(\tau)}),$$

where the first equality follows from the definition of τ , the second from the asymptotic property of the quantiles^[36] and the last from the convergence of the empirical cdfs^[47]. Thus, for large M , the integral in (18) is approximated by:

$$\varphi_e = \int_{c_{1-\alpha}}^{\infty} dF_{T_a}(t) \approx \frac{1}{M} \sum_{k=1}^M I(T_a^k \geq c_{1-\alpha}). \quad (20)$$

The accuracy of the MC approximation improves as M increases. In addition, the MC sample size M can even be selected to ensure that the MC approximation achieves required accuracy^[48]. These are well known facts and are not further discussed.

Note that evaluation of the exact power function using the empirical distributions requires that the bivariate distribution of (x_t, y_t) be known. In most real studies, this joint distribution is unknown. Further, investigators may be unwilling to assume a specific distribution for modeling (x_t, y_t) . Thus, the exact approach is probably more useful as a tool to evaluate the accuracy of the asymptotic results, rather than as an alternative to the asymptotic approach. In Section 3, we use the MC-based exact approach to assess the accuracy of the asymptotic approach under the surrogacy assumption with normal and non-normal data distributions.

3 Illustrations

In this section, we illustrate the proposed methods with both real and simulated data. As noted earlier, all results are based on the multivariate Fisher's transformation (4). We start with a real study in psychosocial research, which actually motivated the development of the paper.

Example 1. In an intervention study on PTSD patients conducted by investigators at the School of Medicine, University of Pennsylvania, one primary hypothesis is to determine whether the correlation between a biological and psychological measure would decrease from pre- to post-treatment. Let ρ_1 and ρ_2 denote the correlations at pre- and post-treatment. Then, the hypothesis can be expressed in terms of (1) as follows:

$$H_0 : \rho_2 = \rho_1, \quad \text{vs.} \quad H_a : \rho_2 = \rho_1 - \delta, \quad (21)$$

where $\delta > 0$.

Table 1 goes about here.

Shown in Table 1 are the minimum detectable difference δ as a function of the correlation at pre-treatment, ρ_1 , to achieve 80% power based on a one-sided type I error level, $\alpha = 0.05$. The estimates were computed based on (5) and (10) for both an intent-to-treat sample and a completer sample. The intent-to-treat sample is the proposed sample size of the study, while the completer sample is the number of patients who will have completed a minimum number of therapy sessions required by the study. The latter is the analogue of a treatment compliant sample in medication studies.

The estimates in the table show that the detectable differences decrease as either of the within-variable correlation increases. For example, the detectable difference decreases from 0.37 to 0.22 when both ρ_x and ρ_y increase from 0.2 to 0.8. The detectable difference also decreases as the pre-treatment correlation ρ_1 increases. Thus, for this particular study design, these three input parameters have quite a dramatic effect on power.

There are two approximation steps taken in calculating the power estimates in Example 1. The first is in approximating the sampling distribution using the asymptotic result (2), while the second step is in approximating the asymptotic variance by imposing the surrogacy assumption (6). In the next two examples, the performance of the approximation in each step is examined using simulated data.

We first examined the accuracy of the asymptotic result by comparing the power estimates derived based on this distribution with those obtained from the MC-based exact method. Since the result in (8) is derived under a linear regression model, only continuous distributions were considered. We considered several non-normal distributions. In particular, we applied the approach to data simulated under the following four distributions:

$$\begin{aligned}
\text{Normal } (N) & : \mathbf{x} \sim N(\boldsymbol{\mu}_x; \Sigma_x), \quad \boldsymbol{\epsilon} \sim N(0, \Sigma_\epsilon), \quad \mathbf{x} \perp \boldsymbol{\epsilon}, & (22) \\
\text{Normal mixture } (N_{MIX}) & : \mathbf{x} \sim N_{MIX}(-\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \Sigma_1, \Sigma_2), \quad \boldsymbol{\epsilon} \sim N(0, \Sigma_\epsilon), \quad \mathbf{x} \perp \boldsymbol{\epsilon}, \\
\text{Joint } t \ (t_{JT}) & : \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim t\left(\boldsymbol{\mu}_{x\epsilon} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{0} \end{pmatrix}, \Sigma_{x\epsilon} = \begin{pmatrix} \Sigma_x & \mathbf{0} \\ \mathbf{0} & \Sigma_\epsilon \end{pmatrix}, v\right), \\
\text{Independent } t \ (t_{IND}) & : \mathbf{x} \sim t(\boldsymbol{\mu}_x, \Sigma_x, v_x), \quad \boldsymbol{\epsilon} \sim t(\mathbf{0}, \Sigma_\epsilon, v_\epsilon), \quad \mathbf{x} \perp \boldsymbol{\epsilon},
\end{aligned}$$

where $t(\boldsymbol{\mu}, \Sigma, v)$ denotes a multivariate t distribution^[49], $N(\boldsymbol{\mu}; \Sigma)$ a multivariate normal distribution, $N_{MIX}(-\boldsymbol{\mu}_1, \boldsymbol{\mu}_2; \Sigma_1, \Sigma_2)$ a equal mixture of two normals with each normal component having mean and variance $(\boldsymbol{\mu}_k, \Sigma_k)$ ($1 \leq k \leq 2$), and \perp denotes stochastic independence. Compared to the normal distribution, the normal mixture has two modes. In addition, it is generally not symmetric, unless $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\Sigma_1 = \Sigma_2$. For the two t distributions, their tails approach 0 at much a slower rate, yielding heavy tails^{[49]–[52]}. In addition,

although similar in appearance, the parameters for the two t models have quite different interpretations. For t_{JT} , $\Sigma_{x\epsilon}$ no longer has the interpretation of being the var-covariance matrix of the joint distribution (\mathbf{x}, ϵ) . In fact, the var-covariance is given by $\frac{v}{v-2}\Sigma_{x\epsilon}$. For t_{IND} , \mathbf{x} has the variance $\frac{v_x}{v_x-2}\Sigma_x$, while ϵ has the variance $\frac{v_\epsilon}{v_\epsilon-2}\Sigma_\epsilon$. Thus, only for large v , does Σ becomes close to the var-covariance of the respective variables. Finally, it should be noted that while t_{JT} implies that \mathbf{x} and y_t have a joint t distribution^[50], even the marginal y_t does not have a t distribution under t_{IND} ^[52].

For convenience, we focused on the following two hypotheses:

$$\text{Between-time difference} : H_0 : \rho_2 = \rho_1 = 0.5, \quad \text{vs.} \quad H_a : \rho_1 = 0.5, \rho_2 = 0.6, \quad (23)$$

$$\text{Within-time difference} : H_0 : \rho_t = 0.5, \quad \text{vs.} \quad H_a : \rho_t = 0.6, \quad 1 \leq t \leq 5.$$

The first hypothesis tests for change in correlation between two time points, $t = 1, 2$, while the second tests for a constant difference over time defined by five assessments. For convenience, we assumed a normalized variable g in the simulations so that $E(g) = 0$ and $Var(g) = 1$ ($= \frac{v_g}{v_g-2}$) for the normal and normal mixture (for the two t distributions) ($g = x, y$). In addition, we assumed a uniform compound symmetry for both within-variable correlations so that $P_g = C(0.5)$ ($g = x, y$).

Example 2. Data in this example were simulated from the linear regression model (7) under the four distributions in (22). It is readily shown that the surrogacy condition (6) holds true for all these models (Appendix A). In addition, given $\boldsymbol{\rho}$, P_x and P_y , this linear model can be expressed in terms of these input parameters for the normal and two t distributions (Appendix B). For the normal mixture, we set the parameters in (22) as follows:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \omega\mu_x\mathbf{1}_m, \quad \Sigma_1 = \omega^2C(\tilde{\rho}_x), \quad \Sigma_2 = \omega^2\kappa^2C(\tilde{\rho}_x), \quad (24)$$

where $\omega, \mu_x > 0$. By imposing the constraint $Var(\mathbf{x}) = P_x$, it follows (Appendix B) that

$$\omega^2 = \left[\frac{1}{2} (1 + \kappa^2) + \mu_x^2 \right]^{-1}, \quad \tilde{\rho}_x = \frac{2}{1 + \kappa^2} \left\{ \rho_{xst} \left[\frac{1}{2} (1 + \kappa^2) + \mu_x^2 \right] - \mu_x^2 \right\}. \quad (25)$$

As part of the input parameters for the normal mixture, μ_x and κ^2 , have the interpretation of measuring the spread and skewness of the distribution; as μ_x increases, the distribution becomes more variable, while larger values of κ^2 give rise to more skewed distribution.

Thus, the linear models used to simulate the various distributions with the desired P_g and $\boldsymbol{\rho}$ are given by:

$$\begin{aligned}
\mathbf{y} &= D\mathbf{x} + \boldsymbol{\epsilon}, \quad \mathbf{y} = (y_1, \dots, y_m)^\top, \quad \mathbf{x} = (x_1, \dots, x_m)^\top, \quad D = \text{diag}(\rho_t), \quad (26) \\
N &: \quad \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} P_x & \mathbf{0} \\ \mathbf{0} & P_y - DP_x D \end{pmatrix} \right), \\
N_{MIX} &: \quad \mathbf{x} \sim N \left(-\omega\mu_x \mathbf{1}_m, \omega\mu_x \mathbf{1}_m; \omega^2 C(\tilde{\rho}_x), \omega^2 \kappa^2 C(\tilde{\rho}_x) \right), \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, P_y - DP_x D), \\
t_{JT} &: \quad \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim t \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} P_x & \mathbf{0} \\ \mathbf{0} & \frac{v-2}{v} P_y - DP_x D \end{pmatrix}, v \right), \\
t_{IND} &: \quad \mathbf{x} \sim t(\mathbf{0}, P_x, v_x), \quad \boldsymbol{\epsilon} \sim t \left(\mathbf{0}, \Sigma_\epsilon = \frac{v_\epsilon - 2}{v_\epsilon} \frac{v_x}{v_x - 2} \left(\frac{v_x - 2}{v_x} P_y - DP_x D \right), v_x \right).
\end{aligned}$$

In addition, by the surrogacy assumption, the off-diagonal elements of P_{xy} can be expressed in terms of P_x and $\boldsymbol{\rho}$ as in (8).

Figures 2 and 3 go about here.

Shown in Figures 2 and 3 are power curves as a function of sample size for testing the between- and within-time difference hypotheses in (23). The smooth curve was obtained based on the asymptotic result (2) and (5), while the other saw-toothed curves were obtained based on the exact power function implemented with MC approximations using an MC sample size $m = 500$. The jagged patterns of the latter are the result of the sampling variability among the MC samples. Despite the imperfect approximations, trend in each power curve as a function of sample size is clearly discernible. In general, the exact power curves based on the normal distribution are all consistently closer to the asymptotic power curves than either the t - or the normal mixture. The asymptotic power curve tends to overestimate the exact power for the t -distribution, while underestimate it for the normal mixture, especially as the latter distribution becomes more skewed (as measured by larger values of κ^2).

Between the two t distributions, the asymptotic power curves seems to provide a better approximation for the independent than the joint t model. This differential accuracy may be in part due to the different asymptotic behaviors of the estimate, $\hat{\rho}_t$, under the two t models; $\hat{\rho}_t$ has an asymptotic normal distribution under the former, while an asymptotic t distribution under the latter t model^[53]. Thus, while accuracy in approximation improves

for t_{INT} as sample size increases, no such improvement is expected of t_{JT} if the degree of freedom remains small.

We next devised an approach to address the impact of the surrogacy assumption (6). To this end, we considered a class of auto-regressive type models for which the surrogacy assumption fails. By simulating data under this scenario, we are able to assess the robustness of the power estimates against departures from the surrogacy assumption.

Example 3. Consider the following linear model:

$$y_t = \sum_{|l-t| \leq \Delta} \beta_t \zeta^{|l-t|} x_l + \epsilon_t, \quad 0 \leq \zeta < 1, 1 \leq t \leq m, \quad (27)$$

where Δ is a known integer, ζ a known constant and $(\mathbf{x}, \boldsymbol{\epsilon})$ follows one of the four distributions considered in Example 2. The integer, Δ , determines the number of neighboring x terms upon which y_t depends in addition to x_t , i.e.,

$$[y_t | x_1, \dots, x_m] = [y_t | x_{t-\Delta}, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_{t+\Delta}].$$

In the special case when $\Delta = 0$, (27) reduces to (26). Given Δ , ζ indicates the strength in which y_t depends on such neighboring x terms, i.e., the smaller the ζ , the weaker the dependence. By replacing D with G given below,

$$g_{tl} = \begin{cases} \beta_t \zeta^{|l-t|} & \text{if } \max(t, t - \Delta) \leq l \leq \min(t, t + \Delta) \\ 0 & \text{if otherwise} \end{cases}, \quad 1 \leq t, l \leq m, \quad (28)$$

(27) is also readily expressed in the matrix form as in (26) (Appendix C). Note that although similar in appearance, (27) is not an auto-regressive model in the time series literature^[54]. The latter only involves one variable y_t at various times t , but (27) involves two different variables x_t and y_t .

It follows from (27) that (Appendix C):

$$\beta_t = \left(\sum_{l \in \Gamma_t} \zeta^{|l-t|} \rho_{x_t l} \right)^{-1} \rho_t, \quad \Gamma_t = \{l; |l-t| \leq \Delta\}, \quad (29)$$

Thus, unlike the models in Example 2, β_t no longer has the interpretation of being the correlation between x_t and y_t . Further, since

$$\rho_{xyt} = \left(\sum_{l \in \Gamma_t} \zeta^{|l-t|} \rho_{x_t l} \right)^{-1} \left(\sum_{l \in \Gamma_t} \zeta^{|l-t|} \rho_{x_s l} \right) \rho_t \neq \rho_{x_s l} \rho_t, \quad (30)$$

condition (6) fails for this class of models. Note that the expression above shows that the dependence of y_t on the neighboring x terms diminishes as s gets further apart from t . In addition, as $\zeta \rightarrow 0$, $\rho_{xyt} \rightarrow \rho_{xsl}\rho_t$, i.e., model (26) is the asymptote of (27) as $\zeta \rightarrow 0$.

Figure 4 and 5 go about here.

Shown in Figures 4 and 5 are power curves as a function of sample size for testing the two hypotheses (23) based on the exact and asymptotic power functions. Again, the exact power function was implemented with the MC method using a sample size $m = 500$. Since condition (6) fails for these models, two approaches were used to obtain the asymptotic power curves. One was based on (2) and (5), while the other was obtained by imposing the surrogacy assumption in addition to these conditions. Thus, the comparison of the two asymptotic power curves provides information on the robustness of the power estimates obtained under the surrogacy assumption.

As in Example 2, power curves based on the two asymptotic approaches provide good estimates for the normal distribution. Again, the asymptotic approaches overestimate power for the t -distributions, while underestimate it for the normal mixture. Between the two t curves, t_{IND} again is closer to the asymptotic power curves than t_{JT} . As shown in Figures 4 and 5, the power function does not seem to depend much on Δ and ζ of the auto-regressive models, which characterize the degree of departure from the surrogacy assumption (6). Between the two asymptotic power curves, they are very close to each other. Thus, at least for the auto-regressive type models (27), the power estimates under the surrogacy assumption are pretty robust.

4 Discussion

By introducing a surrogacy condition, we have proposed an approach for effectively reducing the parameters in the asymptotic variance of the joint correlated Pearson-type correlation estimates from clustered study designs under the normal data assumption for power analysis. By reducing the input parameters to the minimum, the proposed approach not only improves the robustness of power and sample estimates, but also makes it practical to use in most real studies.

Although the general impact is unknown, our simulation results seem to show that power estimates obtained under the surrogacy assumption are robust against departures from this assumption. Indeed, for the class of hypotheses and models considered in our simulation

study, the difference between the two power estimates obtained with and without this surrogacy assumption is almost negligible for all practical purposes. Thus, the proposed approach may be applied to more general models other than those that satisfy the surrogacy condition. The simulation results also show that the asymptotic distribution of the correlated correlation estimates is not robust against departures from the normal data distribution. Thus, if data are suspected to follow some non-normal distribution, the asymptotic distribution may not be used and power/sample size estimates may be calculated using other alternative methods such as the one based on the exact power function, as discussed in Section 3.

Acknowledgment

This research was supported in part by NIH/NIMH grants P50-MH45178-08 (Crits-Christoph, Gallop and Tu) and R01-DA018935 (Crits-Christoph and Tu), and by an NIH/NIAD award K22AI 51186 (Kowalski).

Appendix

A. We show that the surrogacy condition (6) is satisfied for the class of linear regression models (7) under the distribution assumptions in (22).

For simplicity and without loss of generality, assume $E(x_t) = E(y_t) = 0$ for $1 \leq t \leq m$. In addition, assume $Var(x_t) = 1$ for the two normal models and $Var(x_t) = \frac{v_x}{v_x - 2}$ for the two t -distributions ($1 \leq t \leq m$). Under these assumptions, $\beta_{0t} = 0$. Also, denote β_{1t} by β_t .

By representing t_{INT} in (22) in a hierarchical form^[52], we have:

$$\mathbf{x} | \lambda_x \sim N(\mathbf{0}, \lambda_x P_x), \quad \boldsymbol{\epsilon} | \lambda_\epsilon \sim t(\mathbf{0}, \lambda_\epsilon \Sigma_\epsilon), \quad \lambda_x \sim IG\left(\frac{v_x}{2}, \frac{2}{v_x}\right), \quad \lambda_\epsilon \sim IG\left(\frac{v_\epsilon}{2}, \frac{2}{v_\epsilon}\right),$$

where $IG(\cdot)$ denotes an inverted gamma distribution. Let

$$\tilde{\mathbf{x}}_t = (x_1, \dots, x_{t-1}, x_{t+1}, \dots, x_m)^\top, \quad P_t = Var(\tilde{\mathbf{x}}_t | \lambda_x), \quad (31)$$

Then,

$$\begin{pmatrix} \tilde{\mathbf{x}}_t \\ x_t \\ \boldsymbol{\epsilon} \end{pmatrix} | \lambda_x, \lambda_\epsilon \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_x P_x & \mathbf{0} \\ \mathbf{0} & \lambda_\epsilon \sigma_{\epsilon t}^2 \end{pmatrix}\right), \quad P_x = \begin{pmatrix} P_t & \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_t^\top & 1 \end{pmatrix}. \quad (32)$$

It follows that

$$\begin{pmatrix} \tilde{\mathbf{x}}_t \\ x_t \\ y_t \end{pmatrix} \mid \lambda_x, \lambda_\epsilon \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_x P_t & \lambda_x \boldsymbol{\alpha}_t & \lambda_x \beta_t \boldsymbol{\alpha}_t \\ \lambda_x \boldsymbol{\alpha}_t^\top & \lambda_x & \lambda_x \beta_t \\ \lambda_x \beta_t \boldsymbol{\alpha}_t^\top & \lambda_x \beta_t & \lambda_x \beta_t^2 + \lambda_\epsilon \sigma_{\epsilon t}^2 \end{pmatrix} \right). \quad (33)$$

By the properties of the multivariate normal distribution^[55], we have:

$$\begin{aligned} [y_t \mid \tilde{\mathbf{x}}_t, x_t, \lambda_x, \lambda_\epsilon] &\sim N(\mu_{y|x}, \sigma_{y|x}^2), \quad \mu_{y|x} = \lambda_x (\beta_t \boldsymbol{\alpha}_t^\top, \beta_t) \lambda_x^{-1} P_t^{-1} \begin{pmatrix} \tilde{\mathbf{x}}_t \\ x_t \end{pmatrix} = \beta_t x_t, \\ \sigma_{y|x}^2 &= \lambda_x \beta_t^2 + \lambda_\epsilon \sigma_{\epsilon t}^2 - \lambda_x (\beta_t \boldsymbol{\alpha}_t^\top, \beta_t) \lambda_x^{-1} P_t^{-1} \lambda_x (\beta_t \boldsymbol{\alpha}_t^\top, \beta_t)^\top = \lambda_\epsilon \sigma_{\epsilon t}^2. \end{aligned}$$

Thus, $[y_t \mid \tilde{\mathbf{x}}_t, x_t, \lambda_x, \lambda_\epsilon] = [y_t \mid x_t, \lambda_\epsilon]$ for all $\lambda_x, \lambda_\epsilon > 0$, from which it follows that

$$[y_t \mid \tilde{\mathbf{x}}_t, x_t] = \int [y_t, \lambda_x, \lambda_\epsilon \mid \tilde{\mathbf{x}}_t, x_t] d\lambda_x d\lambda_\epsilon = \int [y_t \mid x_t, \lambda_\epsilon] dF_{\lambda_\epsilon} = [y_t \mid x_t]. \quad (34)$$

Therefore, the surrogacy condition (6) holds true for (7) under t_{INT} .

By setting $\lambda_x = \lambda_\epsilon$ in the above argument, we obtain (34) for t_{JT} and if, in addition, $\lambda_x = \lambda_\epsilon = 1$, (34) holds true for the normal model. Finally, let $Bernoulli(1, -1; \frac{1}{2})$ denote a Bernoulli random variable taking on values of 1 and -1 with equal probability. Assume that $\lambda_\epsilon = 1$ and $\lambda_x \sim Bernoulli(1, -1; \frac{1}{2})$. Then, it follows immediately from (32) and (33) that (34) holds true for the mixture normal N_{MIX} .

B. For t_{INT} , let:

$$\mathbf{x} \sim t(\mathbf{0}, P_x, v_x), \quad \boldsymbol{\epsilon} \sim t(\mathbf{0}, \Sigma_\epsilon, v_\epsilon),$$

it follows from A and the properties of the multivariate t distribution that^[53]:

$$\begin{aligned} Var(\mathbf{x}) &= \frac{v_x}{v_x - 2} P_x, \quad Var(\boldsymbol{\epsilon}) = \frac{v_\epsilon}{v_\epsilon - 2} \Sigma_\epsilon, \quad E(x_s, y_t) = \frac{v_x}{v_x - 2} \beta_t \rho_{xst}, \quad (35) \\ E(y_s, y_t) &= \frac{v_x}{v_x - 2} \left(\beta_s \beta_t \rho_{xst} + \frac{v_x - 2}{v_x} \frac{v_\epsilon}{v_\epsilon - 2} \sigma_{\epsilon st} \right). \end{aligned}$$

By setting:

$$\beta_t^2 + \frac{v_x - 2}{v_x} \frac{v_\epsilon}{v_\epsilon - 2} \sigma_{\epsilon t}^2 = 1, \quad (36)$$

it follows from (35) that $\beta_t = \rho_t$ and $\rho_{xyt} = \rho_t \rho_{xst}$, and thus,

$$P_y = \text{Corr}(\mathbf{y}, \mathbf{y}) = \frac{v_x}{v_x - 2} \left(DP_x D + \frac{v_x - 2}{v_x} \frac{v_\epsilon}{v_\epsilon - 2} \Sigma_\epsilon \right), \quad (37)$$

from which we obtain the expression for t_{INT} in (26).

For t_{JT} , it is readily checked that (36)-(37) reduce to

$$\beta_t^2 + \sigma_{\epsilon t}^2 = 1, \quad \beta_t = \rho_t, \quad \rho_{xyt} = \rho_t \rho_{xst}, \quad P_y = \frac{v_x}{v_x - 2} (DP_x D + \Sigma_\epsilon). \quad (38)$$

The result for t_{JT} thus follows. By letting $v_x = v_\epsilon \rightarrow \infty$ in (36)-(37), we obtain the expression for the normal distribution.

For N_{MIX} , we first represent this distribution in a hierarchical form as follows:

$$\mathbf{x} | \lambda_x \sim N(\lambda_x \boldsymbol{\theta}_x, \Phi_{\lambda_x x}), \quad \boldsymbol{\epsilon} \sim N(0, \Sigma_\epsilon), \quad \lambda_x \sim \text{Bernoulli}\left(1, -1; \frac{1}{2}\right).$$

Then, it is readily checked that:

$$\begin{aligned} E(\mathbf{x}) &= E(\boldsymbol{\epsilon}) = 0, \quad P_x = \text{Corr}(\mathbf{x}, \mathbf{x}) = \frac{1}{2} (\Phi_{-x} + \Phi_x) + \boldsymbol{\theta}_x \boldsymbol{\theta}_x^\top, \quad (39) \\ E(x_s, y_t) &= \beta_t \left[\frac{1}{2} (\phi_{-xst} + \phi_{xst}) + \theta_{xs} \theta_{xt} \right], \quad E(y_s, y_t) = \left[\beta_s \beta_t \left(\frac{1}{2} (\phi_{-xst} + \phi_{xst}) + \theta_{xs} \theta_{xt} \right) + \sigma_{\epsilon st} \right]. \end{aligned}$$

Thus,

$$\text{Corr}(x_s, y_t) = \beta_t \frac{\frac{1}{2} (\phi_{-xst} + \phi_{xst}) + \theta_{xs} \theta_{xt}}{\sqrt{\beta_t^2 + \sigma_{\epsilon t}^2}} = \frac{\beta_t \rho_{xst}}{\sqrt{\beta_t^2 + \sigma_{\epsilon t}^2}}. \quad (40)$$

By setting $\beta_t^2 + \sigma_{\epsilon t}^2 = 1$, we obtain from (40) that $\rho_{xyt} = \beta_t \rho_{xst}$ and thus under the surrogacy assumption, $\beta_t = \rho_t$. Finally, it follows from (39) that

$$\rho_{yzt} = \beta_s \beta_t \rho_{xst} + \sigma_{\epsilon st},$$

from which the expression in (26) follows.

Under the assumption (24), it is readily checked that

$$\omega^2 \left[\frac{1}{2} (1 + \kappa^2) + \mu_x^2 \right] = \rho_{xtt} = 1, \quad \omega^2 \left[\frac{\tilde{\rho}_x}{2} (1 + \kappa^2) + \mu_x^2 \right] = \rho_{xst}.$$

(25) then follows by solving for ω^2 and $\tilde{\rho}_x$ from the above set of equations.

C. For t_{INT} , we have $E(x_s, y_t) = \beta_t \frac{v_x}{v_x - 2} \sum_{|l-t| \leq \Delta} \zeta^{|l-t|} \rho_{xsl}$. By setting:

$$\beta_t^2 \sum_{|l-t| \leq \Delta, |k-t| \leq \Delta} \zeta^{|k-t|+|l-t|} \rho_{xkl} + \frac{v_x - 2}{v_x} \frac{v_\epsilon}{v_\epsilon - 2} \sigma_t^2 = 1,$$

we obtain

$$\rho_{xyst} = \beta_t \left(\sum_{|l-t| \leq \Delta, |k-t| \leq \Delta} \zeta^{|l-t|} \rho_{xsl} \right), \quad \Sigma_\epsilon = \frac{v_x}{v_x - 2} \frac{v_\epsilon - 2}{v_\epsilon} \left(\frac{v_x - 2}{v_x} P_y - G P_x G^\top \right).$$

The cases for t_{JT} , the normal and normal mixture are obtained by following a similar argument.

References

- [1] Dawson, J.D. (1998). ‘Sample size calculations based on slopes and other summary statistics,’ *Biometrics*, 54: 323-30.
- [2] Diggle, P.J., Heagerty, P.J.; Liang, K.Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data*, 2nd. ed., Oxford University Press, New York.
- [3] Hedeker, D., Gibbons R.D. and Waternaux, C. (1999). ‘Sample Size Estimation for Longitudinal Designs with Attrition: Comparing Time-related Contrasts between Two Groups,’ *J. Educational and Behavioral Statistics*, 24, 70-93.
- [4] Lee, J.W. and DeMets, D.L. (1991). ‘Sequential comparison of changes with repeated measurements data. *Journal of the American Statistical Association*, 86, 757-762.
- [5] Lefante, J.J. (1990). ‘The power to detect differences in averages rates of change in longitudinal studies,’ *Statistics in Medicine*, 9: 437-446.
- [6] Liu, G. and Liang, K.Y. (1997). ‘Sample size calculations for studies with correlated observations,’ *Biometrics*, 53: 937-947.
- [7] Liu, A., Shih, W.J. and Gehan, E. (2002). ‘Sample size and power determination for clustered repeated measurements,’ *Statistics in Medicine*, 21: 1787-801.

- [8] Maxwell, S. (2000). 'Sample size and multiple regression analysis,' *Psychological Methods*, 5, 434-458.
- [9] Muller, K.E., LaVange, L.M., Ramey, S.L. and Ramey, C.T. (1992). 'Power calculations for general linear multivariate models including repeated measures applications,' *Journal of the American Statistical Association*, 87, 1209-1226.
- [10] Overall, J.E., Shobaki, G. and Anderson, C.B. (1998). 'Comparative evaluation of two models for estimating sample sizes for tests on trends across repeated measurements,' *Controlled Clinical Trials*, 19: 188-97.
- [11] Rochon, J. (1991). 'Sample Size Calculations for Two-group, Repeated Measures Experiments,' *Biometrics*, 47, 1383-1398.
- [12] Rochon, J. (1998). 'Application of GEE procedures for sample size calculations in repeated measures experiments,' *Statistics in Medicine*, 17: 1643-1658.
- [13] Schouten, H.J. (1999). 'Planning group sizes in clinical trials with a continuous outcome and repeated measures,' *Statistics in Medicine*, 18: 255-64.
- [14] Tu, X.M., Kowalski, J., Zhang, J., Lynch, K. and Crits-Christoph, P. (2004). 'Power Analysis for Clinical Trials and Other Clustered Study Design,' *Statistics in Medicine*, 23, 2799-2815.
- [15] Wu, M.C. (1988). 'Sample Size for Comparison of Changes in the Presence of Right Censoring Caused by Death, Withdrawal, and Staggered Entry,' *Controlled Clinical Trials*, 9, 32-46.
- [16] Dunn, O.J. and Clark, V.A. (1969). 'Correlation Coefficients Measured on the Same Individuals,' *Journal of the American Statistical Association*, 64, 366-377.
- [17] Dunn, O.J. and Clark, V.A. (1971). 'Comparison of Tests of the Equality of Dependent Correlation Coefficients,' *Journal of the American Statistical Association*, 66, 904-908.
- [18] Steiger, J.H. (1980). 'Tests for Comparing Elements of a Correlation Matrix,' *Psychological Bulletin*, 87, 245-251.
- [19] Olkin, I. and Finn, J. (1990). 'Testing Correlated Correlation,' *Psychological Bulletin*, 108, 330-333.

- [20] Meng, X.L., Rosenthal, R. and Rubin, D.B. (1992). 'Comparing Correlated Correlation Coefficients,' *Psychological Bulletin*, 111, 172-175.
- [21] Foa, E. B., Dancu, C. V., Hembree, E. A. et al. (1999a). 'A Comparison of Exposure Therapy, Stress Inoculation Training and their Combination for Reducing PTSD in Female Assault Victims,' *Journal of Consulting and Clinical Psychology*, 70, 715-723.
- [22] Foa, E. B., Ehlers, A., Clark, D., & Tolin, D. F. (1999b). 'Posttraumatic Cognitions Inventory (PTCI): Development and Comparison with other Measures,' *Psychological Assessment*, 11, 303-314.
- [23] Schroder, K.E.E., Carey, M.P. and Vanable, P.A. (2003). 'Methodological Challenges in Research on Sexual Risk Behavior: II. Accuracy of Self-reports,' *Annals of Behavioral Medicine*, 26, 104-123.
- [24] Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions*, Wiley, New York.
- [25] Rosenthal, R. and Rosnow, R.L. (1991). *Essentials of Behavioral Research Methods and Data Analysis*, McGraw-Hill, MA.
- [26] Hamilton, M. (1960). 'A Rating Scale for Depression,' *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.
- [27] Beck, A.T., Epstein, N., Brown, G. and Steer, R.A. (1988). 'An Inventory for Measuring Clinical Anxiety: Psychometric Properties,' *Journal of Consulting and Clinical Psychology*, 56, 893-897.
- [28] Boik, R.J. (1998). 'A local parameterization of orthogonal and semi-orthogonal matrices with applications,' *Journal of multivariate Analysis*, 67, 244-276.
- [29] Browne, MW. and Shapiro, A. (1986). 'The asymptotic covariance matrix of sample correlation coefficients under general conditions,' *Linear Algebra and its Applications*, 82, 169-176.
- [30] Elston, R.C. (1975). 'On the correlation between correlation,' *Biometrika*, 62, 133-140.
- [31] Hsu, P.L. (1949). 'The limiting distribution of functions of sample means and applications to testing hypotheses,' *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, 359-402.

- [32] Kollo, T. and Neudecker, H. (1994). ‘Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices,’ *Journal of Multivariate Analysis*, 47, 283-300.
- [33] Neudecker, H. and Wesselman, A.M. (1990). ‘The asymptotic distribution of the sample correlation matrix,’ *Linear Algebra and its Applications*, 127, 589-599.
- [34] Olkin, I. and Siotani, M. (1976). ‘Asymptotic Distribution of Functions of a Correlation Matrix,’ In S.Ikeda (editor), *Essays in Probability and Statistics*, Chapter 16 (pp.235-251), Shinko Tsusho, Tokyo, Japan.
- [35] Steiger, J.H. & Hakstian, A.R. (1982). ‘The asymptotic distribution of elements of a correlation matrix: theory and applications,’ *British Journal of Mathematical and Statistical Psychology*, 35, 208-215.
- [36] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- [37] Jennrich, R.I. and Schluchter, M.D. (1986). ‘Unbalanced Repeated Measures Models with Structured Covariance Matrices,’ *Biometrics*, 65, 1582-1589.
- [38] Machin, D., Campbell, M.J., Fayers, P.M. and Pinol, A.P.Y. (1997). *Sample size tables for clinical studies*, Blackwell Science, Oxford.
- [39] Mudholkar, G.S. (1983). ‘Fisher’s z -Transformation,’ in *Encyclopedia of Statistical Sciences*, Vol. 3, pp 130-135, Ed. Samuel, K. and Johnson, N.L., Wiley, New York.
- [40] Billingsley, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.
- [41] Hoel, P.G., Port, S.C. and Stone C.J. (1987). *Introduction to Stochastic Processes*, Waveland Press, Prospect Heights, IL.
- [42] Fuller WA. (1987). *Measurement Error Models*. Wiley, New York.
- [43] Liang, K.Y. and Liu, X.H. (1991). ‘Estimating Equations in Generalized Linear Models with Measurement Error,’ In *Estimating Functions*, edited by Godambe, V.P., Oxford: Oxford University Press, pp.47-63.

- [44] Kowalski, K. and Tu, X.M. (2001). ‘A Generalized Estimating Equation Approach to Modeling Incompatible Data Formats with Covariate Measurement Error: Application to Human Immunodeficiency Virus Immune Marker,’ *Journal of Royal Statistical Society, Series C*, 51, 91-114.
- [45] Kennedy, W.J. and Gentle, J.E. (1980). *Statistical Computing*, Marcel and Dekker, New York.
- [46] Tanner, M.A. (1993). *Tools for Statistical Inference — Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag, New York.
- [47] Chung, K.L. (1974). *A Course in Probability Theory*, 2 ed, Academic Press, CA.
- [48] Geweke, J. (1989). ‘Bayesian Inference In Econometric Models Using Monte Carlo Integration,’ *Econometrica*, 57, 1317-1339.
- [49] Maronna, R.A. (1976). ‘Robust M-estimators of Multivariate Location and Scatter,’ *Annals of Statistics*, 4, 51-67.
- [50] Zellner, A. (1976). ‘Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms,’ *Journal of the American Statistical Association*, 71, 400-405.
- [51] Liu, C.H. and Rubin, D.B. (1994). ‘The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence,’ *Biometrika*, 81, 633-648.
- [52] Kowalski J., Tu X.M., Day R.S. and Mendoza-Blanco J.R. (1997) ‘On the Rate of Convergence of the ECME for Multiple Regression Models with t-distributed Errors,’ *Biometrika*, 84, 269-281.
- [53] Kowalski J., Mendoza-Blanco J.R., Tu X.M. and Gleser L.R. (1999). ‘On the Difference in Inference and Prediction between the Joint and Independent t-error Models for Seemingly Unrelated Regressions,’ *Communications in Statistics, Theory and Methods*, 28, 2119-2140.
- [54] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York.
- [55] Seber, G.A.F. (1984). *Multivariate Observations*, Wiley, New York.

Cross-variable (x, y) correlation at pre-treatment ρ_1 : 0.4 0.6 0.8 0.9												
Within-variable	Within-variable x correlation ρ_x											
y correlation ρ_y	0.2			0.5				0.8				
Intent-to-treat sample ($n = 72$)												
0.2	0.37	0.32	0.22	0.13	0.36	0.31	0.22	0.14	0.35	0.31	0.23	0.17
0.5					0.33	0.28	0.19	0.11	0.29	0.25	0.17	0.12
0.8									0.22	0.19	0.12	0.07
Completer sample ($n = 60$)												
0.2	0.41	0.36	0.25	0.15	0.39	0.35	0.25	0.16	0.38	0.34	0.26	0.21
0.5					0.36	0.31	0.21	0.13	0.32	0.28	0.18	0.10
0.8									0.25	0.21	0.14	0.08

Table 1: Minimum detectable difference δ between pre- and post-treatment to achieve 0.80 power based on a one-sided $\alpha = 0.05$ under a uniform compound symmetry structure for both within-variable correlations P_g ($g = x, y$).

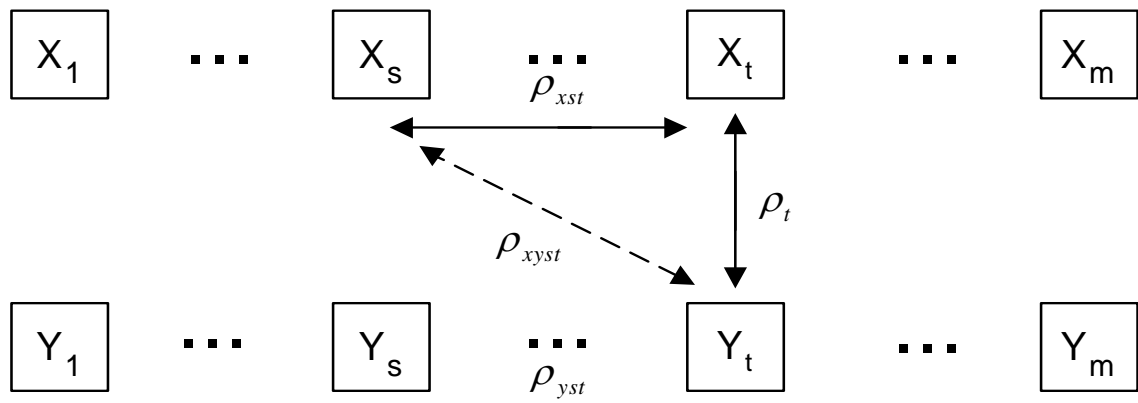


Figure 1: Diagram illustrating the calculation of CVC correlation ρ_{xyst} using WVC correlations ρ_{xst} and ρ_t .

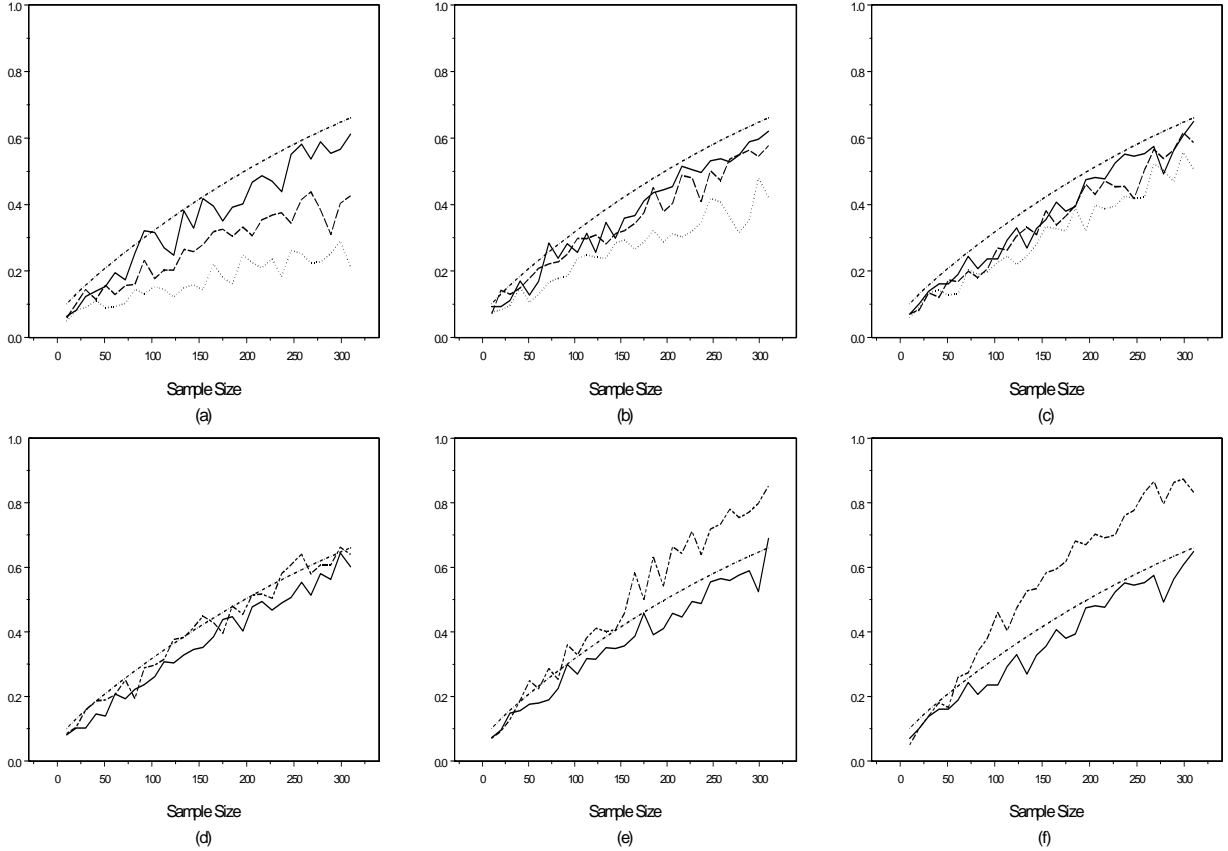


Figure 2: Asymptotic (dot-dashed line) and exact power curves as a function of sample size for testing between-time difference (23) under t - ((a)-(c)) and normal mixture ((d)-(f)) distribution. For t -distributions with $v_e = v_x = v$, (a) $v = 4$; (b) $v = 7$; and (c) $v = 10$. For normal mixture, (d) $\mu = 0.5$, $\kappa^2 = 1$; (e) $\mu = 1$, $\kappa^2 = 2$; and (f) $\mu = 2$, $\kappa^2 = 9$. Exact power curves for t_{INT} (dashed line), t_{JT} (dotted line), N (solid line) and N_{MIX} (short-long dashed line) are based on Monte Carlo method.

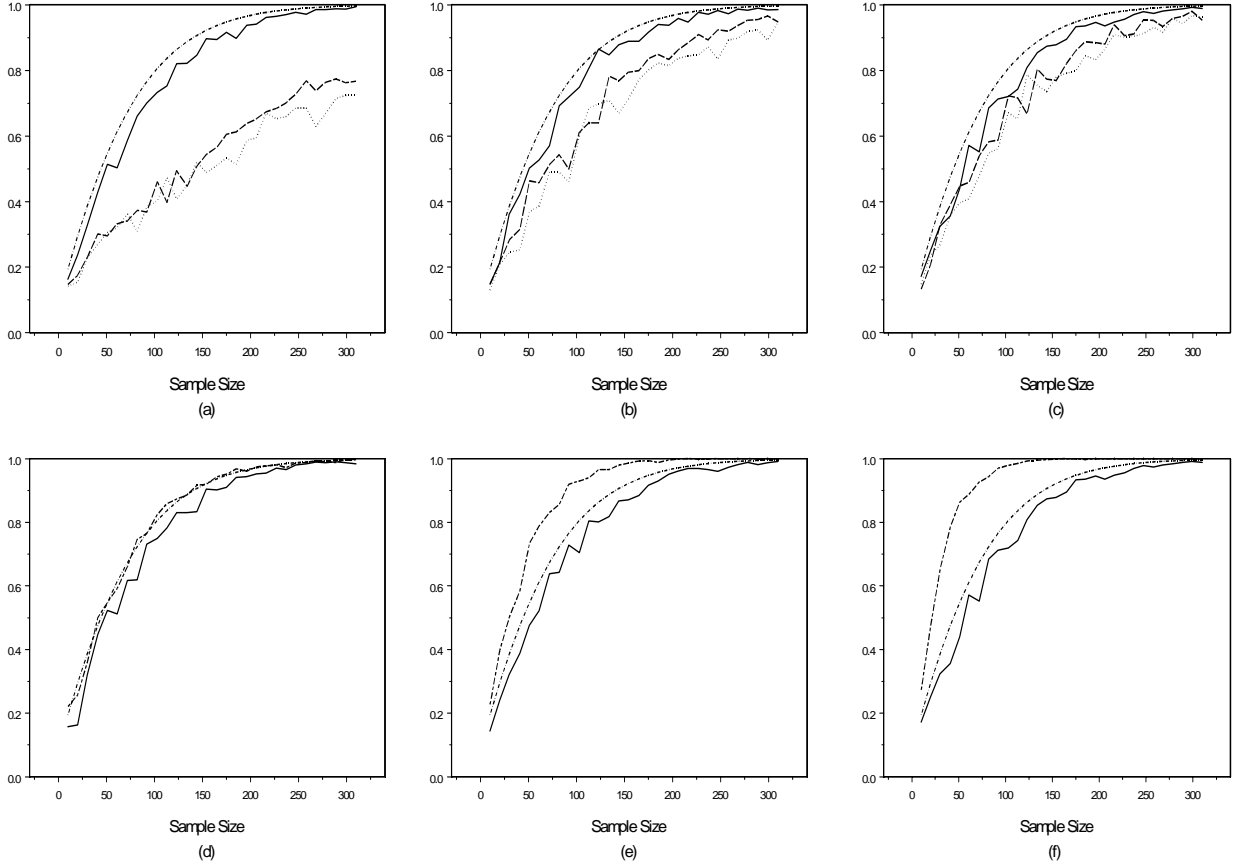


Figure 3: Asymptotic (dot-dashed line) and exact power curves as a function of sample size for testing within-time difference (23) under t - ((a)-(c)) and normal mixture ((d)-(f)) distribution. For t -distributions with $v_\epsilon = v_x = v$, (a) $v = 4$; (b) $v = 7$; and (c) $v = 10$. For normal mixture, (d) $\mu = 0.5, \kappa^2 = 1$; (e) $\mu = 1, \kappa^2 = 2$; and (f) $\mu = 2, \kappa^2 = 9$. Exact power curves for t_{INT} (dashed line), t_{JT} (dotted line), N (solid line) and N_{MIX} (short-long dashed line) are based on Monte Carlo method.

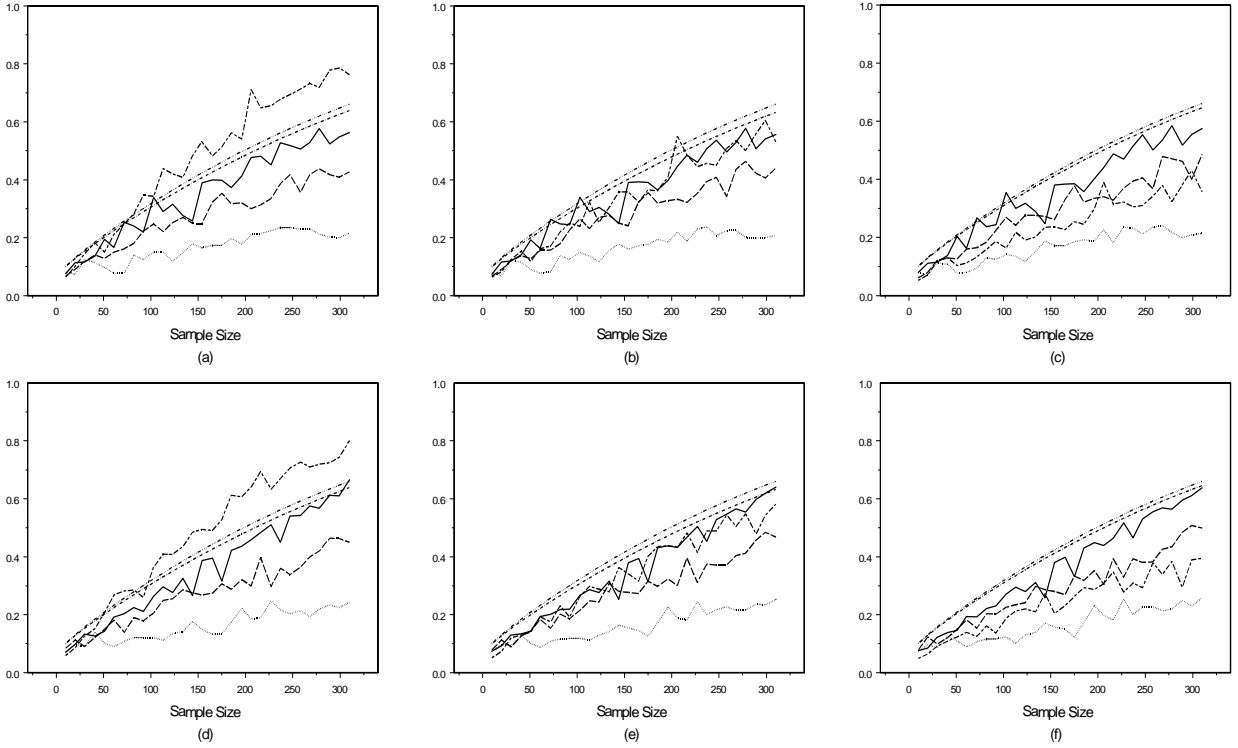


Figure 4: Asymptotic and exact power curves as a function of sample size for testing between-time difference (23) based on auto-regressive model (27) under t - and normal mixture distribution, with (a)-(c) $\Delta = 3$; (d)-(f) $\Delta = 4$; (a) and (d) $\zeta = 0.2$; (b) and (e) $\zeta = 0.5$; and (c) and (f) $\zeta = 0.9$. For t -distributions, $v_\epsilon = v_x = v = 4$ and for normal mixture, $\mu = 2$, $k^2 = 9$. Asymptotic curves are obtained under (dot-dashed line) and without (dots-dashed line) surrogacy condition, while exact power curves for t_{INT} (dashed line), t_{JT} (dotted line), N (solid line) and N_{MIX} (short-long dashed line) are based on Monte Carlo method.

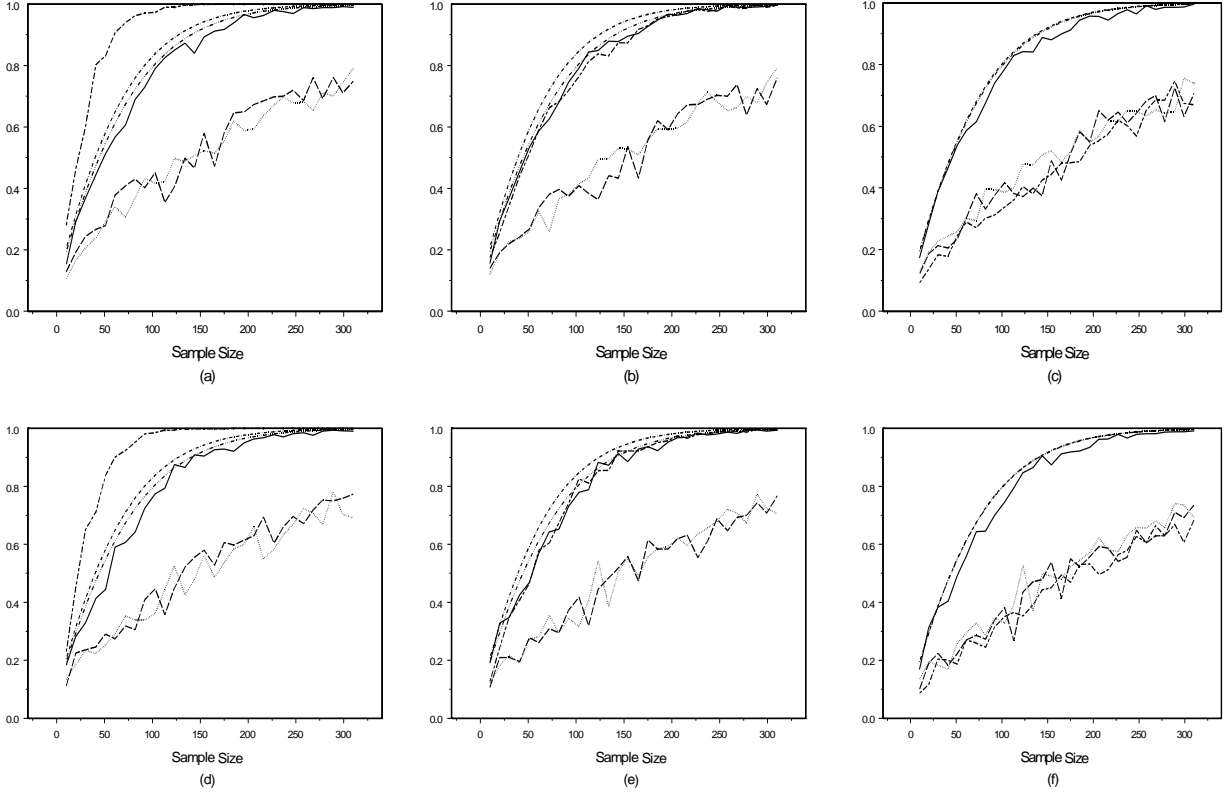


Figure 5: Asymptotic and exact power curves as a function of sample size for testing within-time difference (23) based on auto-regressive model (27) under t - and normal mixture distribution, with (a)-(c) $\Delta = 3$; (d)-(f) $\Delta = 4$; (a) and (d) $\zeta = 0.2$; (b) and (e) $\zeta = 0.5$; and (c) and (f) $\zeta = 0.9$. For t -distributions, $v_\epsilon = v_x = v = 4$ and for normal mixture, $\mu = 2$, $k^2 = 9$. Asymptotic curves are obtained under (dot-dashed line) and without (dots-dashed line) surrogacy condition, while exact power curves for t_{INT} (dashed line), t_{JT} (dotted line), N (solid line) and N_{MIX} (short-long dashed line) are based on Monte Carlo method.