Published in final edited form as:

Stat Med. 2014 February 28; 33(5): 812–827. doi:10.1002/sim.5986.

Power calculation for overall hypothesis testing with high-dimensional commensurate outcomes

Yueh-Yun Chi[a,*,†], **Matthew J. Gribbin**[b], **Jacqueline L. Johnson**[c], and **Keith E. Muller**[d]

[a]Department of Biostatistics, University of Florida, Gainesville, FL, U.S.A

[b]Department of Biostatistics, MedImmune, Gaithersburg, MD, U.S.A

[c]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, U.S.A

[d]Department of Health Outcomes and Policy, University of Florida, Gainesville, FL, U.S.A

Abstract

The complexity of system biology means that any metabolic, genetic, or proteomic pathway typically includes so many components (e.g., molecules) that statistical methods specialized for overall testing of high-dimensional and commensurate outcomes are required. While many overall tests have been proposed, very few have power and sample size methods. We develop accurate power and sample size methods and software to facilitate study planning for high-dimensional pathway analysis. With an account of any complex correlation structure between high-dimensional outcomes, the new methods allow power calculation even when the sample size is less than the number of variables. We derive the exact (finite-sample) and approximate non-null distributions of the 'univariate' approach to repeated measures test statistic, as well as power-equivalent scenarios useful to generalize our numerical evaluations. Extensive simulations of group comparisons support the accuracy of the approximations even when the ratio of number of variables to sample size is large. We derive a minimum set of constants and parameters sufficient and practical for power calculation. Using the new methods and specifying the minimum set to determine power for a study of metabolic consequences of vitamin B6 deficiency helps illustrate the practical value of the new results. Free software implementing the power and sample size methods applies to a wide range of designs, including one group pre-intervention and post-intervention comparisons, multiple parallel group comparisons with one-way or factorial designs, and the adjustment and evaluation of covariate effects.

Keywords

MANOVA; metabolomics; genomics; proteomics

*Correspondence to: Yueh-Yun Chi, Department of Biostatistics, University of Florida, Gainesville, FL, U.S.A.
†yychi@ufl.edu

1. Introduction

1.1. Motivation

Sample size determination is one of the important aspects of study planning. An overestimation of sample size misuses research resources, while an underestimation results in lack of adequate statistical power. High-throughput methods used in metabolomic, genomic, and proteomic research generate high-dimensional data. When planning such studies, scientists increasingly emphasize pathway and set-based analysis, as opposed to simultaneous analysis of each individual metabolite, gene, and protein. Pathway analysis of functionally or structurally related molecules provides interpretable results that may not be easily accessible from simultaneous, individual testing. Goeman and Bühlmann [1] indicated that, in microarray analysis, the shift of the level of analysis from single genes to sets of related genes allows biologists to make use of previously accumulated biological knowledge for a more biology driven analysis.

Owing to the complexity of system biology, any metabolic, genetic, and proteomic pathways typically includes so many components (e.g., molecules) that statistical methods specialized for overall testing of high-dimensional outcomes are required. Concentrations or relative abundances of molecules are typically measured from one single high-throughput bioassay, leading to commensurate measurements sharing the same scale and units. An adequate account of the correlation between the high-dimensional commensurate outcomes can increase statistical power and in turn reduce the sample size required for a targeted power. Traditionally, multivariate methods were developed to account for such within-subject correlation (of any structure); however, the classical multivariate statistics become undefined when the number of variables exceeds sample size.

Many statistical methods have been proposed for a single overall test of high-dimensional differences in a biological pathway or gene set due to, for example, treatment and type of cell ([1–9], among others). However, most approaches apply to a narrow range of designs. Even fewer have power and sample size methods.

With high-dimensional outcomes, typical of genetic research, for example, sample size is often smaller than the number of variables because of cost considerations. Existing power and sample size methods for overall testing either are based on an assumption that sample size is greater than number of variables [10, 11] or rely on large sample Gaussian approximations [7, 8]. Neither approach may be suitable to allow samples with more variables than subjects. Moreover, the lack of free, user-friendly power and sample size software impedes study planning. In this paper, accurate power and sample size methods and software, applicable whether sample size is smaller or larger than the number of variables, will be developed. The new methods will be useful for planning studies for metabolic, genetic, and proteomic pathway analysis, with one-way or factorial between-subject design, adjustment for fixed covariates. Furthermore, the methods apply to any collection of fixed predictors in a general linear multivariate model, including incomplete designs. The methods allow computing power for any kind of general linear multivariate model test, including overall tests of mean vectors, within-subject contrasts, between-subject contrasts, and between-by-within interaction contrasts.

1.2. Previous related work

Driven by a collaboration to compare metabolic profiles before and after a vitamin B6 depleted diet, Chi *et al.* [5] proposed a new test that extends the existing 'univariate approach' to repeated measures (UNIREP) to analyzing high dimension, low sample size data (i.e., number of variables greater than sample size). On the basis of the framework of the general linear multivariate model, their test is flexible and can be applied to general designs, including parallel group comparison and evaluation of continuous covariates. Through a combination of analytic results and extensive simulations, they provided evidence that the new test has better control of type I error rate and power, when compared with methods proposed by Ahmad *et al.* [12], Srivastava and Du [7], and Srivastava and Fujikoshi [8]. Their method is implemented in a free, SAS/IML program called LINMOD, as well as being available in the SAS procedure PROC GLM (using uepsdef = CM option on the REPEATED statement).

Computer simulations provide a general tool for power analysis when the exact or an approximate null distribution of the statistic is available and the non-null distribution is unknown. However, simulations can introduce a substantial computational burden in high-dimensional settings. We derive the exact non-null distribution and a corresponding approximate distribution for the UNIREP statistic. The results provide fast and accurate power calculations. By the method of moments, our F approximation has properly adjusted degrees of freedom and reduces to the F approximation derived by Chi *et al.* [5] when the noncentrality parameter is zero. We describe the exact and approximate non-null distributions in terms of minimum sufficient sets of parameters, which specify a particular power computation. The theory assumes fixed predictors (i.e., with values known before data collection). As discussed by Glueck and Muller [13], different theory and computation of power analysis is needed when random predictors are present.

1.3. New sample size method and software

A minimum set of constants and parameters sufficient to specify the exact non-null distributions consists of (1) error degrees of freedom, (2) number of between-subject contrasts, (3) number of within-subject contrasts, (4) scaled variances of principal components of the hypothesis variables, and (5) multiple semi-partial correlations between principal components and the set of predictors tested in the hypothesis (adjusted for predictors in the model but not included in the test). We derive power-equivalent hypothesis testing scenarios that share the exact non-null distribution. We simulate only simple scenarios with canonical design matrices and diagonal covariance matrices. The analytic results on power equivalence guarantee that the conclusions from the simulations apply to scenarios with complex design matrices and any covariance structure.

The new power approximation uses only three parameters in lieu of the roughly p^2 needed for exact calculations. Simulation studies support the accuracy of the approximation. An updated version of SAS/IML software POWERLIB implements the new approximation in two ways. First, the program provides approximate power based on knowing the population parameters, including the entire covariance matrix and mean matrix, as is typical of power calculation software. Second, the program also allows using sample data (even with more

variables than subjects) from a previous study to estimate the unknown parameters in the approximation and provide an estimated power for a new study.

We organize the rest of the paper as follows. In Section 2, we review the existing methods for testing general linear hypotheses and their distributional properties, especially with more variables than subjects. We then describe new exact and approximate results for the non-null distribution, highlight a number of power-equivalent hypothesis testing scenarios, and discuss power calculation when the population parameters are unknown. In Section 3, we provide a practical example for planning a study of metabolic consequences of vitamin B6 deficiency. In Section 4, we summarize results from extensive simulations for evaluating the accuracy of the approximations. In Section 5, we discuss implications of the new results and future research directions. We include all mathematical theorems and proofs in the Appendix.

2. Approach

2.1. Existing results for testing the general linear hypothesis

With rows of $\boldsymbol{Y}$ ($N \times p$) corresponding to subjects (independent sampling units) and columns of $\boldsymbol{Y}$ corresponding to repeated measures or commensurate outcomes, the general linear multivariate model is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}. \quad (1)$$

Here, $\boldsymbol{X}$ ($N \times q$) is the design matrix, and $\boldsymbol{B}$ ($q \times p$) is the primary parameters matrix. The Gaussian assumption leads to independently and identically distributed $\text{row}_i\,(\boldsymbol{E})' \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, equivalently $\boldsymbol{E} \sim \mathcal{N}_{N,p}(\boldsymbol{0}, \boldsymbol{I}_N, \boldsymbol{\Sigma})$ for a matrix Gaussian distribution, as defined in [14, Chapter 8]. Table II summarizes the notation for the parameters and constants of the model and associated general linear hypothesis $H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$, with $\boldsymbol{\Theta} = \boldsymbol{C}\boldsymbol{B}\boldsymbol{U}$ the matrix of secondary parameters, $\boldsymbol{C}$ ($a \times q$) the between-subject contrast matrix, and $\boldsymbol{U}$ ($p \times b$) the within-subject contrast matrix.

A set of regularity conditions define estimable and testable $\boldsymbol{\Theta}$: (1) $\text{rank}(\boldsymbol{C}) = a \le q$, (2) $\text{rank}(\boldsymbol{U}) = b \le p$, and (3) $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}(\boldsymbol{X}'\boldsymbol{X})$. Rows of the between-subject contrast matrix $\boldsymbol{C}$ define a contrasts in the predictor space, while columns of the within-subject contrast matrix $\boldsymbol{U}$ define b contrasts in the outcome space. In turn, the first two regularity conditions lead to defining an estimable and testable subspace. The third regularity condition is needed for the linear estimability of $\boldsymbol{\Theta}$ ([15, p. 137] and [14, Section 11.4]). For an estimable $\boldsymbol{\Theta}$, both least squares and maximum likelihood estimation methods give $\hat{\boldsymbol{\Theta}} = \boldsymbol{C}\,\boldsymbol{B}\,\tilde{\boldsymbol{U}}$ with $\tilde{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}$ and $(\boldsymbol{X}'\boldsymbol{X})^{-}$ the generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. The general linear hypothesis $H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ for testable $\boldsymbol{\Theta}$ covers many designs for one, two, or many samples, including variations or analogs of repeated measures and multivariate analysis of variance, multivariate regression, discriminant analysis with two or more groups, and canonical correlation.

The UNIREP test statistic for the general linear multivariate model is proportional to the ratio of the trace of the hypothesis sum of squares matrix, $\hat{\boldsymbol{\Delta}} = (\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)\,\boldsymbol{M}^{-1}\,(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$, to the trace of error sum of squares matrix, $\nu_e\hat{\boldsymbol{\Sigma}}_* = \nu_e\boldsymbol{U}'\hat{\boldsymbol{\Sigma}}\boldsymbol{U}$:

$$t_u = \frac{\mathrm{tr}(\hat{\mathbf{\Delta}})/a}{\mathrm{tr}(\hat{\mathbf{\Sigma}}_*)}, \quad (2)$$

with $\hat{\mathbf{\Sigma}} = (\boldsymbol{Y} - \boldsymbol{X}\,\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\,\boldsymbol{B})/\nu_e$ and $\boldsymbol{M} = \boldsymbol{C}\,(\boldsymbol{X}'\,\boldsymbol{X})^{-}\,\boldsymbol{C}'$. The statistic t_u is a one-to-one function of the sample estimate $\hat{\eta}_u = \mathrm{tr}\,(\hat{\mathbf{\Delta}}) = [\mathrm{tr}\,(\hat{\mathbf{\Delta}}) + \nu_e \mathrm{tr}(\hat{\mathbf{\Sigma}}_*)]$ for the UNIREP measure of multivariate association:

$$\eta_u = \frac{\mathrm{tr}(\mathbf{\Delta})}{\mathrm{tr}(\mathbf{\Delta}) + N\mathrm{tr}(\mathbf{\Sigma}_*)}. \quad (3)$$

Chi *et al.* [5] showed that under the null hypothesis $H_0 : \mathbf{\Theta} = \mathbf{\Theta}_0$, regardless of the ratio of sample size to the number of variables, the distribution function of t_u can be expressed exactly as

$$\Pr\{t_u \le f_0\} = \Pr\left\{\sum_{k=1}^{b}\lambda_k y_{kh} - f_0 a \nu_e^{-1}\sum_{k=1}^{b}\lambda_k y_{ke} < 0\right\} \quad (4)$$

$$= \Pr\left\{\sum_{k=1}^{b}\pi_k y_{kh} - f_0 a \nu_e^{-1}\sum_{k=1}^{b}\pi_k y_{ke} < 0\right\}, \quad (5)$$

with $y_{kh} \sim \chi^2(a)$ independent of $y_{ke} \sim \chi^2(\nu_e)$ for $k \in \{1, 2, \cdots, b\}$. With $\pi_k = \lambda_k / \left(\sum_{k=1}^{b}\lambda_k\right)$, the exact null distribution of t_u is invariant to global scale change in outcome space. From the prospective of principal component analysis, eigenvalues $\{\lambda_k\}$ can be interpreted as variances of the principal components of hypothesis variables $\boldsymbol{Y}_u = \boldsymbol{YU}$, while the scaled eigenvalues $\{\pi_k\}$ can be interpreted as proportions of variances explained by the principal components.

Chi *et al.* [5] extended the Box *F* approximation to the high dimension, low sample size setting:

$$\Pr\{t_u \le f_0\} \approx \Pr\{F(ab\varepsilon, \nu_e b\varepsilon) \le f_0\}. \quad (6)$$

Here, $\varepsilon = \left(\sum_{k=1}^{b}\lambda_k\right)^2 / \left(b\sum_{k=1}^{b}\lambda_k^2\right)$ is the sphericity parameter quantifying the spread of population eigenvalues $\{\lambda_k\}$ of $\mathbf{\Sigma}_*$. They proposed a new estimator of ε, based on matching moments of the dual of the error covariance matrix. Their estimator is

$$\tilde{\varepsilon} = \max\{\min[(\nu_a - 2)(\nu_a - 4)\hat{\tau}_1/(\nu_a^2 b\hat{\tau}_2), 1], 1/b\} \quad (7)$$

with

$$\hat{\tau}_1 = \nu_e\left[(\nu_e + 1)\mathrm{tr}^2\left(\hat{\mathbf{\Sigma}}_*\right) - 2\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_*^2\right)\right]/(\nu_e^2 + \nu_e - 2), \quad (8)$$

$$\hat{\tau}_2=\nu_e\left[\nu_e \mathrm{tr}\left(\hat{\mathbf{\Sigma}}_*^2\right)-\mathrm{tr}^2\left(\hat{\mathbf{\Sigma}}_*\right)\right]/(\nu_e^2+\nu_e-2), \quad (9)$$

$$\nu_a=(\nu_e-1)+\nu_e(\nu_e-1)/2. \quad (10)$$

Simulation results reported in Section 4 (including Tables III– VIII and Figure 2) of Chi *et al.* [5] demonstrate that their method (studied in the present paper) accurately controls type I error rate and has reasonable power even with a handful of subjects and thousands or more outcome variables. Chi *et al.* [5] reported empirical type I error rates for data simulated with $b < \nu_e$ (low dimension, high sample size) or $b > \nu_e$ (high dimension, low sample size). The range of designs included one-sample comparisons, two-sample comparisons, and three-sample comparisons with a covariate. Their test (T_2 in [5], also studied here) better controlled the type I error rate than the classic UNIREP tests when $b > \nu_e$. The new test also outperformed the asymptotic tests proposed by Srivastava and Du [7] and Srivastava and Fujikoshi [8], especially with a small sample size.

2.2. New results: Exact non-null distribution

Analytic power calculation requires knowing the distribution of the statistic under the alternative hypothesis (e.g., $H_1 : \mathbf{\Theta} \neq \mathbf{\Theta}_0$). In this section, we describe the exact noncentral distribution of t_u and its properties and delay the discussion about approximation to the next section.

Theorem 1 gives a simple expression for the exact noncentral distribution of t_u, which has a parallel form to the null distribution in Equation (5), except with $y_{kh} \sim \chi^2(a, \omega_k)$ for $k \in \{1, 2, \cdots, b\}$. The noncentral distribution is fully determined by specifying 3 constants and $2 \cdot b$ parameters. The constants are determined by the design and hypothesis: (1) the error degrees of freedom, ν_e, (2) the number of between-subject contrasts, a, and (3) the number of within-subject contrasts, b. The unknown sets of sufficient parameters are $\{\pi_k\}$ and $\{\omega_k\}$.

The noncentrality parameters $\{\omega_k\}$ are the diagonal elements of $\mathbf{\Omega}_t$, the transformed, scaled non-centrality for the hypothesis sums of squares matrix $\hat{\mathbf{\Delta}}$. Theorem 2(b) states that each noncentrality parameter is a one-to-one function of ρ_k^2, the squared multiple semi-partial correlation between principal components of hypothesis variables $\boldsymbol{Y}_u$, and the set of predictors tested, with the predictors adjusted for all untested predictors in the model. The relationship is

$$\omega_k=N\rho_k^2/(1-\rho_k^2). \quad (11)$$

Theorem 2(c) expresses the UNIREP measure of multivariate association in Equation (3) as

$$\eta_u=\frac{\sum_{k=1}^b \pi_k\rho_k^2/(1-\rho_k^2)}{\sum_{k=1}^b \pi_k/(1-\rho_k^2)}. \quad (12)$$

The squared multiple correlations $\{\rho_k^2\}$ are the univariate-multivariable coefficients of determination. They describe the proportion of variance in principal component k explained by the set of a predictors tested in the hypothesis (adjusted for predictors in the model but not included in the test). The interpretability of $\{\rho_k^2\}$ allows straightforward elicitation of $\{\omega_k\}$. Together, $\{\pi_k\}$ and $\{\rho_k^2\}$ define one set of sufficient parameters for exact noncentral distribution of t_u. They are the variance and correlation parameters characterizing the population structure in the principal component space. The sufficiency allows simplifying study planning and simulation designs because eliciting high-dimensional variance and covariance parameters can be a daunting task. The number of parameters greatly reduces from $pq + p(p + 1)/2$ in the original parameter space for $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ to $2 \cdot b$ in the reduced space for $\{\pi_k\}$ and $\{\rho_k^2\}$.

Any test that is based on a general linear multivariate model and its associated UNIREP statistic shares the same noncentral distribution and thus is power equivalent when N, ν_e, a, b, $\{\pi_k\}$, and $\{\omega_k\}$ are the same (Corollary 1). The equivalence holds if $\{\lambda_k\}$ replaces $\{\pi_k\}$ or $\{\rho_k^2\}$ replaces $\{\omega_k\}$.

In Table II, three power-equivalent hypothesis testing scenarios are defined and depicted. Theorem 2(a) gives details of their statistical properties. The three scenarios share both $\{\pi_k\}$ and $\{\omega_k\}$ in addition to constants N, ν_e, a, and b. Scenario S_1 has the original multivariate model defined in Equation (1), while scenarios S_2 and S_3 are based on transformed versions of the original model. Scenarios S2 has outcome variables $\boldsymbol{Y}_2 = \boldsymbol{YU}$ $\boldsymbol{\Upsilon} = \boldsymbol{Y}_u$ $\boldsymbol{\Upsilon}$. With $\mathcal{V}(\text{row}_i(\boldsymbol{Y}_2)) = \text{Dg}(\lambda)$, outcome variables for scenario S_2 are principal components of the hypothesis variables. The power equivalence between scenarios S_1 and S_2 allows assuming a diagonal hypothesis error covariance matrix $\boldsymbol{\Sigma}_*$ without loss of generality.

Following the notation in Table I, scenario S_3 in Table II has outcome variables $\boldsymbol{Y}_3 = \boldsymbol{T}\,\boldsymbol{Y}_2$ with $\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R}_D' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\nu_e} \end{bmatrix} \boldsymbol{L}_X'$ and $\boldsymbol{R}_D$ the eigenvectors of $\boldsymbol{D}'\,\boldsymbol{D}$ for $\boldsymbol{D} = \boldsymbol{C}\boldsymbol{R}_{X+}\text{Dg}(\boldsymbol{s}_{X+})^{-1}$. Also, it has a simple, generic design matrix $\begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix}$ and between-subject contrast matrix $[\boldsymbol{I}_a\ \boldsymbol{0}]$.With $\mathcal{V}(\text{row}_i(\boldsymbol{Y}_3)) = \text{Dg}(\lambda)$, outcome variables for scenario S_3 are transformed components of the hypothesis variables. The power equivalence between scenarios S_1 and S_3 allows assuming a diagonal hypothesis error covariance matrix $\boldsymbol{\Sigma}_*$ and canonical design and contrast matrices without loss of generality. An important implication is that simulation results for diagonal $\boldsymbol{\Sigma}_*$ and canonical structure generalize to models with any covariance and design structure.

2.3. New results: Approximate non-null distribution

The lack of a closed form for the exact distribution of t_u leads to considering an approximation. By the method of moments, the distribution of t_u under the alternative hypothesis $H_1 : \boldsymbol{\Theta} \neq \boldsymbol{\Theta}_0$ can be approximated by a noncentral F distribution, namely

$$\Pr\{t_u \leq f_0\} \approx \Pr\{F(ab\varepsilon_n, b\nu_e\varepsilon_d, \omega_u) \leq f_0\}. \quad (13)$$

Theorem 3 provides detailed expressions for parameters ε_n, $\varepsilon_d = \varepsilon$, and ω_u, all functions of $\{\lambda_k\}$ and $\{\omega_k\}$. Under the null hypothesis, $\omega_u = 0$ and $\varepsilon_n = \varepsilon$, leading to the central F approximation in Equation (6). At the test size α, power can be approximated by computing $1 - \Pr\{F(ab\varepsilon_n, b\nu_e\varepsilon_d, \omega_u) \le f_{crit}\}$ with $\Pr\{F(ab\tilde{\mathrm{E}}(\varepsilon), \nu_e b\tilde{\mathrm{E}}(\varepsilon)) \le f_{crit}\} = 1 - \alpha$. We extend the results in [11] to approximate $\tilde{\mathrm{E}}(\varepsilon)$ by $[(\nu_a - 2)(\nu_a - 4)/\nu_a^2]E(\tilde{\varepsilon}_{\mathrm{HF}})$ with $\tilde{\varepsilon}_{\mathrm{HF}} = \hat{\tau}_1/(b\hat{\tau}_2) = [\nu_e + 1)b\hat{\varepsilon} - 2] / [b(\nu_e - b\hat{\varepsilon}]$ and $\hat{\varepsilon} = b^{-1}\mathrm{tr}^2\left(\hat{\mathbf{\Sigma}}_*\right)/\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_*^2\right)$.

The noncentral F approximation is fully determined by only three parameters, namely ε_n, $\varepsilon_d = \varepsilon$, and ω_u. Here, ω_u can be expressed as a function of ε_n and η_u, the UNIREP measure of multivariate association in Equation (12). Both ε_n and ω_u contains information pertaining to the magnitude and location of the association between the principal components of the hypothesis variables and the predictors tested. The minimum, sufficient parameter set for approximation (3 parameters) can be fully determined by the minimum, sufficient parameter set for exact power calculation ($2b$ parameters), which in turn can be fully determined by the original parameter set ($pq + p(p + 1)/2$ parameters).

2.4. New results: Power calculation with estimated population parameters

As discussed in [16], statisticians frequently fix mean values and calculate power or sample size using a variance estimate from an existing study. For the general linear multivariate model given in Equation (1), fixed primary parameter matrix $\boldsymbol{B}$ leads to fixed secondary parameter matrix $\boldsymbol{\Theta}$ and in turn to fixed hypothesis sum of squares matrix $\boldsymbol{\Delta}$. Adding a well-estimated $\boldsymbol{\Sigma}$ is sufficient for calculating an estimated power. A full rank estimate of $\boldsymbol{\Sigma}$ from an existing study allows estimating scaled eigenvalues of $\boldsymbol{\Sigma}_*$, which in turn leads to specifications of the minimum sufficient parameters $\{\pi_k\}$ and $\{\omega_k\}$ for the exact noncentral distribution. When the existing data have more variables than subjects, the estimator of $\boldsymbol{\Sigma}$ is singular and does not have adequate information to accurately provide estimates of the population eigenvalues [17].

In order to use data from an existing study with more variables than subjects, additional steps must be taken to estimate power for a new study. In practice, the three parameters $\{\varepsilon_d, \varepsilon_n, \omega_u\}$ suffice to specify the approximate noncentral F distribution (Equation (13)). By using Equations (8)–(10), we can first compute $\hat{\tau}_{10}$, $\hat{\tau}_{20}$, and ν_{a0} from the existing data (hence the subscript 0) of sample size N_0, design matrix $\boldsymbol{X}_0$, error degrees of freedom $\nu_{e0} = N_0 - \mathrm{rank}(\boldsymbol{X}_0)$, and hypothesis error covariance matrix $\hat{\Sigma}_{*0}$. We can then use Equations (7), (20), and (22) to estimate $\{\varepsilon_d, \varepsilon_n, \omega_u\}$ as

$$\hat{\varepsilon}_d = \max\{\min[(\nu_{a0} - 2)(\nu_{a0} - 4)\hat{\tau}_{10}/\left(b\nu_{a0}^2\hat{\tau}_{20}\right), 1], 1/b\} \quad (14)$$

$$\hat{\varepsilon}_n = \frac{1 + 2\mathrm{tr}[(\mathbf{\Delta}_t)/a]/\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_{*0}\right)}{\hat{\varepsilon}_d^{-1} + 2b\mathrm{tr}\left[\hat{\mathbf{\Sigma}}_{*0}(\mathbf{\Delta}_t)/a\right]/\mathrm{tr}^2\left(\hat{\mathbf{\Sigma}}_{*0}\right)} \quad (15)$$

$$\hat{\omega}_u = b\hat{\varepsilon}_n \mathrm{tr}[(\boldsymbol{\Delta}_t)]/\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{*0}\right), \quad (16)$$

with $\boldsymbol{\Delta}_t = (\boldsymbol{CBU} - \boldsymbol{\Theta}_0)' [\boldsymbol{C} (\boldsymbol{X}'_t \boldsymbol{X}_t)^- \boldsymbol{C}']^{-1} (\boldsymbol{CBU} - \boldsymbol{\Theta}_0)$ and $\boldsymbol{X}_t$ the design matrix for the planned study. We note that the primary parameter matrix $\boldsymbol{B}$ is generally known and does not require estimation. The existing data are used only to estimate $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\Sigma}}_0$, leading to which immediately gives estimates of $\boldsymbol{\Sigma}_*$ and functions of $\boldsymbol{\Sigma}_*$, such as $\mathrm{tr}(\boldsymbol{\Sigma}_*)$, $\mathrm{tr}^2(\boldsymbol{\Sigma}_*)$, and ε_d.With estimates of $\{\varepsilon_d, \varepsilon_n, \omega_u\}$ in hand, power can be approximated by computing $1 - \Pr\{F(ab\hat{\varepsilon}_n, b\nu_e\hat{\varepsilon}_d, \hat{\omega}_u) \le \hat{f}_{crit}\}$ with $\Pr\{F(ab\tilde{\varepsilon}, \nu_e b\tilde{\varepsilon}) \le \hat{f}_{crit}\} = 1 - \alpha$ at the test size α. The form for the estimated critical value $\hat{f}_{crit}$ is based on the results from the extensive simulations in [18].

Variance estimates are known to have wide sampling distributions. Browne [19], Taylor and Muller [16, 20], and Muller and Pasour [21] described how to account for using estimates in power calculations in various settings with the univariate model. In the univariate setting, the $100(1 - \beta)$th percentile of the variance estimates can be used to achieve the planned power in at least $100(1 - \beta)\%$ of such studies. In the multivariate setting, which is the focus of the present manuscript, estimates for the $b \times b$ covariance matrix Σ_* lead to estimates of the parameters $\{\varepsilon_d, \varepsilon_n, \omega_u\}$ and the critical value f_0 (Equation (13)). Gribbin *et al.* [18] addressed the question for traditional low-dimensional designs with the sample size greater than the number of variables. Extending their results to the high dimension, low sample size design remains a topic for future research. Computing confidence bounds for power and sample size provides a natural solution to account for the variation and ensure achieving the planned power.

3. Simulations

3.1. Design

Designing simulations with high-dimensional outcomes can be difficult as the number of variance and covariance parameters is of the magnitude of the squared number of outcomes, that is, $b(b + 1)/2$. On the basis of Corollary 2, power equivalence between scenarios S_1 and S_4 allows greatly simplifying the task. Scenario S_1 uses the general linear multivariate model $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$ with no constraints on the mean of $\boldsymbol{Y}$, namely $\boldsymbol{XB}$, or on the covariance structure of the errors, namely $\boldsymbol{\Sigma}$. In contrast, scenario S_4 is based on the model $\boldsymbol{Y}_4 = \boldsymbol{X}_4\boldsymbol{B}_4 + \boldsymbol{E}_4$, which has four constraints: (1) independent errors, $\mathcal{V}[\mathrm{row}_i(\boldsymbol{E}'_4)] = \mathrm{Dg}(\pi)$; (2) orthonormal predictors, $\boldsymbol{X}'_4\boldsymbol{X}_4 = \boldsymbol{I}_r$; (3) primary parameter matrix $\boldsymbol{B}_4$ is sparse with all zeros except in the first row; and (4) simple contrast matrices, $\boldsymbol{C}_4 = [\boldsymbol{I}_a \ \boldsymbol{0}]$ and $\boldsymbol{U}_4 = \boldsymbol{I}_b$. The power equivalence between scenarios S_1 and S_4 results from sharing a, b, ν_e, $\{\pi_k\}$, and $\{\omega_k\}$ (or $\{\rho_k^2\}$).

We designed our simulations on the basis of the data structures in scenario S_4. For both one-sample and two-sample comparisons, we adopted a five-way complete factorial design with factors $b \in \{64, 256, 1024\}$, $N \in \{10, 20, 40\}$, $\varepsilon \in \{0.27, 0.56, 0.76\}$, number of nonzero $\rho_k^2 \in \{4, 32\}$, and the location of nonzero ρ_k^2 at either the most dominant or middle components in hypothesis space. We also simulated a three-sample comparison with the same factorial design except for using $N \in \{12, 24, 48\}$ with equal size groups. We varied the number of nonzero ρ_k^2 in order to address the power differences expected to result

between concentrated and diffuse patterns of effect [22]. The location of nonzero ρ_k^2 dictates the sources of variation that are accounted for by predictors tested. The ordered, scaled variances $\pi_1, \ldots, \pi_b$ and τ were selected with $\pi_k = (b - k + 1)^\tau$ such that $\varepsilon \in \{0.27, 0.56, 0.76\}$, and $\sum_k \pi_k = 1$. The values of nonzero ρ_k^2 were set equal to each other and selected to achieve target power $P \in \{0.2, 0.5, 0.8, 0.9\}$ for $b = 64$, and $P \in \{0.8, 0.9\}$ for $b \in \{256, 1024\}$. All simulations used 100,000 replications and $\alpha = 0.05$.

When $b \in \{256, 1024\}$ and $\varepsilon = 0.27$, some of the variance ratios π_k/π_1 are numerical zeros, leading to an untestable hypothesis. The observation can be formalized as a generalization of Corollary 15.1 in [14] in the following way. A test for $\boldsymbol{\Theta}_4 = \boldsymbol{C}_4\boldsymbol{B}_4\boldsymbol{U}_4$ is indistinguishable from a test for $\boldsymbol{\Theta}_{4*} = \boldsymbol{C}_4\boldsymbol{B}\boldsymbol{U}_{4*}$, with $\boldsymbol{U}_{4*} = [\boldsymbol{I}_{b*}\ \boldsymbol{0}]'$ and b_* the number of the variance ratios π_k/π_1 greater than zeros, numerically. In order to assess whether the result was of practical concern, we compared power approximations for testing $\boldsymbol{\Theta}_4$ and $\boldsymbol{\Theta}_{4*}$. We observed that the approximated power values for the b-dimensional and b_*-dimensional tests coincide to at least the fourth decimal place. The expectation $E(\tilde{\varepsilon})$ varies at the third decimal place, leading to small differences in the critical values expected. We therefore report the simulation results for testing $\boldsymbol{\Theta}_4$ when the entire outcome space is considered.

3.2. One-sample problem

Table III displays a summary of absolute differences between the empirical and approximated power values. Performance of our proposed approximation was evaluated and summarized across a range of sample sizes N, covariance structures (governed by ε), and patterns of non-null effects (i.e., number and location of nonzero ρ_k^2). Across the board, our approximation produced absolute biases with both means and standard deviations less than 0.005. As the number of outcomes increased, the mean and standard deviation decreased. Overall, accuracy improved with number of outcomes, as well as sample size. The results are particularly useful for planning nutritional research when one-sample design is commonly used for studies comparing pre-intervention and post-intervention profiles.

3.3. Two-sample problem

Two-sample comparisons are very common in genomic and microarray research. Table IV summarizes the absolute bias when approximating power for testing differences between two groups. Both the means and standard deviations of the absolute biases were less than 0.01 when target power $P \in \{0.2, 0.5, 0.8, 0.9\}$ and number of hypothesis outcomes $b \in \{64, 256, 1024\}$. Table V shows a detailed summary for $b = 64$ and $P = 0.9$ for a variety of sample sizes, noncentrality pattern, and covariance pattern. In general, as sample size increased, the bias decreased. The approximation worked slightly better when the group differences arose from the most dominant principal components of the hypothesis variables than from the middle dominant principal components. Overall, all documented biases were less than 0.03, giving approximated power in the range between 0.90 and 0.93.

3.4. Three-sample problem

Table VI summarizes the power accuracy for three-sample comparisons. Average bias was below 0.03, slightly higher than for one-sample and two-sample designs. Despite that, for power and sample size calculation, a bias at the level of 0.03 is usually acceptable.

4. Study of vitamin B6 deficiency

Chi *et al.* [5] illustrated an example data analysis by considering static concentrations of 19 amino acids from 12 healthy participants. The 19 concentrations were collected before and after a 4-week diet low in vitamin B6. The objective was to assess the metabolic consequences of marginal deficiency of vitamin B6. The data were logarithmically transformed (to meet the Gaussian distributional assumption) before computing pre/post difference scores. Testing the mean vector equal to zero gave a small *p*-value indicating significant change in metabolic profile after vitamin B6 restriction.

With 19 variables, fixed values for their correlations and variances (190 parameters) are not readily available. The existing data provide a means to estimate the covariance and correlation parameters required to compute approximated power for a future study intended to replicate the results of the previous study. As discussed by Lenth [23] and others, using observed means and observed variances for power calculation as an adjunct to data analysis has no value. It does have value for planning a new study to replicate results, perhaps with a larger or smaller sample size. In the example discussed in the current section, we perform power calculations for a number of interesting and plausible patterns of mean differences. Muller and Stewart [14] (Section 20.5) provided additional discussion, especially in the context of research reported in [20] on the bias introduced by conducting power analysis only if the previous research did (or did not) give a significant result.

Equations (14)–(16) were used to estimate $\{\varepsilon_d, \varepsilon_n, \omega_u\}$ with $a = 1$ (for one group pre-intervention and post-intervention comparison) and $b = 19$ after computing $\hat{\tau}_{10}$, $\hat{\tau}_{20}$, and ν_{a0} using the existing data of sample size $N_0 = 12$. At the 0.05 significance level, our calculation led to an overall power of 0.83 when the target sample size is 10, and all amino acids exhibit a mean absolute difference (diffused effects), after logarithmic transformation, of 0.14 log(μmol/L). In contrast, with the same level of effect across all amino acids, the overall power increases to 0.96 for a target sample size of 15. Power was also calculated for a concentrated effect with a difference only in cystathionine, the amino acid that has been shown in the literature to inversely relate to the B6 abundance. At a mean difference of 0.55 log(μmol/L) for cystathionine, the estimated power is 0.86 for a target sample size of 10 and 0.98 for a target sample size of 15.

Figure 1 shows the approximate power curves as a function of the mean difference for diffused (top) and concentrated (bottom) effects. Both cases use a single multiplier for the entire set of outcome variables. Nonetheless, our methods apply to calculations that require different multipliers for different sets of outcome variables. The free, downloadable POWERLIB software implementing our methods was specifically configured to allow any pattern of multipliers. Our collaborators were able to make an informed decision in selecting sample size by comparing the power values for different sample sizes and patterns of effects.

In the example, the nuisance parameters for covariance and correlation were estimated from the existing data and used in power approximation to compute estimated power. Simulations were conducted to evaluate the accuracy of power approximation as a result of estimating population parameters. With 100,000 replications, the differences between the empirical and our approximated power range from 0.001 to 0.065 for the six designs of varied sample and effect sizes. The results demonstrate the accuracy of our power approximation and estimation methods when estimates of covariance and correlation parameters are used.

5. Discussion

Power calculation for a sample size less than the number of variables is required for planning studies involving high-throughput methods. We provided exact results to help understand the problem and derive properties. An accurate approximation was developed and has been implemented in the existing free SAS/IML software called POWERLIB (the updated version with the new methods will be made available when the paper is accepted for publication: https://github.com/SampleSizeShop/POWERLIB). We also described a moment-based approach for using existing data in planning a replication study and applied the method to a study of vitamin B6 deficiency.

Further advancements on approximating power with estimated parameters require future analytical and numerical work. Extensions can be sought on two aspects. First, methods for computing confidence bounds for power and sample size are needed to quantify the uncertainty as a result of using observed data in a power analysis. Second, different theory and computation of power analysis are needed when random predictors (i.e., with values unknown before data collection) are present.

With massive data collected from high-throughput platforms, we urge anyone conducting data analysis, power analysis, or simulations to practice safe computing. Numerical inaccuracy and computer memory issues can easily emerge when the number of variables grows into the thousands. We aim to continue upgrading our software for its computational efficiency and user interface.

Acknowledgments

We thank Dr. Jesse F. Gregory III for permission of using study of vitamin B6 deficiency as an example. Chi's support included NIDCR R01-DE020832, NICHD P01-HD065647, NIDDK R01-DK072398, NIDA R01-DA031017, and UF CTSI core grant NCATS UL1000064. Muller's support included NIDCR R01-DE020832, NIDDK R01-DK072398, NIDCR U54-DE019261, NCRR K30-RR022258, NHLBI R01-HL091005, NIAAA R01-AA013458-01, NIDA R01-DA031017, and NICHD P01-HD065647.

Appendix A

Notation in Tables I and II is used throughout the Appendix without specific references.

Theorem 1

For the model in hypothesis testing scenario S_1 in Table II, the distribution function of $t_u = [\mathrm{tr}\,(\hat{\boldsymbol{\Delta}}) / a] / [\mathrm{tr}\,(\hat{\boldsymbol{\Sigma}}_*)]$ can be expressed exactly as

$$\Pr\{t_u \le f_0\} = \Pr\left\{\sum_{k=1}^{b} \pi_k y_{kh} - f_0 a \nu_e^{-1} \sum_{k=1}^{b} \pi_k y_{ke} < 0\right\}, \quad (17)$$

with $y_{kh} \sim \chi^2(a, \omega_k)$ independent of $y_{ke} \sim \chi^2(\nu_e)$ for $k \in \{1, 2, \cdots, b\}$, and $\omega_k = \boldsymbol{\upsilon}'_k \boldsymbol{\Delta} \boldsymbol{\upsilon}_k / \lambda_k$ for $\boldsymbol{\upsilon}_k$ and λ_k a corresponding eigenvector and eigenvalue of Σ_*. Under the null hypothesis, the noncentrality parameter $\omega_k = 0$, leading to $y_{kh} \sim \chi^2(a)$ for $k \in \{1, 2, \cdots, b\}$. The theorem applies for data with $b \le \nu_e$ or $b > \nu_e$. *The exact noncentral distribution is fully determined by a, b,* ν_e, $\{\pi_k\}$, *and* $\{\omega_k\}$.

Proof

Under the Gaussian assumption $\boldsymbol{E} \sim \mathcal{N}_{N,p}(\boldsymbol{0}, \boldsymbol{I}_N, \boldsymbol{\Sigma})$ with $b \le \nu_e$ or $b > \nu_e$, Chi *et al.* [5] proved that both the hypothesis and error sum of square matrices, $\hat{\boldsymbol{\Delta}} = (\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)' M^{-1} (\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$ and $\nu_e \hat{\boldsymbol{\Sigma}}_* = \nu_e \boldsymbol{U}' \boldsymbol{\Sigma} \hat{\boldsymbol{U}}$, follow a Wishart distribution and are mutually independent, namely $\hat{\boldsymbol{\Delta}} \sim (\mathcal{S})\, \mathcal{W}_b(a, \boldsymbol{\Sigma}_*, \boldsymbol{\Omega}) \perp \nu_e \hat{\boldsymbol{\Sigma}}_* \sim (\mathcal{S})\, \mathcal{W}_b(\nu_e, \boldsymbol{\Sigma}_*, \boldsymbol{0})$. With $\boldsymbol{\Sigma}_* = \boldsymbol{\Upsilon} \mathrm{Dg}(\lambda) \boldsymbol{\Upsilon}'$,

$\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\right) = \mathrm{tr}\left(\boldsymbol{\Upsilon}' \hat{\boldsymbol{\Delta}} \boldsymbol{\Upsilon}\right) = \sum_{k=1}^{b} \lambda_k y_{kh}$ for $\boldsymbol{\Upsilon}' \hat{\boldsymbol{\Delta}} \boldsymbol{\Upsilon} \sim (\mathcal{S})\, \mathcal{W}_b(a, \mathrm{Dg}(\lambda), \boldsymbol{\Omega}_t)$, $y_{kh} \sim \chi^2(a, \omega_k)$, and $\omega_k = \boldsymbol{\upsilon}'_k \boldsymbol{\Delta} \boldsymbol{\upsilon}_k / \lambda_k$ the kth diagonal element of $\boldsymbol{\Omega}_t = (\boldsymbol{\Upsilon}' \boldsymbol{\Delta} \boldsymbol{\Upsilon}) \mathrm{Dg}(\lambda)^{-1}$ [24]. Similarly,

$\nu_e \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_*\right) = \nu_e \mathrm{tr}\left(\boldsymbol{\Upsilon}' \hat{\boldsymbol{\Sigma}}_* \boldsymbol{\Upsilon}'\right) = \sum_{k=1}^{b} \lambda_k y_{ke}$ for $\nu_e \boldsymbol{\Upsilon}' \hat{\boldsymbol{\Sigma}}_* \boldsymbol{\Upsilon}' \sim (\mathcal{S})\, \mathcal{W}_b(\nu_e, \mathrm{Dg}(\lambda), \boldsymbol{0})$, and $y_{ke} \sim \chi^2(\nu_e)$. Independence between $\hat{\boldsymbol{\Delta}}$ and $\hat{\boldsymbol{\Sigma}}_*$ leads to independence between $\{y_{kh}\}$ and $\{y_{ke}\}$ and between $\mathrm{tr}(\hat{\boldsymbol{\Delta}})$ and $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_*)$. Finally

$$\begin{aligned}
&\Pr\{t_u \le f_0\} \\
&\quad = \Pr\left\{\left[\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\right)/a\right] / \left[\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_*\right)\right] \le f_0\right\} = \Pr\left\{\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\right) - a f_0 \left[\nu_e \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_*\right)\right]/\nu_e \le 0\right\} \\
&\quad = \Pr\left\{\sum_{k=1}^{b} \lambda_k y_{kh} \right. \\
&\qquad \left. - f_0 a \nu_e^{-1} \sum_{k=1}^{b} \lambda_k y_{ke} \le 0\right\} \\
&\quad = \Pr\left\{\sum_{k=1}^{b} \pi_k y_{kh} \right. \\
&\qquad \left. - f_0 a \nu_e^{-1} \sum_{k=1}^{b} \pi_k y_{ke} < 0\right\}.
\end{aligned}$$

Corollary 1

- **a.** Any distinct scenario with a model and hypothesis giving the same a, b, ν_e, $\{\pi_k\}$, and $\{\omega_k\}$ has a test statistic with the same distribution, the same type I error rate, and the same power.
- **b.** It is sufficient (but not necessary) for power equivalence that two scenarios have the same a, b, ν_e, transformed, scaled noncentrality for the hypothesis sums of squares, $\boldsymbol{\Omega}_t$, and $\{\lambda_k\}$.

Proof

The truth of each part of the corollary follows directly from the fact that the parameters listed fully specify the distribution of t_u.

Theorem 2

a. In Table II, testing $H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ in scenario S_1 is power equivalent to testing the hypothesis $H_0 : (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y} = \mathbf{0}$ in scenarios S_2, which is power equivalent to testing the hypothesis $H_0 : \boldsymbol{T}_D(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y} = \mathbf{0}$ in scenarios S_3 for $\boldsymbol{T}_D = \mathrm{Dg}(\boldsymbol{s}_{D+})^{-1}\boldsymbol{L}'_{D+}$.

b. Each noncentrality parameter needed to apply Theorem 1 is a one-to-one function of ρ_k^2, namely $\omega_k = \boldsymbol{v}_k'\boldsymbol{\Delta}\boldsymbol{v}_k/\lambda_k = N\rho_k^2/\left(1-\rho_k^2\right)$, while $\rho_k^2 = \omega_k/(\omega_k + N)$ for $k \in \{1, 2, \cdots, b\}$. With y_{2k} column k of $\boldsymbol{Y}_2$, ρ_k^2 equals the squared multiple semi-partial correlation between y_{2k} and the set of predictors tested, with the predictors adjusted for all untested predictors in the model.

c. The population value of the UNIREP measure of multivariate association $\eta_u = \left[\sum_{k=1}^{b}\pi_k\rho_k^2/(1-\rho_k^2)\right] / \left[\sum_{k=1}^{b}\pi_k/(1-\rho_k^2)\right]$.

Proof

a. By inspection, it is clear that the three scenarios in Table II share (1) the same number of between-subject contrasts, (2) the same number of within-subject contrasts, (3) the same sample size, and (4) the same rank of design matrix, which ensure the three scenarios have the same a, b, N, and ν_e. In scenario S_2, $\boldsymbol{B}_2 = [\boldsymbol{BU} - \boldsymbol{C}'(\boldsymbol{CC}')^{-1}\boldsymbol{\Theta}_0]\,\mathbf{Y}$ gives the secondary parameter matrix $\boldsymbol{CB}_2\boldsymbol{I}_b = \boldsymbol{CBU}\mathbf{Y} - \boldsymbol{CC}'\boldsymbol{CC}')^{-1}\boldsymbol{\Theta}_0\mathbf{Y} = (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y}$ and the unscaled noncentrality for the hypothesis sums of squares $[(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y}]'\,M^{-1}\,[(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y}] = \boldsymbol{\Delta}_t$. With $E \sim \mathcal{N}_{N,p}(\mathbf{0}, \boldsymbol{I}_N, \boldsymbol{\Sigma})$, Theorem 8.12 in [14] ensures $E_2 = \boldsymbol{EU}\mathbf{Y} \sim \mathcal{N}_{N,b}[\mathbf{0}, \boldsymbol{I}_N, \mathrm{Dg}(\lambda)]$ for $\mathcal{V}[\mathrm{row}_i\,(\boldsymbol{E}_2)]' = \mathbf{Y}'\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}\mathbf{Y} = \mathbf{Y}'\boldsymbol{\Sigma}_*\,\mathbf{Y} = \mathrm{Dg}(\lambda)$. The eigenvalues of $\boldsymbol{\Sigma}_* = \mathbf{Y}\mathrm{Dg}(\lambda)\mathbf{Y}'$ (the hypothesis error covariance matrix for scenario S_1) are also the eigenvalues of $\mathrm{Dg}(\lambda)$ (hypothesis error covariance matrix for scenarios S_2 and S_3).

If $\boldsymbol{D} = \boldsymbol{CR}_{X+}\mathrm{Dg}(\boldsymbol{s}_{X+})^{-1}$, then singular value decomposition gives $\boldsymbol{D} = [\boldsymbol{L}_{D+}\boldsymbol{L}_{D0}]\mathrm{Dg}(\boldsymbol{s}_{D+}, \mathbf{0})[\boldsymbol{R}_{D+}\boldsymbol{R}_{D0}]' = \boldsymbol{L}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})\boldsymbol{R}'_{D+}$ with $\boldsymbol{R}_D = [\boldsymbol{R}_{D+}\boldsymbol{R}_{D0}]$. Also, $\boldsymbol{T}_D = \mathrm{Dg}(\boldsymbol{s}_{D+})^{-1}\boldsymbol{L}'_{D+}$ and $\boldsymbol{T}_D^{-1} = \boldsymbol{L}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})$. For scenario S_3, with $\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R}_D' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{\nu_e} \end{bmatrix}\boldsymbol{L}_X'$, Theorem 8.12 in [14] ensures $E_3 = \boldsymbol{TE}_2 \sim \mathcal{N}_{N,b}[\mathbf{0}, \boldsymbol{I}_N, \mathrm{Dg}(\lambda)]$ for $\mathcal{V}[\mathrm{col}_i(E_3)] = \boldsymbol{T}\boldsymbol{I}_N\boldsymbol{T}' = \begin{bmatrix} \boldsymbol{R}_D' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{\nu_e} \end{bmatrix}\boldsymbol{L}_X'\boldsymbol{L}_X\begin{bmatrix} \boldsymbol{R}_D & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{\nu_e} \end{bmatrix} = \boldsymbol{I}_N$. By Theorem 11.4 in [14], estimability of the secondary parameter matrix gives $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-(\boldsymbol{X}'\boldsymbol{X})$. The result holds for any generalized inverse and therefore for the Moore–Penrose: $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^+(\boldsymbol{X}'\boldsymbol{X}) = \boldsymbol{CR}_{X+}\boldsymbol{R}'_{X+}$. Scenario S_3 has primary parameter matrix $\boldsymbol{B}_3 = \boldsymbol{R}'_D\mathrm{Dg}(\boldsymbol{s}_{X+})\boldsymbol{R}'_{X+}\boldsymbol{B}_2$ and secondary parameter matrix $\boldsymbol{\Theta}_3 = [\boldsymbol{I}_a\ \mathbf{0}]\,\boldsymbol{B}_3\boldsymbol{I}_b = \boldsymbol{R}'_{D+}\,\mathrm{Dg}(\boldsymbol{s}_{X+})\boldsymbol{R}'_{X+})\,\boldsymbol{R}'_{X+}\boldsymbol{B}_2$. In turn, $\boldsymbol{T}_D^{-1}\boldsymbol{\Theta}_3 = \boldsymbol{D}\mathrm{Dg}(\boldsymbol{s}_{X+})\boldsymbol{R}'_{X+}\boldsymbol{B}_2 = \boldsymbol{CR}_{X+}\boldsymbol{R}'_{X+}\boldsymbol{B}_2 = \boldsymbol{CB}_2 = (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\boldsymbol{\Upsilon}$, and $\boldsymbol{\Theta}_3 = \boldsymbol{T}_D(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\mathbf{Y}$. Scenario S_3 has middle matrix $[\ \boldsymbol{I}_a\ \ \mathbf{0}\ ]\left([\ \boldsymbol{I}_r\ \ \mathbf{0}\ ]\begin{bmatrix} \boldsymbol{I}_r \\ \mathbf{0} \end{bmatrix}\right)^-\begin{bmatrix} \boldsymbol{I}_a \\ \mathbf{0} \end{bmatrix} = \boldsymbol{I}_a$. Using the fact that

$$\begin{aligned}\boldsymbol{T}'_{D}\boldsymbol{T}_{D}&=\boldsymbol{L}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})^{-2}\boldsymbol{L}'_{D+}\\&=\{[\boldsymbol{L}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})\boldsymbol{R}'_{D+}][\boldsymbol{R}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})\boldsymbol{L}'_{D+}]\}^{-1}\\&=(\boldsymbol{D}\boldsymbol{D}')^{-1}\left[\boldsymbol{C}\boldsymbol{R}_{X+}\mathrm{Dg}(\boldsymbol{s}_{X+})^{-2}\boldsymbol{R}'_{X+}\boldsymbol{C}'\right]^{-1}\\&=[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{C}']^{-1}\end{aligned}$$

allows writing the unscaled noncentrality for scenario S_3 as $[\boldsymbol{T}_D(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)\mathbf{Y}]'(\boldsymbol{I}_a)^{-1}\boldsymbol{T}_D(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)\mathbf{Y} = [\mathbf{Y}(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{C}']^{-1}[(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)\mathbf{Y} = \boldsymbol{\Delta}_t$. In turn, scenarios S_1 (which uses the original model) and S_2 and S_3 (which use transformed versions of the original model) have the same transformed, scaled noncentrality, $\boldsymbol{\Omega}_t$, for the hypothesis sums of squares. The power equivalence among scenarios S_1, S_2, and S_3 follows from part (b) of the corollary to Theorem 1.

b. The model for S_2, that is, $\boldsymbol{Y}_2 = \boldsymbol{X}\boldsymbol{B}_2 + \boldsymbol{E}_2$, contains b univariate models, namely $y_{2k} = \boldsymbol{X}\boldsymbol{\beta}_{2k} + \boldsymbol{e}_{2k}$ with column k of $\boldsymbol{E}_2$ being $\boldsymbol{e}_{2k} \sim \mathcal{N}(\mathbf{0}, \lambda_k\boldsymbol{I}_N)$ for $k \in \{1, 2, \cdots, b\}$. For each univariate model, the noncentrality parameter [14, Definition 2.6] is

$$\begin{aligned}\omega_k&=\boldsymbol{\beta}'_{2k}\boldsymbol{C}'\boldsymbol{M}^{-1}\boldsymbol{C}\boldsymbol{\beta}_{2k}/\lambda_k\\&=\boldsymbol{v}'_k\,[\boldsymbol{B}\boldsymbol{U}\\&\quad-\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{\Theta}'_0\boldsymbol{C}'\boldsymbol{M}^{-1}\boldsymbol{C}\,[\boldsymbol{B}\boldsymbol{U}\\&\quad-\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{\Theta}_0\big]\,\boldsymbol{v}_k/\lambda_k\\&=\boldsymbol{v}'_k(\boldsymbol{C}\boldsymbol{B}\boldsymbol{U}\\&\quad-\boldsymbol{\Theta}_0)'\boldsymbol{M}^{-1}(\boldsymbol{C}\boldsymbol{B}\boldsymbol{U}\\&\quad-\boldsymbol{\Theta}_0)\boldsymbol{v}_k/\lambda_k\\&=\boldsymbol{v}'_k\boldsymbol{\Delta}\boldsymbol{v}_k/\lambda_k,\end{aligned}\tag{18}$$

which is exactly the noncentrality parameter k for hypothesis testing scenario S_1. From the univariate theory [14, Equation 2.19]), $\omega_k = N\rho_k^2/(1-\rho_k^2)$. Also, $\rho_k^2 = \omega_k/(\omega_k + N)$ with ρ_k the multiple semi-partial correlation between y_{2k} and the set of predictors tested, with the predictors adjusted for all untested predictors in the model.

c. With $\pi_k = \lambda_k/\sum_{k=1}^{b}\lambda_k$ and definition of η_u in Equation (3),

$$\begin{aligned}\eta_u &= \frac{\mathrm{tr}(\boldsymbol{\Delta})}{\mathrm{tr}(\boldsymbol{\Delta})+N\mathrm{tr}(\boldsymbol{\Sigma}_*)} \\ &= \frac{\sum_{k=1}^{b}\lambda_k\omega_k}{\sum_{k=1}^{b}\lambda_k\omega_k+N\sum_{k=1}^{b}\lambda_k} \\ &= \frac{\sum_{k=1}^{b}[\lambda_k\rho_k^2/(1-\rho_k^2)]}{\sum_{k=1}^{b}\lambda_k[\rho_k^2/(1-\rho_k^2)+1]} \\ &= \frac{\sum_{k=1}^{b}\lambda_k\rho_k^2/(1-\rho_k^2)}{\sum_{k=1}^{b}\lambda_k/(1-\rho_k^2)} \\ &= \frac{\sum_{k=1}^{b}\pi_k\rho_k^2/(1-\rho_k^2)}{\sum_{k=1}^{b}\pi_k/(1-\rho_k^2)}\end{aligned}$$

Corollary 2

For any general linear multivariate model $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$ and hypothesis $H_0 : \boldsymbol{\Theta} = \boldsymbol{CBU} = \boldsymbol{\Theta}_0$ (scenario S_1), there exists a power-equivalent scenario (called scenario S_4) with model $\boldsymbol{Y}_4 = \boldsymbol{X}_4\boldsymbol{B}_4 + \boldsymbol{E}_4$ and hypothesis $H_0 : \boldsymbol{\Theta}_4 = \boldsymbol{0}$. Here, $\boldsymbol{X}_4'\boldsymbol{X}_4 = \boldsymbol{I}_r$, $\boldsymbol{B}_4' = [\ \beta_4 \quad \boldsymbol{0}\ ]$ for $b \times 1$ vector $\boldsymbol{\beta}_4$ with $\boldsymbol{\beta}_{4k} = (\pi_k\omega_k)^{1/2}$, $\mathcal{V}(\boldsymbol{E}_4) = \mathrm{Dg}(\pi)$, $\boldsymbol{C}_4 = [\boldsymbol{I}_a\ \boldsymbol{0}]$, $\boldsymbol{U}_4 = \boldsymbol{I}_b$, and $\boldsymbol{\Theta}_4 = \boldsymbol{C}_4\boldsymbol{\beta}_4\boldsymbol{U}_4$.

Proof

Scenario S_4 has unscaled noncentrality for hypothesis sum of squares $\boldsymbol{\Delta}_4 = \boldsymbol{\Theta}_4'[\boldsymbol{C}_4(\boldsymbol{X}_4'\boldsymbol{X}_4)\boldsymbol{C}_4']^{-1}\boldsymbol{\Theta}_4 = \boldsymbol{\beta}_4\boldsymbol{\beta}_4'$, and transformed, scaled noncentrality $\boldsymbol{\Omega}_4 = \boldsymbol{\beta}_4\boldsymbol{\beta}_4'\mathrm{Dg}(\pi)^{-1}$. In turn, diagonal element k of $\boldsymbol{\Omega}_4$ is $\omega_{4k} = \boldsymbol{\beta}_{4k}^2/\pi_k = \omega k$. Hence, by part (a) of Corollary 1, scenario S_4 is power equivalent to scenario S_1.

Theorem 3

a. By the method of moments, the distribution of t_u can be approximated by an F distribution with numerator degrees of freedom $ab\varepsilon_n$, denominator degrees of freedom $b\nu_e\varepsilon_d$, and noncentrality ω_u, that is

$$\Pr\{t_u \le f_0\} \approx \Pr\{F(ab\varepsilon_n, b\nu_e\varepsilon_d, \omega_u) \le f_0\}, \quad (19)$$

with

$$\varepsilon_n = \frac{\left(a+2\sum_{k=1}^{b}\pi_k\omega_k\right)}{ab\sum_{k=1}^{b}\pi_k^2+2b\sum_{k=1}^{b}\pi_k^2\omega_k} = \frac{\mathrm{tr}^2(\boldsymbol{\Sigma}_*)+2\mathrm{tr}(\boldsymbol{\Sigma}_*)\mathrm{tr}(\boldsymbol{\Delta}/a)}{b[\mathrm{tr}(\boldsymbol{\Sigma}_*^2)+2\mathrm{tr}(\boldsymbol{\Sigma}_*\boldsymbol{\Delta}/a)]} \quad (20)$$

$$\varepsilon_d = 1/\left(b\sum_{k=1}^{b}\pi_k^2\right) = \varepsilon = \frac{\mathrm{tr}^2(\boldsymbol{\Sigma}_*)}{b\mathrm{tr}(\boldsymbol{\Sigma}_*^2)} \quad (21)$$

$$\omega_u = \left(\sum_{k=1}^{b} \pi_k \omega_k\right) b\varepsilon_n = \frac{\mathrm{tr}(\boldsymbol{\Delta})}{\mathrm{tr}(\boldsymbol{\Sigma}_*)} b\varepsilon_n. \quad (22)$$

Under the null hypothesis, all $\omega_k = 0$, leading to $\varepsilon_n = \varepsilon$ and a central F approximation coinciding with the result given by Chi *et al*. [5].

b. One minimum set of sufficient constants and parameters for the F approximation consists of a, b, ν_e, ε_n, ε_d, and ω_u.

Proof

a. With $\pi_k = \lambda_k / \left(\sum_{k=1}^{b} \lambda_k\right)$, we apply Theorem 2 from [25] and show that for the form in (2), $\Pr\{t_u \le f_0\} \approx \Pr\{F(\nu_{*1}, \nu_{*2}, \omega_u) \le f_0 \varphi\}$ with $\phi = \lambda_{*2} \lambda_{*1}^{-1} a \nu_{*1}^{-1} \nu_{*2} \nu_e^{-1}$ and

$$\nu_{*1} = \frac{(a\sum_{k=1}^{b} \lambda_k)(a\sum_{k=1}^{b} \lambda_k + 2\sum_{k=1}^{b} \lambda_k \omega_k)}{a\sum_{k=1}^{b} \lambda_k^2 + 2\sum_{k=1}^{b} \lambda_k^2 \omega_k} = \frac{a^2 + 2a\sum_{k=1}^{b} \pi_k \omega_k}{a\sum_{k=1}^{b} \pi_k^2 + 2\sum_{k=1}^{b} \pi_k^2 \omega_k} \quad (23)$$

$$\nu_{*2} = \nu_e \left(\sum_{k=1}^{b} \lambda_k\right)^2 / \left(\sum_{k=1}^{b} \lambda_k^2\right) = \nu_e / \left(\sum_{k=1}^{b} \pi_k^2\right) \quad (24)$$

$$\omega_{*u} = \left(\sum_{k=1}^{b} \lambda_k \omega_k\right) \nu_{*1} / (a\sum_{k=1}^{b} \lambda_k = \left(\sum_{k=1}^{b} \pi_k \omega_k\right) \nu_{*1} / a \quad (25)$$

$$\lambda_{*1} = \left(a\sum_{k=1}^{b} \lambda_k\right) / \nu_{*1} \quad (26)$$

$$\lambda_{*2} = \left(\sum_{k=1}^{b} \lambda_k^2\right) / \left(\sum_{k=1}^{b} \lambda_k\right). \quad (27)$$

$$\begin{aligned} \varepsilon_n &= \nu_{*1}/(ab) \\ &= \left(a + 2\sum_{k=1}^{b} \pi_k \omega_k\right) / \left(ab\sum_{k=1}^{b} \pi_k^2 \right. \\ &\quad \left. + 2b\sum_{k=1}^{b} \pi_k^2 \omega_k\right), \varepsilon_d = \nu_{*2}/(b\nu_e) \\ &= 1 / \left(b\sum_{k=1}^{b} \pi_k^2\right) \\ &= \varepsilon, \omega_u \\ &= \left(\sum_{k=1}^{b} \pi_k \omega_k\right) b\varepsilon_n \end{aligned}$$

Hence, , and $\varphi = (\lambda_{*2}/\lambda_{*1})(ab/\nu_{*1})\nu_{*2}/(b\nu_e) = 1$. Under the null hypothesis, all $\omega_k = 0$, leading to $\varepsilon_n = 1 / \left(b\sum_{k=1}^{b} \pi_k^2\right) = \varepsilon$ and $\omega_u = 0$. (b) With Equation (19), it follows that {a, b,

ν_e, ε_n, ε_d, $\omega_u\}$ are a minimum-dimension set of constants and parameters sufficient for the F approximation.

References

1. Goeman J, Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics. 2007; 23:980–987. [PubMed: 17303618]
2. Brunner, E. Proceedings of the 6th St. Petersburg Workshop on Simulation. VVM.com Ltd.; 2009. Repeated measures under non-sphericity; p. 605-609.
3. Chen LS, Debashis P, Prentice RL, Wang P. A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. Journal of the American Statistical Association. 2011; 106:1345–1360. [PubMed: 23997374]
4. Chen SX, Qin Y. A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics. 2010; 38:808–835.
5. Chi YY, Gribbin M, Lamers Y, Gregory JF, Muller KE. Global hypothesis testing for high dimensional repeated measures outcomes. Statistics in Medicine. 2012; 31:724–742. [PubMed: 22161561]
6. Wu MC, Zhang L, Wang Z, Christiani DC, Lin X. Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection. Bioinformatics. 2009; 25:1145–1151. [PubMed: 19168911]
7. Srivastava MS, Du M. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis. 2008; 99:386–402.
8. Srivastava MS, Fujikoshi Y. Multivariate analysis of variance with fewer observations than the dimension. Journal of Multivariate Analysis. 2006; 97:1927–1940.
9. Warton DI. Penalized normal likelihood and ridge regularization of correlation and covariance matrices. Journal of the American Statistical Association. 2008; 103:340–349.
10. Muller KE, Barton CN. Approximate power for repeated-measures ANOVA lacking sphericity. Journal of the American Statistical Association. 1989; 84:549–555. Corrigenda 1991, 86: 255–256.
11. Muller KE, Edwards LJ, Simpson SL, Taylor DJ. Statistical tests with accurate size and power for balanced linear mixed models. Statistics in Medicine. 2007; 26:3639–3660. [PubMed: 17394132]
12. Ahmad MR, Werner C, Brunner E. Analysis of high-dimensional repeated measures designs: the one sample case. Computational Statistics and Data Analysis. 2008; 53:416–427.
13. Glueck DH, Muller KE. Adjusting power for a baseline covariate in a linear model. Statistics in Medicine. 2003; 22:2535–2551. [PubMed: 12898543]
14. Muller, KE.; Stewart, PW. Linear Model Theory: Univariate, Multivariate, and Mixed Models. New York: Wiley; 2006.
15. Srivastava, MS.; Khatri, CG. An Introduction to Multivariate Statistics. North-Holland: New York: 1979.
16. Taylor DJ, Muller KE. Computing confidence bounds for power and sample size of the general linear univariate model. American Statistician. 1995; 49:43–47. [PubMed: 24039272]
17. Chi YY, Muller KE. Two-Step hypothesis testing when the number of variables exceeds the sample size. Communications in Statistics – Simulation and Computation. 2013; 42:1113–1125. [PubMed: 24855328]
18. Gribbin MJ, Chi YY, Stewart PW, Muller KE. Confidence regions for repeated measures ANOVA power curves based on estimated covariance. BMC Medical Research Methodology. 2013; 13(1): 57. [PubMed: 23586676]
19. Browne RH. On the use of a pilot sample for sample size determination. Statistics in Medicine. 1995; 14:1933–1940. [PubMed: 8532986]
20. Taylor DJ, Muller KE. Bias in linear model power and sample size calculation due to estimating noncentrality. Communications in Statistics - Theory and Methods. 1996; 25:1595–1610.
21. Muller KE, Pasour VB. Bias in linear model power and sample size due to estimating variance. Communications in Statistics - Theory and Methods. 1997; 26:839–851.

22. Muller KE, LaVange LM, Ramey SL, Ramey CT. Power calculations for general linear multivariate models including repeated measures applications. Journal of the American Statistical Association. 1992; 87:1209–1226. [PubMed: 24790282]

23. Lenth RV. Some practical guidelines for effective sample size determination. American Statistician. 2011; 55:187–193.

24. Glueck DH, Muller KE. On the trace of a Wishart. Communications in Statistics - Theory and Methods. 1998; 27:2137–2141. Corrigendum 2002; 31: 159–160.

25. Kim H, Gribbin MJ, Muller KE, Taylor DJ. Analytic and computational forms for the ratio of a noncentral chi square and a Gaussian quadratic form. Journal of Computational and Graphical Statistics. 2006; 15:443–459. [PubMed: 23843686]

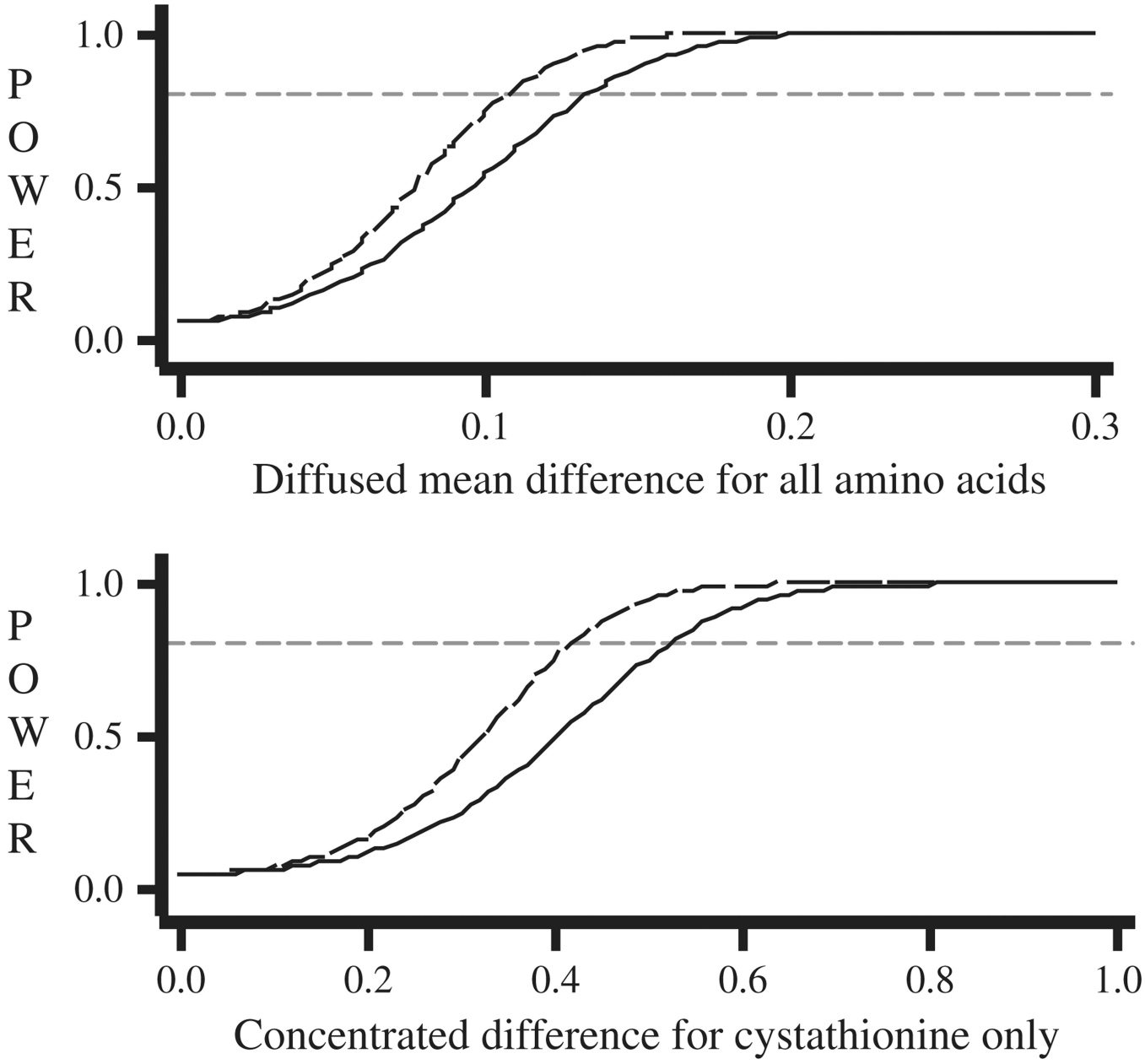

Figure 1.

Approximate overall power for replicating the vitamin B6 study when sample size is 10 (solid curve) or 15 (dashed curve). The top plot is for diffused effects occurring across all amino acids, and the bottom plot is for a concentrated effect on cystathionine only. The horizontal axis is the mean difference of logarithmically transformed concentration.

Table I

Parameters and constants for general linear multivariate model $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$ and associated general linear hypothesis $H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$.

Symbol	Size	Definition and properties
N	1×1	Sample size
p	1×1	Number of outcome variables
q	1×1	Number of predictors
$\boldsymbol{X} = \boldsymbol{L}_{X+}\mathrm{Dg}(\boldsymbol{s}_{X+})\boldsymbol{R}'_{X+}$	$N \times q$	Fixed, known design matrix
$\boldsymbol{R}_{X+}$	$q \times r$	Range eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{R}'_{X+}\boldsymbol{R}_{X+} = \boldsymbol{I}_r$
$\boldsymbol{L}_X = [\boldsymbol{L}_{X+}\ \boldsymbol{L}_{X0}]$	$N \times N$	All eigenvectors of $\boldsymbol{XX}'$, $\boldsymbol{L}'_X\boldsymbol{L}_X = \boldsymbol{I}_N$
$r = \mathrm{rank}(\boldsymbol{X})$	1×1	Rank of design matrix
$\nu_e = N - r$	1×1	Error degrees of freedom
$\boldsymbol{B}$	$q \times p$	Primary (mean) parameters
a	1×1	Number of between-subject contrasts
b	1×1	Number of within-subject contrasts
$\boldsymbol{C}$	$a \times q$	Between-subject contrast matrix
$\boldsymbol{U}$	$p \times b$	Within-subject contrast matrix
$\boldsymbol{\Theta} = \boldsymbol{CBU}$	$a \times b$	Secondary parameters
$\boldsymbol{\Theta}_0$	$a \times b$	Null values for testing $\boldsymbol{\Theta}$
$\boldsymbol{M} = \boldsymbol{C}\,(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$	$a \times a$	Middle matrix
$\boldsymbol{\Delta} = (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)'\boldsymbol{M}^{-1}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)$	$b \times b$	Unscaled noncentrality of $\hat{\boldsymbol{\Delta}}$
$\boldsymbol{\Sigma}$	$p \times p$	Error covariance, $\mathcal{V}\{[\mathrm{row}_i(\boldsymbol{E})]'\}$
$\boldsymbol{\Sigma}_* = \boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{\Upsilon}\mathrm{Dg}(\boldsymbol{\lambda})\boldsymbol{\Upsilon}'$	$b \times b$	Hypothesis error covariance, $\mathcal{V}\{[\mathrm{row}_i(\boldsymbol{EU})]'\}$
$\boldsymbol{\lambda} = [\lambda_1\ \lambda_2 \cdots \lambda_b]'$	$b \times 1$	Vector of eigenvalues of $\boldsymbol{\Sigma}_*$
$\boldsymbol{\Upsilon} = [\upsilon_1\ \upsilon_2 \cdots \upsilon_b]$	$b \times b$	Eigenvectors of $\boldsymbol{\Sigma}_*$, $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}' = \boldsymbol{\Upsilon}'\boldsymbol{\Upsilon} = \boldsymbol{I}_b$
$\lambda_+ \sum_{k=1}^{b} \lambda_k$	1 ×1	Sum of eigenvalues of $\boldsymbol{\Sigma}_*$
$\pi_k = \lambda_k/\lambda_+$	1 ×1	The kth scaled eigenvalue of $\boldsymbol{\Sigma}_*$
$\varepsilon = \left(\sum_{k=1}^{b} \lambda_k\right)^2 / \left(b \sum_{k=1}^{b} \lambda_k^2\right)$	1 ×1	Sphericity parameter
$\boldsymbol{\Omega} = \boldsymbol{\Delta}\boldsymbol{\Sigma}_*^{-1}$	$b \times b$	Scaled noncentrality of $\hat{\boldsymbol{\Delta}}$
$\boldsymbol{\Delta}_t = \boldsymbol{\Upsilon}'\boldsymbol{\Delta}\boldsymbol{\Upsilon}$	$b \times b$	Transformed, unscaled noncentrality of $\hat{\boldsymbol{\Delta}}$
$\boldsymbol{\Omega}_t = \boldsymbol{\Delta}_t\mathrm{Dg}(\boldsymbol{\lambda})^{-1}$	$b \times b$	Transformed, scaled noncentrality of $\hat{\boldsymbol{\Delta}}$
$\omega_k = \boldsymbol{\upsilon}_k'\boldsymbol{\Delta}\boldsymbol{\upsilon}_k/\lambda_k = N\rho_k^2/(1-\rho_k^2)$	1 ×1	The kth diagonal element of $\boldsymbol{\Omega}_t$
$\rho_k^2 = \omega_k/(\omega_k + N)$	1 ×1	Squared semi-partial correlation for ω_k

Table II

UNIREP power-equivalent hypothesis testing scenarios based on transformed models sharing the same N, ν_e, a, and b.

	Hypothesis testing scenario		
Feature	$\boldsymbol{S_1}$	$\boldsymbol{S_2}$	$\boldsymbol{S_3}$
Model	$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$	$\boldsymbol{Y}_2 = \boldsymbol{XB}_2 + \boldsymbol{E}_2$ $\boldsymbol{YU\Upsilon} = \boldsymbol{XB}_2 + \boldsymbol{EU\Upsilon}$	$\boldsymbol{Y}_3 = \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{B}_3 + \boldsymbol{E}_3$ $\boldsymbol{Y}_3 = \begin{bmatrix} \boldsymbol{R'}_D \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{B}_2 + \boldsymbol{E}_3$ $\boldsymbol{TY}_2 = \boldsymbol{TXB}_2 + \boldsymbol{TE}_2$
Outcome space	Original data	Principal components in hypothesis space	Transformed components in hypothesis space
Number of outcomes	p	b	b
Number of predictors	q	q	r
Primary parameters	$\boldsymbol{B}$	$\boldsymbol{B}_2 = [\boldsymbol{BU} - \boldsymbol{C}'(\boldsymbol{CC}')^{-1}\boldsymbol{\Theta}_0]\boldsymbol{\Upsilon}$	$\boldsymbol{B}_3 = \boldsymbol{R}'_D \mathrm{Dg}(\boldsymbol{s}_{X+}) \boldsymbol{R}'_{X+} \boldsymbol{B}_2$
Error covariance	$\mathcal{V}(\boldsymbol{E}) = \boldsymbol{\Sigma}$	$\mathcal{V}(\boldsymbol{E}_2) = \mathrm{Dg}(\boldsymbol{\lambda})$	$\mathcal{V}(\boldsymbol{E}_3) = \mathrm{Dg}(\boldsymbol{\lambda})$
Between-subject contrast	$\boldsymbol{C}$	$\boldsymbol{C}$	$[\boldsymbol{I}_a \ \boldsymbol{0}]$
Within-subject contrast	$\boldsymbol{U}$	$\boldsymbol{I}_b$	$\boldsymbol{I}_b$
Secondary parameters	$\boldsymbol{CBU} = \boldsymbol{\Theta}$	$(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\boldsymbol{\Upsilon}$	$\boldsymbol{T}_D(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)\boldsymbol{\Upsilon}$
Hypothesis error covariance	$\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{\Sigma}_*$	$\boldsymbol{I}_b' \mathrm{Dg}(\boldsymbol{\lambda}) \boldsymbol{I}_b = \mathrm{Dg}(\boldsymbol{\lambda})$	$\mathrm{Dg}(\boldsymbol{\lambda})$
Null hypothesis	$H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$	$H_0 : (\boldsymbol{\Theta} = \boldsymbol{\Theta}_0)\boldsymbol{\Upsilon} = \boldsymbol{0}$	$H_0 : \boldsymbol{T}_D(\boldsymbol{\Theta} = \boldsymbol{\Theta}_0)\boldsymbol{\Upsilon} = \boldsymbol{0}$
Unscaled noncentrality	$\boldsymbol{\Delta}$	$\boldsymbol{\Upsilon}'\boldsymbol{\Delta}\boldsymbol{\Upsilon} = \boldsymbol{\Delta}_t$	$\boldsymbol{\Delta}_t$

$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R}'_D & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\nu_e} \end{bmatrix} \boldsymbol{L}'_X$ for $\boldsymbol{R}_D$ eigenvectors of $\boldsymbol{D}'\boldsymbol{D}$ with $\boldsymbol{D} = \boldsymbol{CR}_{X+}\mathrm{Dg}(\boldsymbol{s}_{X+})^{-1}$.

$\boldsymbol{T}_D = \mathrm{Dg}(\boldsymbol{s}_{D+})^{-1}\boldsymbol{L}'_{D+}$ for singular value decomposition $\boldsymbol{D} = \boldsymbol{L}_{D+}\mathrm{Dg}(\boldsymbol{s}_{D+})\boldsymbol{R}'_{D+}$.

Table III

Absolute difference between empirical and approximated power for a one-sample design.

			Empirical, absolute bias in power			
Number of outcomes	**Number of conditions**	**Approximated power**	**Min**	**Max**	**Mean**	**Standard deviation**
64	36	0.20	0.000	0.010	0.004	0.003
64	36	0.50	0.000	0.024	0.004	0.005
64	36	0.80	0.000	0.005	0.002	0.001
64	36	0.90	0.000	0.013	0.002	0.003
256	36	0.80	0.000	0.005	0.002	0.001
256	36	0.90	0.000	0.005	0.002	0.001
1024	36	0.80	0.000	0.004	0.001	0.001
1024	36	0.90	0.000	0.003	0.001	0.001

Each line summarizes 36 simulation conditions in a four-way factorial: $N \in \{10, 20, 40\}$, $\varepsilon \in \{0.27, 0.56, 0.76\}$, number of nonzero ρ_k^2 of 4 or 32, and location of nonzero ρ_k^2 at either the most dominant or middle components in hypothesis space (100,000 replications per condition).

Table IV

Absolute difference between empirical and approximated power for a two-sample design.

			Empirical, absolute bias in power			
Number of outcomes	**Number of conditions**	**Approximated power**	**Min**	**Max**	**Mean**	**Standard deviation**
64	36	0.20	0.001	0.001	0.006	0.003
64	36	0.50	0.000	0.034	0.011	0.010
64	36	0.80	0.001	0.037	0.011	0.010
64	36	0.90	0.001	0.028	0.009	0.006
256	36	0.80	0.001	0.031	0.010	0.009
256	36	0.90	0.001	0.024	0.009	0.006
1024	36	0.80	0.000	0.027	0.011	0.009
1024	36	0.90	0.001	0.022	0.009	0.006

Each line summarizes 36 simulation conditions in a four-way factorial: $N \in \{10, 20, 40\}$, $\varepsilon \in \{0.27, 0.56, 0.76\}$, number of nonzero ρ_k^2 of 4 or 32, and location of nonzero ρ_k^2 at either the most dominant or middle components in hypothesis space (100,000 replications per condition).

Table V

Empirical and approximated power for a two-sample design with $b = 64$, and approximated power is 0.90.

Sample size	Sphericity parameter ε	Number of nonzero ρ_k^2	Location of nonzero ρ_k^2	Empirical power	Absolute bias
10	0.27	4	Top	0.909	0.009
10	0.27	4	Middle	0.929	0.029
10	0.27	32	Top	0.914	0.014
10	0.27	32	Middle	0.921	0.021
10	0.56	4	Top	0.909	0.009
10	0.56	4	Middle	0.916	0.016
10	0.56	32	Top	0.912	0.012
10	0.56	32	Middle	0.916	0.016
10	0.76	4	Top	0.910	0.010
10	0.76	4	Middle	0.914	0.014
10	0.76	32	Top	0.911	0.011
10	0.76	32	Middle	0.914	0.014
20	0.27	4	Top	0.904	0.004
20	0.27	4	Middle	0.920	0.020
20	0.27	32	Top	0.906	0.006
20	0.27	32	Middle	0.912	0.012
20	0.56	4	Top	0.905	0.005
20	0.56	4	Middle	0.905	0.005
20	0.56	32	Top	0.904	0.004
20	0.56	32	Middle	0.905	0.005
20	0.76	4	Top	0.903	0.003
20	0.76	4	Middle	0.905	0.005
20	0.76	32	Top	0.906	0.006
20	0.76	32	Middle	0.906	0.006
40	0.27	4	Top	0.903	0.003
40	0.27	4	Middle	0.919	0.019
40	0.27	32	Top	0.904	0.004

Sample size	Sphericity parameter ε	Number of nonzero ρ_k^2	Location of nonzero ρ_k^2	Empirical power	Absolute bias
40	0.27	32	Middle	0.909	0.009
40	0.56	4	Top	0.903	0.003
40	0.56	4	Middle	0.903	0.003
40	0.56	32	Top	0.904	0.004
40	0.56	32	Middle	0.905	0.005
40	0.76	4	Top	0.903	0.003
40	0.76	4	Middle	0.901	0.001
40	0.76	32	Top	0.903	0.003
40	0.76	32	Middle	0.903	0.003

Location of nonzero ρ_k^2 can be at either the most or middle dominant components in hypothesis space.

Table VI

Absolute difference between empirical and approximated power for a three-sample design.

			Empirical, absolute bias in power			
Number of outcomes	**Number of conditions**	**Approximated power**	**Min**	**Max**	**Mean**	**Standard deviation**
64	36	0.20	0.000	0.023	0.007	0.008
64	36	0.50	0.000	0.025	0.008	0.009
64	36	0.80	0.000	0.015	0.005	0.005
64	36	0.90	0.000	0.009	0.004	0.003
256	36	0.80	0.000	0.076	0.027	0.031
256	36	0.90	0.000	0.057	0.020	0.023
1024	36	0.80	0.000	0.076	0.027	0.031
1024	36	0.90	0.000	0.056	0.020	0.023

Each line summarizes 36 simulation conditions in a four-way factorial: $N \in \{10, 20, 40\}$, $\varepsilon \in \{0.27, 0.56, 0.76\}$ number of nonzero ρ_k^2 of 4 or 32, and location of nonzero ρ_k^2 at either the most dominant or middle components in hypothesisspace (100,000 replications per condition).