Published in final edited form as:

Commun Stat Theory Methods. 2010 December 1; 39(20): 3717–3738. doi:10.1080/03610920903353709.

Combining an Internal Pilot with an Interim Analysis for Single Degree of Freedom Tests

John A. Kairalla^{1,*}, Keith E. Muller¹, and Christopher S. Coffey²

¹ Division of Biostatistics, Department of Epidemiology and Health Policy Research, College of Medicine, University of Florida. PO Box 100177, Gainesville, FL 32610, USA

² Department of Biostatistics, School of Public Health, University of Iowa, Iowa City, IA 52240, USA

Abstract

An internal pilot with interim analysis (IPIA) design combines interim power analysis (an internal pilot) with interim data analysis (two stage group sequential). We provide IPIA methods for single df hypotheses within the Gaussian general linear model, including one and two group t tests. The design allows early stopping for efficacy and futility while also re-estimating sample size based on an interim variance estimate. Study planning in small samples requires the exact and computable forms reported here. The formulation gives fast and accurate calculations of power, type I error rate, and expected sample size.

Keywords

Adaptive designs; Power; Sample size re-estimation

1. Introduction

Choosing a sample size often reflects a mix of ethical, monetary, and practical considerations. A study of diet in children with autism (Elder et al., 2006) illustrates the tradeoffs. The randomized, crossover design compared autistic symptoms of either a gluten and casein free diet or a placebo diet. Ethically, scientists may wish to stop early for efficacy or futility if the effects of the intervention are clear. Time and resource are minimized by using as small a sample size as possible, while statistical power will be maximized with as large a sample size as possible. As often happens with truly innovative science, the investigators may have substantial uncertainty about nuisance parameters, such as the variance for a t test in the example given. The desire for minimal sample sizes and correctly powered designs have led statisticians to develop group sequential (GS) and internal pilot (IP) designs.

GS and IP designs have distinct strengths and weaknesses in planning and conducting research. GS designs reduce sample size under the alternative hypothesis by frequently monitoring studies to stop early if effect sizes differ from those planned. GS designs, however, are vulnerable to power loss if nuisance parameter assumptions do not hold. IP designs use one interim power analysis, which corrects for mis-specification of nuisance parameters in order to insure study power or avoid unnecessarily large sample size. IP designs, however, do not reduce expected sample size when effect sizes are larger or smaller than planned. While both designs have advantages, neither simultaneously achieves the goals of informative early stopping ability and assurance of a properly powered study.

*jak@biostat.ufl.edu.

Our goal is a two stage design that maintains and protects study power by updating sample size needs while incorporating early stopping procedures, especially for one or two group designs. Exact distributional results with computable formulae for power and sample size would allow researchers to accurately and efficiently explore design properties without simulations before conducting a study, even in small sample situations. Additionally, the exact results will assist in the refinement and comparisons of current and future methodological procedures.

The NIH Roadmap explicitly highlighted the importance of small sample theory (e.g., Clinical and Translational Science Awards, RFA-RM-07-002 U54). The analytic results in this article can help determine small sample properties for many study designs in a two stage framework. The exact theory can facilitate studying and comparing properties of new methods to ascertain ones with the most desirable features for a particular study. For example, unadjusted large sample critical values may cause inflation to the type I error rate in small sample studies. A new method for determining boundary points promises to lead to a more unbiased testing procedure.

We restrict attention to designs which combine interim analysis and sample size re-estimation. Stein's (1945) two stage design, which uses variance information only from the first stage, provides the foundation for early sequential testing methods (Baker, 1950; Hall, 1962; Arghami and Billard, 1992; Hochberg and Marcus, 1983). All of these have the disadvantage of incorporating only early stage variance information in the test statistics. Facey (1992) described a Phase 2 trial design using the triangular test stopping bounds (Whitehead and Straton, 1983). Gould and Shih (1998) and Whitehead et al. (2001) used simulation to examine the use of a blinded variance estimate from the initial stage to possibly increase future sample sizes. Denne and Jennison (2000) proposed a group sequential t -test with sample size update based on variance for a two-sided one group test of mean with early stopping under the null. Morgan (2003) considered by simulation a sample size re-estimation method in group sequential trials in order to extend the idea for use in group-sequential response-adaptive designs for Gaussian data. Another approach to clinical trial monitoring with nuisance parameter based sample size adjustment is the large sample information-based approach described by Mehta and Tsiatis (2001) and Tsiatis (2006). They described a method for use within group sequential designs that allow for early stopping while updating the estimated maximum sample size at each analysis stage as nuisance parameter estimates are updated.

The remainder of the paper consists of five parts. Section 2 contains details of the IPIA model, section 3 includes the specifics of the recommended procedure, section 4 provides key analytic results, section 5 describes study characteristics through an example, and section 6 is a closing discussion that centers on the advantages and disadvantages of the design, as well as their implications for future research.

2. The IPIA Model

2.1 Notation

Notational conventions follow Muller and Stewart (2006, Chapter 1). An $r \times 1$ vector (always a column) is written $\mathbf{a}$, and an $r \times c$ matrix is written $\mathbf{A} = \{a_{j,k}\}$, with transpose $\mathbf{A}'$. For full rank matrix $\mathbf{A}$, we will use $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})' = \mathbf{A}^{-t}$. Here $\mathbf{1}_r$ always represents an $r \times 1$ vector of 1's and define $\mathbf{I}_r$ as the $r \times r$ identity matrix with $\mathbf{I}_r = \text{Dg}(\mathbf{1}_r)$. The direct (Kronecker) product is defined as $\mathbf{A} \otimes \mathbf{B} = \{a_{j,k}\mathbf{B}\}$.

Detailed information about all basic random variables discussed can be found in Johnson et al. (1994, 1995). Writing $\mathbf{x} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ shows that $\mathbf{x}$ ($n \times 1$) has vector Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

When Σ may be non-full rank, $\mathbf{x}$ has (possibly) singular Gaussian distribution, written $\mathbf{x} \sim (S) \mathcal{N}_s(\boldsymbol{\mu}, \Sigma)$. Writing $X \sim \chi^2(\nu, \omega)$ means X is a chi-square variable with ν degrees of freedom (df) and noncentrality ω . Likewise, writing $X \sim F(\nu_1, \nu_2, \omega)$ indicates that X follows a F distribution with numerator df ν_1 denominator df ν_2 , and noncentrality ω . Writing $\chi^2(\nu)$ or $F(\nu_1, \nu_2)$ implies $\omega = 0$. Writing $X \sim \chi^2_T(\nu; t_L, t_U)$ indicates X follows a doubly-truncated central chi-square distribution with ν df, truncated to $[t_L, t_U]$ (Coffey and Muller, 2000). For random U and parameters $\gamma_1 \dots \gamma_k$, the cumulative distribution function (CDF) taken at u is $F_U(u; \gamma_1 \dots \gamma_k)$. As a special case, $\Phi(z)$ indicates the CDF of the scalar Gaussian distribution with mean zero and unit variance. In general, $F_U^{-1}(\alpha; \gamma_1 \dots \gamma_k)$ indicates the α quantile of a random variable U with parameters $\gamma_1 \dots \gamma_k$.

2.2 The IPIA Model

The IPIA model generalizes the two stage internal pilot introduced in Wittes and Brittain (1990). In turn, the exact, small sample IPIA theory for the general linear univariate model with Gaussian errors generalizes the results in Coffey and Muller (1999). The introduction of early stopping requires a more general notation. In an IP design, N_+ with $n_{+, \min} \leq N_+ \leq n_{+, \max}$ indicates the random final sample size calculated using $\hat{\sigma}_1^2$, the variance estimate from the interim sample, bounded by predetermined limits. For the IPIA model, random N_+ with constraints $n_1 \leq N_+ \leq n_{+, \max}$ (with n_1 the interim sample size) also uses $\hat{\sigma}_1^2$ and determines the random variable $N_2 = N_+ - n_1$. However, because of the possible early stopping, N_+ may not be the final sample size for the study. In an IPIA design, the random final sample size is $N_w = n_1 + N_2 \cdot \mathcal{I}$ (continue) with $\mathcal{I}$ an event indicator equal to 1 if study continues after the first stage. Therefore $N_w = \begin{cases} n_1 & \text{if study stopped after first stage} \\ N_+ & \text{otherwise} \end{cases}$. The design leads to interest in three different but intimately connected models. The combined model for the final analysis may be written as

$$\begin{matrix} \mathbf{y}_+ \\ N_+ \times 1 \end{matrix} = \begin{matrix} \mathbf{X}_+ \boldsymbol{\beta} \\ N_+ \times q \times 1 \end{matrix} + \begin{matrix} \mathbf{e}_+ \\ N_+ \times 1 \end{matrix}, \text{ or} \quad (1)$$

$$\begin{bmatrix} \mathbf{y}_1 \\ n_1 \times 1 \\ \mathbf{y}_2 \\ N_2 \times 1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ n_1 \times q \\ \mathbf{X}_2 \\ N_2 \times q \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{e}_1 \\ n_1 \times 1 \\ \mathbf{e}_2 \\ N_2 \times 1 \end{bmatrix}, \quad (2)$$

with partitioning corresponding to the fixed n_1 and random N_2 observations in the first and second samples, respectively. The second sample of size $N_2 = N_+ - n_1$ shown above is taken only if in analysis of the first sample a) no stopping rules are achieved and b) $\hat{\sigma}_1^2$ and the sample size re-estimation rule imply the need for $N_+ > n_1$ ($N_2 > 0$). The special case of $N_+ = n_1$ collapses the full model to the interim model. Model components include random observed data $\mathbf{y}_+$ ($N_+ \times 1$) (independent sampling units as rows), design matrix of fixed form $\mathbf{X}_+$, and unobserved $\mathbf{e}_+ \sim \mathcal{N}_s(\mathbf{0}, \sigma^2 \mathbf{I}_{N_+})$. For computational convenience, random values of sample size, $N_+ = n_1 + N_2$, increase only in multiples of a replication factor, m , in order to maintain the desired group allocation. For example, a balanced 2-group study design would have $m = 2$, while a 2-group study with a planned 2:1 sample size ratio would have $m = 3$. Given the restriction, for some $m \times q \mathbf{X}_0$ it follows that $\mathbf{X}_1 = \mathbf{1}_{k_1} \otimes \mathbf{X}_0$ and $\mathbf{X}_2 = \mathbf{1}_{k_2} \otimes \mathbf{X}_0$, with fixed k_1 and random k_2 the number of replications in the first and second samples, respectively. Consequently the columns

of X_1 and X_2 span the same space (when $K_2 > 0$), which allows writing $r = \text{rank}(X_1) = \text{rank}(X_2) = \text{rank}(X_+)$. In order to simplify computations and some discussions, attention will usually be restricted to a full rank design, that is $r = \text{rank}(X_0) = q$. The principles of linearly equivalent models allow the restriction without meaningful loss of generality (Muller and Stewart, 2006, Chapter 11).

The test of interest is $H_0: \theta = \theta_0$, with C fixed $a \times q$ contrast and $\theta = C\beta$. Without loss of generality we assume $\theta_0 = 0$ (see Lemma A.1, in Appendix). For a ‘scientifically important’ effect of interest ($\theta = \theta_1$), we seek a design that ensures a target type I error rate (α_t) with sample size appropriate to achieve target power (P_t). Throughout, subscript $s \in \{1, +\}$ indicates a value for either the model based on the internal pilot (first) sample or the total combined sample (conditioned on $N_+ = n_+$). Error df are $v_s = n_s - r$. We use functional notation in many places to emphasize the dependence on n_s . For example, for the model based on sample s , the ‘hat’ matrix is $H_s = X_s(X_s'X_s)^{-1}X_s'$, while $\hat{\theta}(n_s) = C(X_s'X_s)^{-1}X_s'y_s$ and $\hat{\sigma}^2(n_s) = y_s'(I_{n_s} - H_s)y_s/v_s$ represent the unadjusted estimates of θ and σ^2 . Similarly, the ‘middle’ matrix is $M_s = C(X_s'X_s)^{-1}C'$ and the noncentrality parameter is $\delta_s = \theta' M_s^{-1} \theta$. In turn, for the model based on sample s , $\hat{\delta}(n_s) = \hat{\theta}(n_s)' M_s^{-1} \hat{\theta}(n_s)$ is the observed hypothesis sum of squares and the corresponding unadjusted test statistic for stage s is $F(n_s) = [\hat{\delta}(n_s)/a]/\hat{\sigma}^2(n_s)$. When there is no confusion, the functional aspects of the forms will be implied with a subscript, e.g., $\hat{\theta}(n_s)$ may be written $\hat{\theta}_s$ and $F(n_s)$ may be written $F_s = (\hat{\delta}_s/a)/\hat{\sigma}_s^2$. We consider only testable hypotheses, which require full rank C as well as $C = (X_s'X_s)^{-1}(X_s'X_s) = C$ (the latter needed only for less than full rank designs). Table 1 summarizes relevant dimensions, Table 2 summarizes parameters and constants, and Table 3 summarizes design factors for the study. The appendix also details some model properties and their derivations.

3. The IPIA Procedure

Figure 1 describes via flowchart the general procedure for the IPIA model. The order of the steps matters in specifying the distributions. First, an initial study design is specified with target power and type I error rate, design matrix, hypotheses, effect size of interest, and best guess at variability. Next, a first stage sample size is determined based on the design. After then collecting the first stage sample, the updated variability information from the sample is used to calculate the second stage sample to be used if the study is continued. The first stage test statistic and critical values for the first and second stages are then computed. The decision is made to either to stop the study under the null hypothesis, stop the study under the alternative hypothesis, or to continue the study and collect the second stage sample. If continued, the second stage sample is collected, the final test statistic using all information is solved for, and a decision is made as to whether the null hypothesis can be rejected.

The value of the sample size for the internal pilot stage, n_1 , should be made at the design stage of the study. Lower values give less certain estimates of σ^2 while higher values reduce possible savings. Although there has been discussion concerning appropriate choices for n_1 in IP settings (Birkett and Day, 1994; Coffey and Muller, 1999; Denne and Jennison, 1999), no simple rules apply. Rather, the particular scientific setting seems to dictate reasonable choices. Expecting a similar conclusion in the current context, we will default to a choice of roughly half of the pre-planned sample size and explore the effect of the initial sample size choice in the IPIA design in future work.

Calculation of the three critical values for the study must follow rules pre-specified in the study protocol. The critical values may depend on n_+ , the realized value of N_+ ; however, when clear, they will be referred to as f_L (lower first stage), f_U (upper first stage), and f_+ (upper second

stage). Ideally, they should control the type I error rate and have good power and sample size properties. The α -spending technique used in group-sequential designs provides a reasonable first approach. The example in section 5 uses O'Brien-Fleming (1979) type boundaries to determine critical values. The sample size fraction at the interim stage (known once n_+ has been determined), specifies the α levels for critical value calculations. The theory here allows for stopping under the null at the interim analysis if $F_1 < f_L$. This can greatly reduce expected sample size with effects near the null by allowing the study to stop early for a "lost cause". If no early stopping under the null is allowed, then $f_L = 0$ for all $n_+ \neq n_1$. When $n_+ = n_1$, no further subjects are sampled. Hence $f_L = f_U$, which guarantees stopping for acceptance or rejection of the null. More detailed exploration and refinement of methods for selecting critical values will be undertaken in future papers. In this paper we are using unadjusted Z-based critical values calculated with O'Brien-Fleming like boundaries. Small studies could use t (or F) based critical values to account for variance estimation. Additionally, we believe a *bounding* method, such as that described by Coffey and Muller (2001) for internal pilots, could be developed for and work well for IPIA. In such a design, the nominal α level is algorithmically decreased until the overall type I error rate is controlled for the worst case scenario, and thus for all cases. The adjusted α is then used as the basis for critical value calculations. For internal pilots, this method protects type I error rate with very little loss of power. Extending the method to the IPIA setting requires additional work.

The choice of sample size re-estimation rule determines the distribution of N_+ . The choice affects the type I error rate, power, and expected sample size. An approach similar to that described by Coffey and Muller (1999) for the IP setting will be used here. As in IPs, an IPIA design determines the final sample size by using the variance estimate from the interim stage to recalculate the sample size need. In order to determine the distribution of the sample size, one solves the cumulative distribution function for possible values of N_+ , that is $\Pr(N_+ \leq n)$, by determining cut-off points using the first stage variance estimate. For a particular value n_+ of N_+ , one solves for the scaled noncentrality parameter $\lambda(n_+)$ that satisfies $P_t = 1 -$

$F_{\chi^2}[f_{\text{crit}}; a, \lambda(n_+)]$ or $P_t = 1 - F_F[f_{\text{crit}}; a, v_+, \lambda(n_+)]$ where $f_{\text{crit}} = F_{\chi^2}^{-1}(1 - \alpha_t; a)$ or $F_F^{-1}(1 - \alpha_t; a, v_+)$ depending on whether or not large sample distributional assumptions are used. The scaled noncentrality parameter $\lambda(n_+)$ represents the minimum value that would lead to a final sample size of n_+ . Since the effect of interest, θ_1 , is used at the planning stage, for $\delta(n_+) = \theta_1' M_+^{-1} \theta_1$ we have $\lambda(n_+) = \delta(n_+)/\sigma^2(n_+)$. Here, $\sigma^2(n_+)$ designates the largest value of $\hat{\sigma}_1^2$ that would produce $N_+ = n_+$. Therefore, since $v_1 \hat{\sigma}_1^2 \sim \sigma^2 W$ for $W \sim \chi^2(v_1)$,

$$\Pr(N_+ \leq n_+) = \Pr[\hat{\sigma}_1^2 \leq \sigma^2(n_+)] = \Pr\{W \leq v_1 \delta(n_+)/[\sigma^2 \lambda(n_+)]\}.$$

The discreteness of sample size implies $\Pr(N_+ = n_+) = \Pr(N_+ \leq n_+) - \Pr(N_+ \leq n_+ - m)$. When restrictions are given for minimum or maximum sample size, the tail probabilities are collapsed into values determined by the smallest or largest allowable values of N_+ , respectively.

A key result of the process is the calculation of cut-off points that determine a range into which continuous $\hat{\sigma}_1^2$ must have fallen in order for a given final sample size to occur. Here $q_1(n_+)$ and $q_2(n_+)$ are such that $q_1(n_+) < v_1 \hat{\sigma}_1^2 / \sigma^2 \leq q_2(n_+)$, with $\Pr(N_+ = n_+) = F_{\chi^2}[q_2(n_+); v_1] - F_{\chi^2}[q_1(n_+); v_1]$. The cut off points determine the probabilities for discrete values of N_+ and hence describe the variable's distribution. When it is unambiguous, q_1 and q_2 are used for $q_1(n_+)$ and $q_2(n_+)$.

One minor difference between this technique and group sequential methods is the lack of sample size inflation factor to account for multiple looks. While it could be integrated without much difficulty, it would have very little effect on this two stage design. For example, using O'Brien-Fleming bounds for $P_t = 0.9$ and $\alpha_t = 0.05$ calls for a sample size inflation factor of 1.007 for an interim look at one half of the maximum information time (Jennison and Turnbull, 2000; Chapter 2). We simplify by leaving it out; it can always be added if deemed important.

4. Key Analytic Results

The two stage IPIA results in this paper are developed to test $H_0: \theta = 0$ with $\theta = C\beta$ for the case where the contrast matrix C has only one row, namely, $a = 1$. Tests falling under this restriction include one and two group mean comparisons, tests for interaction among factors with only two levels, as well as any other single df contrasts. Proofs for all of the results in this section as well as related results and definitions can be found in the accompanying appendix.

In order to compute power (and hence type I error rate) for the study design, the joint distribution of the two stage test statistics is needed. Since critical values and denominator df depend on sample sizes, we derive the joint CDF conditional on an N_+ value, or $F_{F_1, F_+|N_+}(f_1, f_+)$. By using the law of total probability and summing over possible values, this result leads to calculation of unconditional power. *The following theorem gives a convenient expression for the conditional joint CDF of the test statistics for $a = 1$.*

Theorem 1

Here $b(t_e, t_h) = t_h^2 / (\sigma^2 c_+) - t_e$, $d(t_p) = n_+ t_p^2 / (n_2 \sigma^2)$, $h(t_e, t_h) = \sigma(c_1 t_e)^{1/2} - (n_1/n_+)^{1/2} t_h$, $g(t_e, t_h) = -\sigma(c_1 t_e)^{1/2} - (n_1/n_+)^{1/2} t_h$, and $p_{n_+} = F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)$. For $a = 1$,

$$F_{F_1, F_+|N_+}(f_1, f_+) = p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \int_{g(t_e, t_h)}^{h(t_e, t_h)} f_N[t_p; 0, (n_2/n_+) \sigma^2] \times \{1 - F_{\chi^2}[b(t_e, t_h) - d(t_p); n_2 - 1]\} dt_p dt_h dt_e. \quad (4)$$

A result needed with early futility stopping comes directly from Theorem 1 as follows:

$$\Pr(f_L \leq F_1 \leq f_U, F_+ \leq f_+ | N_+ = n_+) = F_{F_1, F_+|N_+}(f_U, f_+) - F_{F_1, F_+|N_+}(f_L, f_+). \quad (5)$$

A special case distribution is the conditional joint distribution of the test statistics when $n_2 = 1$, as given in the following theorem. This result is needed for single group hypothesis tests. Related results and definitions can be found in the appendix.

Theorem 2

For $a = 1$ and $n_2 = 1$, define $b(t_e) = n_+ c_1 t_e$ and $p_{n_+}^{-1} = F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)$, then

$$\begin{aligned}
F_{F_1, F_+ | N_+}(f_1, f_+) = & p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{-\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) \times \left(\Phi \left\{ \min \left[h(t_e, t_h), \sqrt{b(t_e, t_h)} \right] / \sigma \right\} - \Phi \left\{ \max \left[g(t_e, t_h), -\sqrt{b(t_e, t_h)} \right] / \sigma \right\} \right) dt \\
& + p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\sqrt{\sigma^2 t_e c_+}}^{\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) \{ \Phi[h(t_e, t_h)/\sigma] - \Phi[g(t_e, t_h)/\sigma] \} dt_h dt_e \\
& + p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{\sqrt{\sigma^2 t_e c_+}}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \times \left(\Phi \left\{ \min \left[h(t_e, t_h), \sqrt{b(t_e, t_h)} \right] / \sigma \right\} - \Phi \left\{ \max \left[g(t_e, t_h), -\sqrt{b(t_e, t_h)} \right] / \sigma \right\} \right) dt_h dt_e.
\end{aligned}
\tag{6}$$

Another distribution important to power and expected size calculations is the CDF of the first test statistic, conditional on N_+ . This result is in Coffey and Muller (1999) for internal pilots with $N_+ = n_1$. In the IPIA model it has more importance because of early stopping. It is used in subsequent results for power and expected sample size.

Theorem 3

The conditional CDF of the first test statistic can be written

$$F_{F_1 | N_+}(f_1) = \int_{q_1}^{q_2} \frac{F_{\chi^2}(c_1 t_1; a, \lambda_1) f_{\chi^2}(t_1; \nu_1)}{F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)} dt_1. \tag{7}$$

The following corollary adapts Theorem 3 to solve for the probability of the test continuing to the second stage conditional on $N_+ = n_+$ when futility stopping is possible.

Corollary 1

$$\Pr(f_L \leq F_1 < f_U | N_+ = n_+) = \int_{q_1}^{q_2} \frac{[F_{\chi^2}(c_U t_1; a, \lambda_1) - F_{\chi^2}(c_L t_1; a, \lambda_1)] f_{\chi^2}(t_1; \nu_1)}{F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)} dt_1. \tag{8}$$

The results of the following theorems are exact expressions for study power, type I error rate (power under the null hypothesis), and expected sample size. *The following theorem gives the formula for unconditional power.*

Theorem 4

An expression for unconditional power, P_w , can be written

$$P_w = 1 - \sum_{\{N_+ = n_+\}} [F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)] \left\{ F_{F_1 | N_+}[f_L(n_+)] + \Pr[f_L(n_+) \leq F_1 < f_U(n_+), F_+ < f_+(n_+) | N_+ = n_+] \right\}. \tag{9}$$

The results in this section can also be applied to calculate expected sample size for an IPIA study design.

Theorem 5

The expected value of N_w , the random total sample size taken in study, can be written

$$E(N_w) = n_1 + \sum_{\{N_+ = n_+\}} n_2 [F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)] \Pr(f_L \leq F_1 < f_U | N_+ = n_+) \quad (10)$$

5. Example

5.1 Motivation for Example

We illustrate our methods with a study designed to study the effects of a gluten-free, casein-free (GFCF) diet in children with autism (Elder et al., 2006). The study was a randomized, double-blind crossover design exploring the effects on autistic symptoms of the GFCF diet with a placebo diet. The primary analysis used a two group t -test to compare Childhood Autism Rating Scale (CARS) between the diets. Shuster (2007) detailed the advantages of the two group approach employed. With group membership defined by the *order* of diets, the analysis compared the groups on a difference score: second diet CARS score minus first diet CARS score (irrespective of order). Unlike most cross-over studies, this study did not require a wash-out period per the authors' dietary consultants. The sample of 15 participants was expected to be too small to reliably detect a minimal clinically important difference. The Principal Investigator deems this difference to be five points on the CARS. The study was undertaken and published to demonstrate how to conduct a controlled clinical trial of the GFCF diet in children with autism.

In planning a study, we seek the sample size required to ensure a target power (P_t) for a specified effect of interest (θ_1) at a given significance level (α_t). The true sample size required to meet the conditions depends on the unknown variance (σ^2), a nuisance parameter. With $\alpha_t = 0.05$, $P_t = 0.80$, $\theta_1 = 5$ (scalar effect of interest), and using $\sigma_0^2 = 50$ from estimates in the 2006 article, a fixed sample power calculation suggests $n_0 = 66$ children with autism are required. The difficulty and expense of obtaining and researching qualified children leads to a desire to use as few subjects as possible. Hence, we consider an IPIA analysis with the following specifications: 1) a preplanned sample size of $n_0 = 66$ children, and 2) the first $n_1 = 34$ children comprise the internal pilot sample.

The example has two main purposes. First, running the enumerations using both exact theory and simulations compares speed and gives an additional check on accuracy of the theory. Second, the results describe properties of the IPIA procedure compared to a fixed sample design, a special case within the general IPIA framework that does not allow for early stopping or sample size re-estimation ($f_L = 0, f_U = \infty$, and $\Pr(N_+ = n_0) = 1$). To facilitate the comparison, we use a naive, but common approach to critical value selection and study design.

The IPIA procedure is considered both with and without early stopping at the interim stage for futility. No futility stopping implies that the lower first stage critical value, f_L , is 0. The benefit of including futility stopping is that studies unlikely to result in positive outcomes are more often stopped early saving time, resources and patient intervention exposure. Hence, studies with futility stopping rules should have noticeable expected sample size reductions when true effect sizes are low. For futility stopping, we will use a simple p-value cut point of 0.85. In reality, this is not an ideal approach since the first stage may contain only a small fraction of the needed information of the study (especially for high true variance values) and so may not be very informative in some cases. It is used here for simplicity in order to portray characteristics of the procedures.

Critical values for both the fixed sample and IPIA designs will be based on the Gaussian distribution. O'Brien-Fleming stopping rule bounds determined by the design and sample fraction at the interim analysis will be used for IPIA design. The bounds are designed to allow for conservative early stopping while adjusting the final critical value for type I error rate inflation caused by multiple testing. More liberal early stopping rules are available and could have more effect on expected sample size properties. For example, the stopping family proposed by Wang and Tsatis (1987) has a continuous parameter that allows for stopping rules between the conservative O'Brien-Fleming bounds and the liberal Pocock bounds (Pocock, 1977). Examination of various stopping bounds during study planning should be undertaken and would lead to study properties better aligning with study goals.

Using unadjusted final critical values may cause type I error rate inflation in some cases. This is caused by misalignment of test statistic and critical value distributions and a biased variance value used in sample size re-estimation designs. They are used here in order to best compare the procedures to common practice and also to notice the magnitudes of such inflations. Strategies that better control the type I error rate in small samples and maintain power while minimizing expected sample size will be considered in future papers.

In total three design procedures were considered: fixed sample, IPIA with futility stopping, and IPIA without futility stopping. Type I error rate, power, and expected sample size were calculated for each procedure for true variance values determined by $\gamma \in \{0.5, 1, 2\}$ where $\gamma = \sigma^2 / \sigma_0^2$. We used the sampling fraction $\pi = 0.5$ and $n_{+, \max} = \infty$. Values were calculated under three conditions for the true effect size: the null ($\theta = 0$), the alternative of interest ($\theta = \theta_1$), and a true effect twice the effect size of interest ($\theta = 2\theta_1$).

5.2 Computational Methods

All programs were written in SAS/IML (SAS Institute, 2004). All IPIA computations for the examples used the exact theory developed in this paper. Stopping bound computation used the SEQSCALE function and the numeric integrations used the QUAD function within SAS/IML. To avoid numerical instability of the calculated integrals, computation was performed using quantile transformations (Glueck and Muller, 2001) of the distributions derived in Section 4.

For illustration, let $p = F_{\chi^2}(t; u)$. Then the integral transformation using $t = F_{\chi^2}^{-1}(p, v)$ and $dp = f_{\chi^2}(t; v)dt$ allows for integration using finite bounds with better behaved integrands.

Simulations were conducted over a range of conditions in order to check speed and the accuracy of the derivations, programming, and numerical algorithms. One million replications were sampled for each case. All programs were run with a Intel Xeon 3.2 GHz processor. For all cases, the simulated values were within two standard deviations of the analytically calculated values. The comparisons were run in groups of three cases corresponding to variance values of $\gamma \in \{0.5, 1, 2\}$ with $\gamma = \sigma^2 / \sigma_0^2$. Runs were made under the null hypothesis ($\theta = 0$) and assuming the effect of interest ($\theta = \theta_1$) for the IPIA design without early stopping for futility. Explicit timing results are detailed in Table 4.

For all cases, the analytic calculations using the exact theory were much faster than the simulations. While little effort has been made thus far to improve the computational speed of the programs, the simulations took between 30–50 times more computation time. Switching from the null to the alternative case took longer for the analytical calculations but made relatively little difference in the simulations.

5.3 Example Results

Table 5 displays the values for unconditional power and expected sample size for the fixed sample and IPIA designs. Power was calculated under the effect size assumptions of $\theta = 0$ (type I error rate), $\theta = \theta_1$ (power at effect size of interest), and $\theta = 2\theta_1$ (twice the effect size of interest). In order to see the effect of variance misspecification, values were calculated for variances corresponding to $\gamma = \sigma^2 / \sigma_0^2$ with $\gamma \in \{0.5, 1, 2\}$.

For the fixed sample design, type I error rate is somewhat inflated by a constant amount across γ . The IPIA designs have moderate type I error rate inflation caused by both variance estimate bias and the use of Gaussian critical values. Allowing for early stopping for futility causes a very small reduction in the type I error rate. With the target power at 0.8, the fixed sample example can be significantly over or under powered at $\theta = \theta_1$ depending on $\gamma = \sigma^2 / \sigma_0^2$. In contrast, the IPIA designs greatly achieve the target power as a result of using the variance estimate from the first stage for sample size re-estimation. The lowest power value for the IPIA without futility is 79.4% when the true variance is twice the preplanned value. This is 26.3% greater than the fixed sample power under the same conditions. The IPIA design with futility exhibits a small reduction in power compared to the IPIA design without futility, but is still considerably higher at its lowest point than the power for the fixed sample design (0.77 vs. 0.53). As expected, all designs examined had very high (at least 98%) power for a large effect size ($\theta = 2\theta_1$).

For all conditions, the fixed sample design uses the preplanned sample size, 66. The IPIA design gives lower expected samples sizes than the fixed design for low γ , and increases expected sample size for high γ . Under the alternative of interest, ($\theta = \theta_1$), the IPIA design has expected sample size noticeably lower than the fixed sample design whenever $\gamma < 1$. The expected sample size appropriately rises with $\gamma > 1$ owing to the information derived from the variance estimate at the first stage. Here early stopping (group sequential-like) benefits to expected sample size are offset by the cost of sample size recalculation, which protects study power (a characteristic of internal pilots).

Under the null hypothesis, the addition of possible futility stopping at the first stage to the IPIA design gives an across-the-board reduction in expected sample size attributable to the probability of correctly stopping early for no effect. Under the alternative of interest, allowing futility stopping decreases expected sample size since there is a nonzero probability of incorrectly stopping at the first stage. The chance of the event increases with γ owing to the futility-stopping rules used to calculate the first stage futility critical value. Futility stopping makes little difference when $\theta = 2\theta_1$ since the probability of futility stopping for the large effect size is very small.

6. Discussion

A single framework, the IPIA design, encompasses a variety of useful features of adaptive designs. For univariate Gaussian linear model hypotheses with fixed predictors in small and large samples, the theory and practice of group sequential, internal pilot, and fixed sample designs occur as special cases. We have derived the exact theory needed for the two stage IPIA design for single degree of freedom GLM hypotheses, including one and two group t -tests with unknown, common variances. The exact results allow numerical calculation of type I error rate, power, and expected sample size for many useful study designs without the need for time-consuming simulations.

For two stage designs, the techniques detailed in this paper are closely related to information-based monitoring described by Mehta and Tsiatis (2001) and Tsiatis (2006). Information-based

monitoring uses large sample theory to adjust maximum sample size adaptively within a group sequential design framework. Although very useful and general with studies large enough for asymptotic properties to hold, the breakdown point is not generally known. As highlighted in the NIH Roadmap, many prospective research studies and early phase clinical trials are not large enough for asymptotic properties to hold. The exact theory presented here could be used during study planning to assess type I error rate, power, and expected sample size properties in an extremely accurate and time-efficient manner, thus eliminating the need to make a subjective judgement as to whether or not one has a 'large' sample size. As seen in Table 4, computation with the exact theory used up to 50 times less computation time than the comparable simulations.

The example enumerated general IPIA properties and compared them to a fixed sample design. The fixed sample design has a known sample size, but has a power function affected by the unknown variance. The IPIA design seeks to combine the advantage in expected sample size of the group sequential design with the advantage in power protection of internal pilots. Although the given examples use only conservative O'Brien-Fleming stopping bounds, a larger reduction in sample size could be achieved by using more liberal bounds such as those contained in the stopping bound family described by Wang and Tsiatis (1987). Results in Table 5 allow concluding that the IPIA design does in fact greatly achieve its goals of protection of study power.

The IPIA procedure as implemented in this manuscript exhibits type I error rate inflation, with more inflation in small studies where exact theory has the most value. The results indicate that future work should take advantage of small sample (t vs z) critical values as part of eliminating type I error rate inflation in finite samples. Additionally, careful adjustments that control type I errors while maintaining the IPIA benefits must be developed in order for the IPIA procedure to be useful in practice for small samples, and perhaps even moderate samples.

Additionally, the examples contained here only use very simple futility stopping rules based on p -values. Futility stopping can save time, resources, and limit patient exposure during a study with little chance of success. If developed correctly, the benefits can be achieved with little cost to study power. Futility stopping decisions for clinical trials often do not follow strict rules, and mandatory futility stopping bounds are difficult to adhere to. However, if for planning purposes study properties of early futility stopping (such as potential sample size benefits) are to be examined, then at least approximate rules need to be assumed. If study planners have an idea about stopping thresholds, then more accurate planning information can be obtained.

The IPIA procedure was purposefully kept general in many regards. For example, it does not specify methods for selecting critical values, updating sample size, or selecting the interim stage sample size. The theory and software developed so far will make exploring many different methods and designs convenient within the two stage IPIA framework. Such work as well as expansion of the theory into more than two testing stages will be goals of future research.

References

- Arghami NR, Billard L. Some sequential analogs of Stein's 2-sample test. *Theory of Probability and its Applications* 1992;37:115–117.
- Baker AG. Properties of some tests in sequential analysis. *Biometrika* 1950;37:334–346. [PubMed: 14801060]
- Birkett MA, Day SJ. Internal pilot-studies for estimating sample-size. *Statistics in Medicine* 1994;13:2455–2463. [PubMed: 7701146]
- Coffey CS, Muller KE. Controlling test size while gaining the benefits of an internal pilot design. *Biometrics* 2001;57:625–631. [PubMed: 11414593]

- Coffey CS, Muller KE. Some distributions and their implications for an internal pilot study with a univariate linear model. *Communications in Statistics-Theory and Methods* 2000;29:2677–2691.
- Coffey CS, Muller KE. Exact test size and power of a Gaussian error linear model for an internal pilot study. *Statistics in Medicine* 1999;18:1199–1214. [PubMed: 10363340]
- Denne JS, Jennison C. A group sequential t-test with updating of sample size. *Biometrika* 2000;87:125–134.
- Denne JS, Jennison C. Estimating the sample size for a t-test using an internal pilot. *Statistics in Medicine* 1999;18:1575–1585. [PubMed: 10407230]
- Elder JH, Shankar M, Shuster J, Theriaque D, Burns S, Sherill L. The gluten-free, casein-free diet in autism: results of a preliminary double blind clinical trial. *Journal of Autism and Developmental Disorders* 2006;36:413–420. [PubMed: 16555138]
- Facey KM. A sequential procedure for a phase-ii efficacy trial in hypercholesterolemia. *Controlled Clinical Trials* 1992;13:122–133. [PubMed: 1316827]
- Glueck DH, Muller KE. On the expected values of sequences of functions. *Communications in Statistics-Theory and Methods* 2001;30:363–369.
- Gould AL, Shih WJ. Modifying the design of ongoing trials without unblinding. *Statistics in Medicine* 1998;17:89–100. [PubMed: 9463852]
- Hall WJ. Some sequential analogs of steins 2-stage test. *Biometrika* 1962;49:367–378.
- Hochberg Y, Marcus R. 2-phase tests on a normal-mean when variance is unknown. *Journal of Statistical Planning and Inference* 1983;7:233–242.
- Jennison, C.; Turnbull, BW. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton: Chapman & Hall/CRC; 2000.
- Johnson, NL.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions-1*. 2. New York: Wiley; 1994.
- Johnson, NL.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions-2*. 2. New York: Wiley; 1995.
- Mehta CR, Tsiatis AA. Flexible sample size considerations using information-based interim monitoring. *Drug Information Journal* 2001;35:1095–1112.
- Morgan CC. Sample size re-estimation in group-sequential response-adaptive clinical trials. *Statistics in Medicine* 2003;22:3843–3857. [PubMed: 14673942]
- Muller, KE.; Stewart, P. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. New York: Wiley; 2006.
- O'Brien PC, Fleming TR. Multiple testing procedure for clinical-trials. *Biometrics* 1979;35:549–556. [PubMed: 497341]
- Pocock SJ. Group sequential methods in design and analysis of clinical-trials. *Biometrika* 1977;64:191–200.
- SAS Institute. *SAS/IML® 9.1 User's Guide, Volumes 1 and 2*. Cary, NC: SAS Institute; 2004.
- Shuster, JJ. Design and analysis of experiments. In: Ambrosius, WT., editor. *Topics in Biostatistics*. Vol. Chapter 12. Humana Press; Totawa, NJ: 2007. p. 235-260.
- Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* 1945;16:43–58.
- Tsiatis AA. Information-based monitoring of clinical trials. *Statistics in Medicine* 2006;25:3236–3244. [PubMed: 16927248]
- Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* 1987;43:193–200. [PubMed: 3567304]
- Whitehead J, Stratton I. Group sequential clinical-trials with triangular continuation regions. *Biometrics* 1983;39:227–236. [PubMed: 6871351]
- Whitehead J, Whitehead A, Todd S, Bolland K, Sooriyarachchi MR. Mid-trial design reviews for sequential clinical trials. *Statistics in Medicine* 2001;20:165–176. [PubMed: 11169595]
- Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* 1990;9:65–72. [PubMed: 2345839]

Appendix

Lemma A.1

Any testable general linear hypothesis with $\theta_0 \neq \mathbf{0}$ may be expressed in terms of a related model and general linear hypothesis with $\theta_0 = \mathbf{0}$.

Proof

For $\beta_* = \beta - C'(CC')^{-1}\theta_0$, $C\beta_* = C\beta - CC'(CC')^{-1}\theta_0 = C\beta - \theta_0$ which implies $H_0: C\beta = \theta_0 \Leftrightarrow H_0: C\beta_- = \mathbf{0}$.

Lemma A.2

For the model in equation 2 interpreted as a fixed n_+ design, the following $n_+ \times n_+$ matrices are symmetric and idempotent, for any testable hypothesis, and have ranks of a , $(n_+ - r)$, $(n_1 - r)$ and n_2 :

$$\begin{aligned} A_{h+} &= X_+(X'_+X_+)^{-1}C' \left[C(X'_+X_+)^{-1}C'^{-1} \right] C(X'_+X_+)^{-1}X'_+ \\ A_{e+} &= I_{n_+} - X_+(X'_+X_+)^{-1}X'_+ \\ A_{e1} &= \begin{bmatrix} I_{n_1} - X_1(X'_1X_1)^{-1}X'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n_2 \times n_2} \end{bmatrix} \\ A_{ep} &= A_{e+} - A_{e1}. \end{aligned} \quad (11)$$

Furthermore, with detailed calculations given by Coffey and Muller (1999), $A_{h+}A_{e+} = A_{h+}A_{e1} = A_{h+}A_{ep} = A_{e1}A_{ep} = \mathbf{0}$. Extending the notation gives

$$A_{h1} = \begin{bmatrix} X_1(X'_1X_1)^{-1}C' [C(X'_1X_1)^{-1}C']^{-1} C(X'_1X_1)^{-1}X'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n_+ \times n_+}. \quad (12)$$

Equivalently if $X_{1*} = \begin{bmatrix} X_1 \\ \mathbf{0}_{n_2 \times q} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{k_1} \otimes X_0 \\ \mathbf{0}_{n_2 \times q} \end{bmatrix}$ then

$$A_{h1} = X_{1*}(X'_{1*}X_{1*})^{-1}C' \left[C(X'_{1*}X_{1*})^{-1}C' \right]^{-1} C(X'_{1*}X_{1*})^{-1}X'_{1*}. \quad (13)$$

In turn, A_{hs} , for $s \in \{1, +\}$, is idempotent, symmetric, and rank a (testable hypothesis). Hence one can define V_{hs} of dimension $n_+ \times a$ with $A_{hs} = V_{hs}V'_{hs}$ and $V'_{hs}V_{hs} = I_a$. Also, since $A_{ep} = A_{e+} - A_{e1}$ is idempotent, symmetric, and rank n_2 , the matrix V_{ep} of dimension $n_+ \times n_2$ exists such that $A_{ep} = V_{ep}V'_{ep}$ and $V'_{ep}V_{ep} = I_{n_2}$.

Since it is symmetric and rank a , $M_0 = C(X'_0X_0)^{-1}C'$ can be written as $F_0F'_0$ for F_0 of dimension $a \times a$ and rank a . This in turn implies that $M_0^{-1} = F_0^{-1}F_0'^{-1}$. For $k_s = n_s/m$ the number

of replications at stage s , $M_s = k_s^{-1} M_0$ implies that $M_s^{-1} = k_s F_0^{-t} F_0^{-1}$. Also $X_s = 1_{k_s} \otimes X_0$ implies that $X_s' X_s = k_s X_0' X_0$.

The matrices V_{h+} and V_{h1} can now be derived. The result that

$$\begin{aligned} A_{h+} &= X_+(X_+'X_+)^{-t} C' M_+^{-1} C (X_+'X_+)^{-1} X_+' \\ &= X_+(k_+ X_0' X_0)^{-t} C' M_+^{-1} C (k_+ X_0' X_0)^{-1} X_+' \\ &= k_+^{-2} k_+ X_+(X_0' X_0)^{-t} C' F_0^{-1} F_0^{-1} C (X_0' X_0)^{-1} X_+' \\ &= k_+^{-1} \left[X_+(X_0' X_0)^{-1} C' F_0^{-1} \right] \left[F_0^{-1} C (X_0' X_0)^{-1} X_+' \right] \end{aligned} \quad (14)$$

implies $V_{h+} = k_+^{-1/2} X_+(X_0' X_0)^{-1} C' F_0^{-t}$. Similarly,

$A_{h1} = k_1^{-1} [X_{1*}(X_0' X_0)^{-1} C' F_0^{-t}] [F_0^{-1} C (X_0' X_0)^{-1} X_{1*}']$ implies

$V_{h1} = k_1^{-1/2} X_{1*}(X_0' X_0)^{-1} C' F_0^{-t}$. Using the result

$X_+' X_{1*} = \begin{bmatrix} X_1' & X_2' \end{bmatrix} \begin{bmatrix} X_1 \\ \mathbf{0} \end{bmatrix} = X_1' X_1 = k_1 X_0' X_0$, the following can be proven:

$$\begin{aligned} V_{h+}' V_{h1} &= (k_+ k_1)^{-1/2} \left[F_0^{-1} C (X_0' X_0)^{-1} X_+' \right] \left[X_{1*}(X_0' X_0)^{-1} C' F_0^{-t} \right] \\ &= (n_1/n_+)^{1/2} F_0^{-1} F_0 F_0' F_0^{-t} \\ &= (n_1/n_+)^{1/2} I_a \end{aligned} \quad (15)$$

Also, directly from symmetry, $V_{h1}' V_{h+} = V_{h+}' V_{h1} = (n_1/n_+)^{1/2} I_a$.

Since $A_{h+} A_{ep} = V_{h+} V_{h+}' V_{ep} V_{ep}' = \mathbf{0}$, the following are true: $V_{h+}' V_{ep} = \mathbf{0}$, and $V_{ep}' V_{h+} = \mathbf{0}$. Similarly,

since $A_{ep} A_{e1} = V_{ep} V_{ep}' V_{e1} V_{e1}' = \mathbf{0}$, the following are true: $V_{ep}' V_{e1} = \mathbf{0}$ and $V_{e1}' V_{ep} = \mathbf{0}$. Also,

$V_{ep}' = V_{ep}' A_{ep} = V_{ep}' (A_{e+} - A_{e1}) = V_{ep}' A_{e+}$ which implies that

$V_{ep}' \mu_+ = V_{ep}' A_{e+} \mu_+ = V_{ep}' (X_+ - H_+ X_+) \beta = \mathbf{0}$. Also,

$$\begin{aligned} & (I_{n_+} - A_{ep} - A_{h+}) V_{h1} \\ &= (H_+ - A_{h+}) V_{h1} \\ &= k_+^{-1} X_+(X_0' X_0)^{-1} X_+' V_{h1} - (k_1/k_+)^{1/2} V_{h+} \\ &= k_+^{-1} k_1^{1/2} X_+(X_0' X_0)^{-1} C' F_0^{-t} - k_+^{-1} k_1^{1/2} X_+(X_0' X_0)^{-1} C' F_0^{-t} \\ &= \mathbf{0}. \end{aligned} \quad (16)$$

The following definitions are useful:

$$\begin{aligned} y_{h1} &= V_{h1}' y_+ & a \times n_+ \times 1 \\ y_{ep} &= V_{ep}' y_+ & n_2 \times n_+ \times 1 \\ y_{h+} &= V_{h+}' y_+ & a \times n_+ \times 1. \end{aligned} \quad (17)$$

With

$$\begin{bmatrix} V_{h1}' \\ V_{ep}' \\ V_{h+}' \end{bmatrix} I_{n_+} \begin{bmatrix} V_{h1} & V_{ep} & V_{h+} \end{bmatrix} = \begin{bmatrix} I_a & V_{h1}' V_{ep} & V_{h1}' V_{h+} \\ V_{ep}' V_{h1} & I_{n_2} & V_{ep}' V_{h+} \\ V_{h+}' V_{h1} & V_{h+}' V_{ep} & I_a \end{bmatrix} = \begin{bmatrix} I_a & V_{h1}' V_{ep} & (n_1/n_+)^{1/2} I_a \\ V_{ep}' V_{h1} & I_{n_2} & \mathbf{0} \\ (n_1/n_+)^{1/2} I_a & \mathbf{0} & I_a \end{bmatrix}_{(2a+n_2)} \quad (18)$$

and

$$\begin{bmatrix} V_{h1}' \mu_+ \\ V_{ep}' \mu_+ \\ V_{h+}' \mu_+ \end{bmatrix} = \begin{bmatrix} \mu_{h1} \\ \mathbf{0} \\ \mu_{h+} \end{bmatrix}_{(2a+n_2) \times 1}, \quad (19)$$

the following vector can be defined:

$$\mathbf{y}_h = \begin{bmatrix} V_{h1}' \\ V_{ep}' \\ V_{h+}' \end{bmatrix} \mathbf{y}_+ = \begin{bmatrix} \mathbf{y}_{h1} \\ \mathbf{y}_{ep} \\ \mathbf{y}_{h+} \end{bmatrix}. \quad (20)$$

Furthermore

$$\mathbf{y}_h \sim (S) \mathcal{N}_{2a+n_2} \left(\begin{bmatrix} \mu_{h1} \\ \mathbf{0} \\ \mu_{h+} \end{bmatrix}, \sigma^2 \begin{bmatrix} I_a & V_{h1}' V_{ep} & (n_1/n_+)^{1/2} I_a \\ V_{ep}' V_{h1} & I_{n_2} & \mathbf{0} \\ (n_1/n_+)^{1/2} I_a & \mathbf{0} & I_a \end{bmatrix} \right). \quad (21)$$

Temporarily defining

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} I_a & [V_{h1}' V_{ep} \ (n_1/n_+)^{1/2} I_a] \\ \begin{bmatrix} V_{ep}' V_{h1} \\ (n_1/n_+)^{1/2} I_a \end{bmatrix} & \begin{bmatrix} I_{n_2} & \mathbf{0} \\ \mathbf{0} & I_a \end{bmatrix} \end{bmatrix}, \quad (22)$$

using the properties in equation 15 one can write the relation:

$$\begin{aligned} \mathbf{y}_{h1} &= V_{h1}' (A_{ep} + A_{h+}) \mathbf{y}_+ + V_{h1}' (I_{n_+} - A_{ep} - A_{h+}) \mathbf{y}_+ \\ &= V_{h1}' A_{ep} \mathbf{y}_+ + V_{h1}' A_{h+} \mathbf{y}_+ \\ &= V_{h1}' V_{ep} \mathbf{y}_{ep} + (n_1/n_+)^{1/2} \mathbf{y}_{h+} \end{aligned} \quad (23)$$

with $\mathbf{y}_{ep}$ independent of $\mathbf{y}_{h+}$ and both distributed as described in equation 21. That is, the relation in equation 23 holds for independent $\mathbf{y}_{ep}$ and $\mathbf{y}_{h+}$ and $\mathbf{y}_{ep} \sim \mathcal{N}_{n_2}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_2})$ and $\mathbf{y}_{h+} \sim \mathcal{N}_s(\boldsymbol{\mu}_{h+}, \sigma^2 \mathbf{I}_a)$. Additionally,

$$\begin{aligned} \mu_{h+} &= \mathbf{V}_{h+}' \boldsymbol{\mu}_+ \\ &= k_+^{-1/2} \left[\mathbf{F}_0^{-1} \mathbf{C} (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_+' \right] \mathbf{X}_+ \boldsymbol{\beta} \\ &= (n_+/m)^{1/2} \mathbf{F}_0^{-1} \boldsymbol{\theta} \end{aligned} \quad (24)$$

and similarly, $\mu_{h1} = (n_1/m)^{1/2} \mathbf{F}_0^{-1} \boldsymbol{\theta}$. These then imply the relation $\boldsymbol{\mu}_{h1} = (n_1/n_+)^{1/2} \boldsymbol{\mu}_{h+}$.

For $E_s = \nu_s \widehat{\sigma}_s^2 / \sigma^2$, and $E_p = E_+ - E_1$, a requirement for the main results of this paper is the distribution of E_p conditional on $\mathbf{V}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep}$ for the special case of $a = 1$. The following properties will help describe this distribution.

$$\begin{aligned} \mathbf{V}_{h1}' \mathbf{A}_{ep} \mathbf{V}_{h1} &= \mathbf{V}_{h1}' \mathbf{A}_{e+} \mathbf{V}_{h1} \\ &= \mathbf{I}_a - \mathbf{V}_{h1}' \mathbf{H}_+ \mathbf{V}_{h1} \\ &= \mathbf{I}_a - k_1^{-1} \left[\mathbf{F}_0^{-1} \mathbf{C} (\mathbf{X}_0' \mathbf{X}_0)^{-1} (\mathbf{X}_{1*}') \right] \mathbf{H}_+ \left[\mathbf{X}_{1*} (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{C}' \mathbf{F}_0^{-t} \right] \\ &= \mathbf{I}_a - (k_1/k_+) \mathbf{F}_0^{-1} \mathbf{F}_0 \mathbf{F}_0' \mathbf{F}_0^{-t} \\ &= (n_2/n_+) \mathbf{I}_a. \end{aligned} \quad (25)$$

For $a = 1$, this becomes $\mathbf{v}_{h1}' \mathbf{A}_{ep} \mathbf{v}_{h1} = n_2/n_+$. Using equations 21 and 24, the distribution of $\mathbf{V}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep}$ is shown to be $\mathbf{V}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep} \sim \mathcal{N}_a[\mathbf{0}, (n_2/n_+) \sigma^2 \mathbf{I}_a]$. Or, for $a = 1$, $\mathbf{v}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep} \sim \mathcal{N}[0, (n_2/n_+) \sigma^2]$.

For the $a = 1$ case, the joint distribution of $\mathbf{y}_{ep}$ and $\mathbf{V}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep}$ is

$$\begin{bmatrix} \mathbf{y}_{ep} \\ \mathbf{v}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n_2} \\ \mathbf{v}_{h1}' \mathbf{V}_{ep} \end{bmatrix} \mathbf{y}_{ep} \sim (S) \mathcal{N}_{n_2+1} \left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{I}_{n_2} & \mathbf{V}_{ep}' \mathbf{v}_{h1} \\ \mathbf{v}_{h1}' \mathbf{V}_{ep} & (n_2/n_+) \end{bmatrix} \right). \quad (26)$$

Conditional Gaussian theory (see Muller and Stewart, 2006; Chapter 8) then implies for $a = 1$, $\mu_{e,h} = (n_+/n_2) t_0 \mathbf{V}_{ep}' \mathbf{v}_{h1}$, and $\sum_{e,h} = \sigma^2 [\mathbf{I}_{n_2} - (n_+/n_2) \mathbf{V}_{ep}' \mathbf{A}_{h1} \mathbf{V}_{ep}]$, the distribution of $\mathbf{y}_{ep}$ conditional on $\mathbf{v}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep}$ is

$$\mathbf{y}_{e,h} = \mathbf{y}_{ep} | (\mathbf{v}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep} = t_0) \sim (S) \mathcal{N}_{n_2}(\mu_{e,h}, \sum_{e,h}). \quad (27)$$

For $E_p = \mathbf{y}_{ep}' \mathbf{y}_{ep} / \sigma^2$, we can now state in a corollary the distribution of E_p conditional on $\mathbf{V}_{h1}' \mathbf{V}_{ep} \mathbf{y}_{ep}$ for $a = 1$.

Corollary A.1

For $a = 1$ and $E_p = \mathbf{y}'_1(\mathbf{I}_{n_1} - \mathbf{H}_1)\mathbf{y}_1/\sigma^2$, the distribution of E_p conditional on $\mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} = t_0$ is

$$E_p | (\mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} = t_0) = X_p + (n_+ / n_2)(t_0 / \sigma)^2 \quad (28)$$

where X_p is a central χ^2 distributed variable with $n_2 - 1$ degrees of freedom.

Proof

For $\mu_{e,h} = (n_+ / n_2)t_0 \mathbf{V}'_{ep} \mathbf{v}_{h1}$ and

$\sum_{e,h} = \sigma^2 [\mathbf{I}_{n_2} - (n_+ / n_2) \mathbf{V}'_{ep} \mathbf{A}_{h1} \mathbf{V}_{ep}]$, $\mathbf{y}_{e,h} = \mathbf{y}_{ep} | (\mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} = t_0) \sim (S) \mathcal{N}_{n_2}(\mu_{e,h}, \sum_{e,h})$. So, since

$E_p = \mathbf{y}'_{ep} \mathbf{y}_{ep} / \sigma^2$, the variable $E_p | (\mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} = t_0)$ is distributed the same as

$q = \mathbf{y}'_{e,h} \mathbf{y}_{e,h} / \sigma^2 = \mathbf{y}'_{e,h} \mathbf{A} \mathbf{y}_{e,h}$ for $\mathbf{A} = (1/\sigma^2) \mathbf{I}_{n_2}$. Also, for $\mathbf{F}_A = \sigma^{-1} \mathbf{I}_{n_2}$, $\mathbf{A} = \mathbf{F}_A \mathbf{F}_A'$. For

$\mu_A = \mathbf{F}_A' \mu_{e,h} = \sigma^{-1} \mu_{e,h}$ and $\sum_A = \mathbf{F}_A' \sum_{e,h} \mathbf{F}_A = (1/\sigma^2) \sum_{e,h}$ define

$\mathbf{y}_A = \mathbf{F}_A' \mathbf{y}_{e,h} \sim (S) \mathcal{N}_{n_2}(\mu_A, \sum_A)$. Also

$$\begin{aligned} \sum_A &= \mathbf{I}_{n_2} - (n_+ / n_2) \mathbf{V}'_{ep} \mathbf{A}_{h1} \mathbf{V}_{ep} \\ &= \mathbf{I}_{n_2} - (n_+ / n_2) (\mathbf{V}'_{ep} \mathbf{v}_{h1}) (\mathbf{v}'_{h1} \mathbf{V}_{ep}) \\ &= \mathbf{I}_{n_2} - (n_+ / n_2) \left\{ \left[(\mathbf{V}'_{ep} \mathbf{v}_{h1}) (n_2 / n_+)^{-1/2} \right] (n_2 / n_+) \left[(n_2 / n_+)^{-1/2} \mathbf{v}'_{h1} \mathbf{V}_{ep} \right] \right\} \\ &= \mathbf{I}_{n_2} - \mathbf{v}_{x2} \mathbf{v}'_{x2}. \end{aligned} \quad (29)$$

Here $\mathbf{v}_{x1} \mathbf{v}'_{x1}$ has rank 1, with only nonzero eigenvalue $(\mathbf{v}'_{h1} \mathbf{V}_{ep}) (\mathbf{V}'_{ep} \mathbf{v}_{h1}) = n_2 / n_+$ because, in general, $\mathbf{X} \mathbf{X}'$ and $\mathbf{X}' \mathbf{X}$ have same nonzero eigenvalues and the eigenvalue of a scalar is itself.

Hence $\sum_A = \mathbf{I}_{n_2} - (n_+ / n_2) \mathbf{V}'_{ep} \mathbf{A}_{h1} \mathbf{V}_{ep}$ is $n_2 \times n_2$, rank $n_2 - 1$, and idempotent, so has $n_2 - 1$ eigenvalues of 1 and one of 0. Hence, $\sum_A$ has rank $n_2 - 1$.

By spectral decomposition one can write $\sum_A = \begin{bmatrix} \Psi & \mathbf{v}_{x2} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_2-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Psi' \\ \mathbf{v}'_{x2} \end{bmatrix}$. Also $\Psi^+ = (\Psi'' \Psi)^{-1} \Psi'' = \Psi''$ and $\begin{bmatrix} \Psi & \mathbf{v}_{x2} \end{bmatrix} \begin{bmatrix} \Psi' \\ \mathbf{v}'_{x2} \end{bmatrix} = \Psi \Psi' + \mathbf{v}_{x2} \mathbf{v}'_{x2} = \mathbf{I}_{n_2}$. Hence, $\mu_A = \Psi \Psi' \mu_A + \mathbf{v}_{x2} \mathbf{v}'_{x2} \mu_A$.

If $\mu_z = \Psi \mu_A$ and $\mu_0 = \mathbf{v}_{x2} \mathbf{v}'_{x2} \mu_A$, while $\mathbf{y}_0 = \mathbf{y}_A - \mu_A$ and $\mathbf{z} \sim \mathcal{N}_{n_2}(\mathbf{0}, \mathbf{I}_{n_2})$, then $\mathbf{y}_0 \sim (S)^{\mathcal{N}_{n_2}}(\mathbf{0}, \sum_A)$ and

$$\begin{aligned} \mathbf{y}_A &= \mathbf{y}_0 + \mu_A \\ &= \mathbf{y}_0 + \Psi \Psi' \mu_A + \mathbf{v}_{x2} \mathbf{v}'_{x2} \mu_A \\ &= \Psi(\mathbf{z} + \mu_z) + \mu_0. \end{aligned} \quad (30)$$

Now the following can be expressed:

$$\begin{aligned}
q &= \mathbf{y}'_A \mathbf{y}_A \\
&= [\Psi(\mathbf{z} + \mu_z) + \mu_0]' [\Psi(\mathbf{z} + \mu_z) + \mu_0] \\
&= (\mathbf{z} + \mu_z)' \Psi' \Psi (\mathbf{z} + \mu_z) + 2(\mathbf{z} + \mu_z)' \Psi' \mu_0 + \mu_0' \mu_0 \\
&= (\mathbf{z} + \mu_z)' (\mathbf{z} + \mu_z) + 2(\mathbf{z} + \mu_z)' \Psi' \mathbf{v}_{x2} \mathbf{v}'_{x2} \mu_A + \mu_0' \mu_0 \\
&= \sum_{k=1}^{n_2-1} (z_k + \mu_{z(k)})^2 + \mu_0' \mu_0.
\end{aligned} \tag{31}$$

Also, since $\mu_z = \Psi^+ \mu_A = \sigma^{-1} \Psi' \mathbf{v}_{x2} = \mathbf{0}$ and

$$\begin{aligned}
\mu_0' \mu_0 &= \mu_A' \mathbf{v}_{x2} \mathbf{v}'_{x2} \mathbf{v}_{x2} \mathbf{v}'_{x2} \mu_A \\
&= \mu_A' (I - \sum_A) \mu_A \\
&= \sigma^{-2} \mu_{e,h}' (I - \sigma^{-2} \sum_A) \mu_{e,h} \\
&= (n_+/n_2)^2 (t_0/\sigma)^2 \mathbf{V}'_{h1} \mathbf{V}_{ep} \{ (n_+/n_2) \mathbf{V}'_{ep} \mathbf{A}_{h1} \mathbf{V}_{ep} \} \mathbf{V}'_{ep} \mathbf{V}_{h1} \\
&= (n_+/n_2) (t_0/\sigma)^2,
\end{aligned} \tag{32}$$

one can write $q = \sum_{k=1}^{n_2-1} z_k^2 + (n_+/n_2) (t_0/\sigma)^2 = X_p + (n_+/n_2) (t_0/\sigma)^2$ where X_p is a central χ^2 distributed variable with $n_2 - 1$ degrees of freedom.

Theorem 1 Proof

Related definitions are above. Conditional on N_+ , $E_1 = \nu_1 \widehat{\sigma}_1^2 / \sigma^2 \sim \chi^2_T[\nu_1; q_1(n_+), q_2(n_+)]$ (Coffey & Muller, 1999). For $c_s = f_s/\nu_s$, $E_s = \nu_s \widehat{\sigma}_s^2 / \sigma^2$, $E_p = E_+ - E_1$, $G_s = \delta_s / \sigma^2$ with $s \in \{1, +\}$. Here $b(t_e, t_h) = t_e^2 / (\sigma^2 c_+) - t_e$, $d(t_p) = n_+ t_p^2 / (n_2 \sigma^2)$, $h(t_e, t_h) = \sigma(c_1 t_e)^{1/2} - (n_1/n_+)^{1/2} t_h$, $g(t_e, t_h) = -\sigma(c_1 t_e)^{1/2} - (n_1/n_+)^{1/2} t_h$, and $p_{n_+}^{-1} = F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)$.

$$\begin{aligned}
F_{F_1, F_+ | N_+}(f_1, f_+) &= \Pr_{N_+}(F_1 \leq f_1, F_+ \leq f_+) \\
&= \Pr_{N_+}(\widehat{\delta}_1 / \widehat{\sigma}_1^2 \leq f_1, \widehat{\delta}_+ / \widehat{\sigma}_+^2 \leq f_+) \\
&= \Pr_{N_+}\{G_1 \leq c_1 E_1, G_+ \leq c_+(E_1 + E_p)\} \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \times \Pr_{N_+} \left[\left| \mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} + \sqrt{\frac{n_1}{n_+}} y_{h+} \right| \leq \sigma \sqrt{c_1 t_e}, y_{h+}^2 \leq \sigma^2 c_+(t_e + E_p) \right] dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \times \Pr_{N_+}[g(t_e, t_h) \leq \mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} \leq h(t_e, t_h), b(t_e, t_h) \leq E_p] dt_h dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \int_{g(t_e, t_h)}^{h(t_e, t_h)} f_N[t_p; 0, (n_2/n_+) \sigma^2] \times \Pr_{N_+} \left[b(t_e, t_h) \leq E_p | \mathbf{v}'_{h1} \mathbf{V}_{ep} \mathbf{y}_{ep} \right] dt_p dt_h dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \int_{g(t_e, t_h)}^{h(t_e, t_h)} f_N[t_p; 0, (n_2/n_+) \sigma^2] \times \Pr_{N_+} [b(t_e, t_h) - d(t_p) \leq X_p] dt_p dt_h dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; \nu_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \int_{g(t_e, t_h)}^{h(t_e, t_h)} f_N[t_p; 0, (n_2/n_+) \sigma^2] \times \{1 - F_{\chi^2}[b(t_e, t_h) - d(t_p); n_2 - 1]\} dt_p dt_h dt_e
\end{aligned}$$

Theorem 2 Proof

Related definitions are above. For $n_2 = 1$, $c_s = f_s/\nu_s$, $E_s = \nu_s \widehat{\sigma}_s^2 / \sigma^2$, $E_p = E_+ - E_1$, $G_s = \delta_s / \sigma^2$ for $s \in \{1, +\}$,

$b(t_e, t_h) = t_h^2/c_+ - \sigma^2 t_e$, $g(t_e, t_h) = n_1^{1/2} t_h - \sigma(n_+ c_1 t_e)^{1/2}$, $h(t_e, t_h) = n_1^{1/2} t_h + \sigma(n_+ c_1 t_e)^{1/2}$, $v_{1p} = v_{h1}' v_{ep} = -(n_2/n_+)^{1/2}$, and $p_{n_+}^{-1} = F_{\chi^2}(q_2; v_1) - F_{\chi^2}(q_1; v_1)$.

$$\begin{aligned}
F_{F_1, F_+} \Big|_{N_+} (f_1, f_+) &= \Pr_{N_+} (F_1 \leq f_1, F_+ \leq f_+) \\
&= \Pr_{N_+} \left(\widehat{\delta}_1 / \widehat{\sigma}_1^2 \leq f_1, \widehat{\delta}_+ / \widehat{\sigma}_+^2 \leq f_+ \right) \\
&= \Pr_{N_+} \{G_1 \leq c_1 E_1, G_+ \leq c_+ (E_1 + E_p)\} \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \times \Pr_{N_+} [G_1 \leq c_1 t_e, G_+ \leq c_+ (t_e + E_p)] dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \Pr_{N_+} \left[\left| v_{1p} y_{ep} + \sqrt{\frac{n_1}{n_+}} y_{h+} \right| \leq \sigma \sqrt{c_1 t_e}, y_{h+}^2 \leq \sigma^2 c_+ (t_e + E_p) \right] dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{-\infty}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \Pr_{N_+} [g(t_e, t_h) \leq y_{ep} \leq h(t_e, t_h), b(t_e, t_h) \leq y_{ep}^2] dt_h dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{-\infty}^{-\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) \times \\
&\quad \left\{ \Pr_{N_+} [g(t_e, t_h) \leq y_{ep} \leq -\sqrt{b(t_e, t_h)}] + \Pr_{N_+} [\sqrt{b(t_e, t_h)} \leq y_{ep} \leq h(t_e, t_h)] \right\} dt_h dt_e + \\
&\quad p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{-\sqrt{\sigma^2 t_e c_+}}^{\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) \Pr_{N_+} [g(t_e, t_h) \leq y_{ep} \leq h(t_e, t_h)] dt_h dt_e + \\
&\quad p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{\sqrt{\sigma^2 t_e c_+}}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) \times \\
&\quad \left\{ \Pr_{N_+} [g(t_e, t_h) \leq y_{ep} \leq -\sqrt{b(t_e, t_h)}] + \Pr_{N_+} [\sqrt{b(t_e, t_h)} \leq y_{ep} \leq h(t_e, t_h)] \right\} dt_h dt_e \\
&= p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{-\infty}^{-\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) (\max \{ \Phi[-\sqrt{b(t_e, t_h)}/\sigma] - \Phi[g(t_e, t_h)/\sigma], 0 \} + \\
&\quad \max \{ \Phi[h(t_e, t_h)/\sigma] - \Phi[-\sqrt{b(t_e, t_h)}/\sigma], 0 \}) dt_h dt_e + \\
&\quad p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{-\sqrt{\sigma^2 t_e c_+}}^{\sqrt{\sigma^2 t_e c_+}} f_N(t_h; \mu_{h+}, \sigma^2) \{ \Phi[h(t_e, t_h)/\sigma] - \Phi[g(t_e, t_h)/\sigma] \} dt_h dt_e + \\
&\quad p_{n_+}^{-1} \int_{q_1}^{q_2} f_{\chi^2}(t_e; v_1) \int_{\sqrt{\sigma^2 t_e c_+}}^{\infty} f_N(t_h; \mu_{h+}, \sigma^2) (\max \{ \Phi[-\sqrt{b(t_e, t_h)}/\sigma] - \Phi[g(t_e, t_h)/\sigma], 0 \} + \\
&\quad \max \{ \Phi[h(t_e, t_h)/\sigma] - \Phi[\sqrt{b(t_e, t_h)}/\sigma], 0 \}) dt_h dt_e
\end{aligned}$$

Theorem 3 Proof

$$\begin{aligned}
F_{F_1} \Big|_{N_+} (f_1) &= \Pr(F_1 \leq f_1 | N_+ = n_+) \\
&= \Pr\{\widehat{\delta}_1 / (a \cdot \widehat{\sigma}_1^2) \leq f_1 | N_+ = n_+\} \\
&= \Pr(G_1 \leq c_1 E_1 | N_+ = n_+) \\
&= \int_{q_L}^{q_2} \frac{\Pr(G_1 \leq c_1 t_1) f_{\chi^2}(t_1; v_1)}{F_{\chi^2}(q_2; v_1) - F_{\chi^2}(q_1; v_1)} dt_1 \\
&= \int_{q_1}^{q_2} \frac{F_{\chi^2}(c_1 t_1; a, \lambda_1) f_{\chi^2}(t_1; v_1)}{F_{\chi^2}(q_2; v_1) - F_{\chi^2}(q_1; v_1)} dt_1
\end{aligned}$$

Corollary 1 Proof

Let $c_s = f_s/v_s$ for $s \in \{L, U\}$

$$\begin{aligned}
\Pr(f_L \leq F_1 \leq f_U | N_+ = n_+) &= \Pr\left(f_L \leq \frac{\hat{\sigma}_1/a}{\hat{\sigma}_1^2} \leq f_U | N_+ = n_+\right) \\
&= \Pr(c_L E_1 \leq G_1 \leq c_U E_1 | N_+ = n_+) \\
&= \int_{q_1}^{q_2} \frac{\Pr(c_L t_1 \leq G_1 \leq c_U t_1) f_{\chi^2}(t_1; \nu_1)}{\Pr(N_+ = n_+)} dt_1 \\
&= \int_{q_1}^{q_2} \frac{[F_{\chi^2}(c_U t_1; a, \lambda_1) - F_{\chi^2}(c_L t_1; a, \lambda_1)] f_{\chi^2}(t_1; \nu_1)}{F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)} dt_1
\end{aligned}$$

Theorem 4 Proof

An expression for unconditional power, P_w , can be written

$$\begin{aligned}
P_w &= \Pr(\text{Reject } H_0) \\
&= 1 - \Pr(\text{Accept } H_0) \\
&= 1 - \Pr\{\text{Accept } H_0 \text{ at stage 1} \cup (\text{Continue at stage 1} \cap \text{Accept } H_0 \text{ at stage 2})\} \\
&= 1 - \sum_{\{N_+ = n_+\}} \Pr(N_+ = n_+) \left\{ F_{F_1|N_+}[f_L(n_+)] + \Pr[f_L(n_+) \leq F_1 < f_U(n_+), F_+ < f_+(n_+) | N_+ = n_+] \right\}
\end{aligned}$$

$\Pr(N_+ = n_+) = F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)$ finishes the proof.

Theorem 5 Proof

$$\begin{aligned}
E(N_w) &= \sum_{\{N_+ = n_+\}} \{n_1 [1 - \Pr(f_L \leq F_1 < f_U, N_+ = n_+)] + n_+ \Pr(f_L \leq F_1 < f_U, N_+ = n_+)\} \\
&= n_1 + \sum_{\{N_+ = n_+\}} n_2 [F_{\chi^2}(q_2; \nu_1) - F_{\chi^2}(q_1; \nu_1)] \Pr(f_L \leq F_1 < f_U | N_+ = n_+)
\end{aligned}$$

Figure 1.

The general procedure. Shaded is IP only step, dashed is GS only step. All steps combined make IPIA.

Table 1

Dimensions

Symbol	Definition
n_1	Sample size, first stage
N_+	Total random sample size if study continued at first stage
N_2	Random second stage sample size, $N_+ - n_1$
N_w	Total random sample size used in study
q	Number of predictors and columns in X_0
m	minimum sample size addition for desired allocation and number of rows in X_0
r	$\text{rank}(X_0)$, $\text{rank}(X_1)$, $\text{rank}(X_+)$, and $\text{rank}(X_2)$ when $N_2 > 0$
v_s	Error df = $n_s - r$, $s \in \{1, +\}$
a	Number of rows in C , hypothesis df for test statistics
k_1	Number of replications in first stage, n_1/m
K_2	Number of replications in second stage, random, N_2/m

Table 2

Parameters and constants

Symbol	Size	Definition and Properties
X_0	$m \times q$	Fixed, known, base design matrix
X_1	$n_1 \times q$	Fixed, known, first stage design matrix
X_+	$N_+ \times q$	Final design matrix
X_2	$N_2 \times q$	Stage 2 design matrix
β	$q \times 1$	Primary parameters
C	$a \times q$	Between-subject contrast matrix
$\theta = C\beta$	$a \times 1$	Secondary parameters
θ_0	$a \times 1$	Null hypothesis values (can set to 0 WOLOG)
$M_s = C(X'_s X_s)^{-1} C'$	$a \times a$	'Middle' matrix for stage s
$H_1 = X_1(X'_1 X_1)^{-1} X'_1$	$n_1 \times n_1$	'Hat' matrix for first stage
$H_+ = X_+(X'_+ X_+)^{-1} X'_+$	$N_+ \times N_+$	'Hat' matrix for second stage
σ^2	1×1	True variance
$\delta_s = \theta' M_s^{-1} \theta$	1×1	Unscaled noncentrality parameter for stage s
$\lambda_s = \delta_s / \sigma^2$	1×1	scaled noncentrality parameter for stage s

Table 3

Internal pilot with interim analysis notation

	Symbol	Definition
Design Parameters	α_t	Target type I error rate
	P_t	Target power
	θ_1	‘Scientifically Important’ value of θ
	σ_0^2	Variance value used for planning
	n_0	Planned sample size for $\alpha_t, P_t, \theta_1, \sigma_0^2$
Sample Size	π	Proportion of n_0 used in internal pilot
Allocation	$n_1 = \pi n_0$	Internal pilot sample size
	$n_{+, \max}$	Maximum size of final sample
Unknown Parameter	$\gamma = \sigma^2 / \sigma_0^2$	Ratio of true to planning variances

Table 4

Simulation and calculation times (minutes)

	$\theta = 0$ (null)	$\theta = \theta_1$ (alt)
Sim	157.4	159.0
Calc	3.2	4.6

Table 5

Enumeration of power $\times$ 100 and $E(N_w)$

$\gamma = \sigma^2 / \sigma_0^2$	IPIA					
	Fixed Z			w/o Futility		
	Pow	N_w		Pow	$E(N_w)$	Pow
$\theta = 0$	5.4	66	6.0	36.4	6.0	36.0
	5.4	66	6.2	63.6	6.2	58.9
	5.4	66	5.6	126.4	5.3	112.1
$\theta = \theta_1$	98.1	66	85.6	35.0	85.6	35.0
	81.8	66	80.6	57.8	80.2	57.2
	53.1	66	79.4	124.4	76.6	119.3
$\theta = 2\theta_1$	>99.9	66	>99.9	34.0	>99.9	34.0
	>99.9	66	>99.9	39.5	>99.9	39.5
	98.1	66	>99.9	108.1	99.7	108.1