ISSN: 2469-5831

Jin and Kepner. Int J Clin Biostat Biom 2017, 3:013
DOI: 10.23937/2469-5831/1510013
Volume 3 | Issue 1
Open Access

International Journal of
Clinical Biostatistics and Biometrics

RESEARCH ARTICLE

On Sample Size Calculation for Exact Group Sequential Tests for Rare Disease

Man Jin[1]* and James L Kepner[2]

[1]*MRL, Merck & Co., Inc., Rahway, USA*

[2]*Alpha, Illinois, USA*

***Corresponding author:** *Dr. Man Jin, MRL, Merck & Co., Inc., Rahway, NJ 07065, USA, E-mail: man.jin@merck.com*

Abstract

For a rare disease, all the patients having the disease constitute a small population, and the standard single-stage hypergeometric test is uniformly most powerful to evaluate the response probability of a specific treatment regimen. Although exact group sequential designs are widely employed in phase II clinical trials for binomial proportions, it is unknown whether or not similar tests can be employed for hypergeometric proportions. In this manuscript, it is proved that, for hypergeometric proportions, there exist exact group sequential designs that achieve the predesignated significance level and power with maximum total sample size bounded above by the sample size for the corresponding standard exact single-stage test. Additionally, two types of optimal two-stage designs are examined for a range of design parameters; one is optimal in the sense that the expected sample size under the null hypothesis is minimized, and the other is optimal in the sense that the maximum sample size is minimized.

Keywords

Exact group sequential designs, Hypergeometric distribution, Minimax designs, Optimal designs, Small population, Uniformly most powerful test

Introduction

For a rare disease, all the patients having the disease constitute a small population of size N. Phase II trials of Investigational New Drugs (INDs) are performed in order to assess whether a new drug shows some promise of activity for the disease [1]. Imagine that if all the patients with the disease were to be treated, the new drug would show some promise of activity to *M* of them, and therefore the response rate could be defined as p = M/N. Care must be taken to explicitly define what is meant by "show some activity" [1]. For example, the drug is said to show some activity to a cancer patient whose tumor shrinks by at least 50% after the treatment.

Glycogen storage disease type II (also known as Pompe disease) is an autosomal recessive metabolic disorder which damages muscle and nerve cells throughout the body. The disease affects approximately 1 in 140,000 babies and 1 in 60,000 adults a year [2]. Von Hippel-Lindau (VHL) disease is another rare autosomal dominant syndrome which affects 1 in 36,000 babies [3].

To conduct clinical trials for such extremely rare disease, the designs developed here are based on testing a null hypothesis $H_0 : p \le p_0$ that the true response rate is less than some uninteresting level p_0; that is, the new drug shows some activity to fewer than $M_0 = Np_0$ patients. If the null hypothesis is true, then we require that the probability of falsely concluding that the drug is efficacious is less than a user specified α. We also require if a specified alternative hypothesis $H_1 : p \le p_1$ (that is, more than $M_1 = Np_1$ patients response to the drug) is true with $p_0 \le p_1$, then the probability of falsely concluding that the drug is not efficacious is less than a user specified β.

We start our discussion by considering the standard one-stage design for testing

$$H_0 : p \le p_0 \text{ vs. } H_1 : P > p_0, \text{ or equivalently } H_0 : M \le M_0 \text{ vs. } H_1 : M > M_0. \quad (1)$$

In order to test the above hypotheses, a sample of n patients selected randomly from the population of N patients is treated. Let S denote the number of patients

Citation: Jin M, Kepner JL (2017) On Sample Size Calculation for Exact Group Sequential Tests for Rare Disease. Int J Clin Biostat Biom 3:013. doi.org/10.23937/2469-5831/1510013

Received: March 08, 2017: **Accepted:** October 07, 2017: **Published:** October 09, 2017

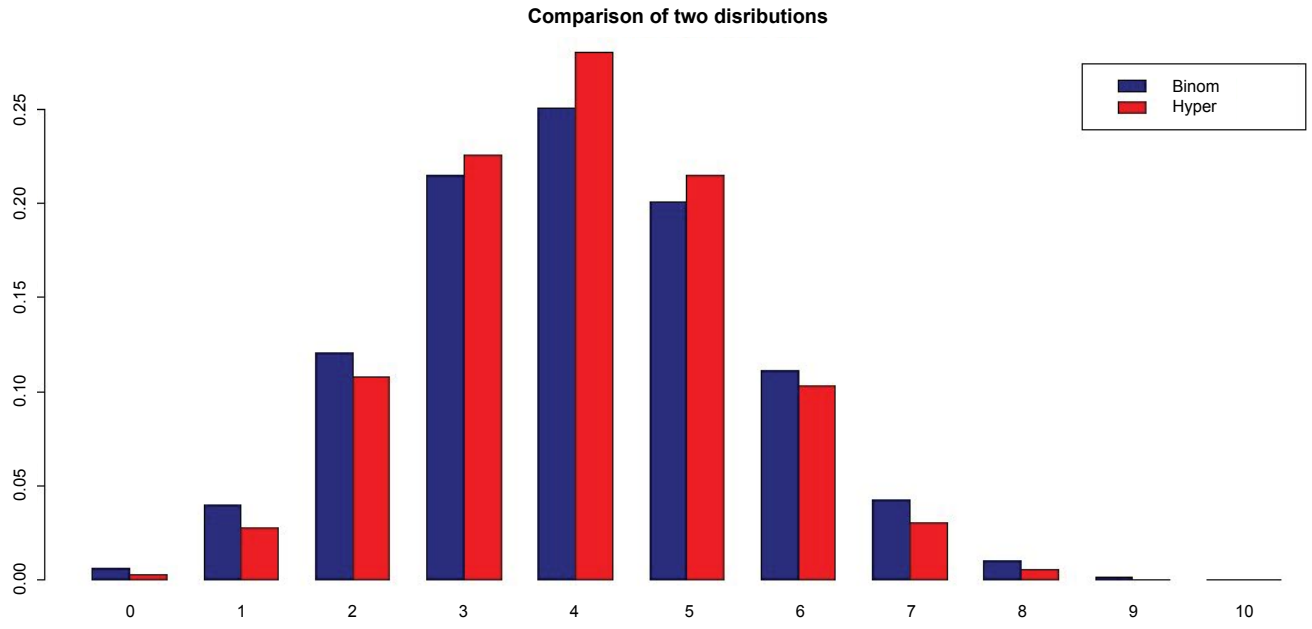

Figure 1: Comparing hypergeometric distribution $h(s;50,20,10)$ and binomial distribution $b(s;10,0.4)$, where $s=0,1,...,10$.

in the sample who respond to the treatment, which is following the hypergeometric distribution, denoted as H (N; M; n),

$$\Pr(S=s)=h(s;N,M,n)=\binom{M}{s}\binom{N-M}{n-s}\Big/\binom{N}{n}, \quad (2)$$

Where $s=\max(0;n+M-N)$; ... ; $\min(n;M)$. Consider the test with reject region $S\le b+1$, where b is an integer such that $\Pr(S\le b+1\mid p=p_0)\le\alpha$. We can show that this test is Uniformly Most Powerful (UMP) test for hypotheses (1).

Group sequential designs are widely employed in phase II and phase III clinical trials. A group sequential test allows early termination of accrual if the treatment response rate of the treatment is quite high or quite low during early stages. Fleming [4], Chang, et al. [5], and Simon [6] pointed out that for binomial distributions; well-designed group sequential tests are more efficient than standard single-stage tests, because on average they require fewer patients to achieve the predesignated significance level and power. However, it is still unknown that for hypergeometric distributions, whether or not there exist exact group sequential tests which are more efficacious than the aforementioned standard single-stage exact test.

It is known that when the disease population size N is large, the hypergeometric distribution $h(s;N,M,n)$ can be approximately well by the binomial distribution $b(s;n,p)$ where $p=M/N$ and then we can use the available designs [6]. But for extremely rare diseases, using designs based on binomial distribution may have incorrect significant level; that is, the true type-I error may be different from the prespecified type-I error. In Figure 1, we display a toy example, showing the difference between two distributions, hypergeometric distribution $h(s;50,20,10)$ and binomial distribution $b(s;10,0.4)$, where $s=0,1,...,10.$

Motivated by Kepner and Chang [7,8], in this manuscript we show that, for hypergeometric distributions, under mild conditions and for a given number $k\ge 2$, there exists at least one k-stage group sequential test which has exactly the same maximum total sample size, significance level, and power as the standard single-stage hypergeometric test. Furthermore, motivated by Simon [6], we examine two types of optimal two-stage designs for a range of design parameters; one is optimal in the sense that the expected sample size under the null hypothesis is minimized, and the other is optimal in the sense that the maximum sample size is minimized.

The remainder of the manuscript is organized as follows. In Section 2, group sequential procedures are described. In Section 3, main theorems are presented and proved. In Section 4, two types of optimal two-stage designs are examined. Some brief discussion is given in Section 5 (Figure 1).

Group Sequential Tests

Group sequential tests are specified by the maximum number of stages, k, the cumulative number of patients to be treated up to each stage i, n_i and the critical values, $\{[a_1,b_1],\cdots,[a_{k-1},b_{k-1}],b_k\}$, where $a_i\le b_i, i=1,...,k-1$. Let S_i be the number of patients who respond positively to the treatment among n_i patients cumulative up to stage i. The distribution of S_i is $\Pr(S_i=s)=h(s;N,M,n_i)$ where $s=\max(0,n+M-N),...,\min(n_i,M)$. The distribution is undetermined up to an unknown parameter, M or equivalently, $p=M/N$.

The group sequential tests are conducted as follows. Start with stage $i=1$. If $S_i\le a_i-1$, stop sampling and

reject H_1; if $S_i \geq b_i + 1$, stop sampling and accept H_1; if $a_i \leq S_i \leq b_i$, continue to stage $i+1$. At the final stage, accept H_1 if $S_k \geq b_k + 1$ and reject H_i if $S_k \leq b_k$.

The power function for any such group sequential test is

$$P(M) = \Pr\{S_1 \geq b_1 + 1 \mid M\} + \Pr\{a_1 \leq S_1 \leq b_1, S_2 \geq b_1 + 1 \mid M\} + \ldots + \Pr\{a_1 \leq S_1 \leq b_1, \ldots, a_{k-1} \leq S_{k-1} \leq b_{k-1}, S_k \geq b_k + 1 \mid M\}. \quad (3)$$

For the required significance level α and power $1-\beta$ at $M_0 + \Delta$, one should select $(n_1, \ldots, n_k)$ and $[(a_1, b_1), \ldots, (a_{k-1}, b_{k-1}), b_k]$ such that $P(M_0) \leq \alpha$ and $P(M_0 + \Delta) \geq 1 - \beta$. The power function can be written as a function of proportion p. To abuse the notation slightly, let the power function (3) be written as $P(p)$. Therefore, for the required significance level α and power $1-\beta$ at $p_0 + \delta$, where $p_0 = M_0 / N$ and $\delta = \Delta / N$, one should select $(n_1, \ldots, n_k)$ and $[(a_1, b_1), \ldots, (a_{k-1}, b_{k-1}), b_k]$ such that $P(p_0) \leq \alpha$ and $P(p_0 + \delta) \geq 1 - \beta$.

As discussed in Kepner and Chang [7], there are three types of group sequential designs. Type 1 designs stop early only to conclude efficacy (i.e., stop early only to accept H_1), Type 2 design stops early only to conclude futility (i.e., stop early only to reject H_1, and Type 3 designs stop early for either efficacy or futility. Using the above notation, Type 1 group sequential tests are those with $a_1 = a_2 = \ldots = a_{k-1} = 0$, Type 2 group sequential tests are those with $b_1 = b_2 = \ldots = b_{k-1} = n_i$, and Type 3 group sequential tests are those with $0 < a_i \leq b_i \leq n_i$, for $i = 1, \ldots, k-1$.

Main Theorems

In this section, three theorems are established, one for each of three types of exact group sequential tests discussed in Section 2. These three theorems are similar to those established in Kepner and Chang [7] which were for binomial distributions.

Assume:

$(A1)\ S \sim\sim H(N, M, n)$;

$(A1)$ n is the smallest sample size for which there is an integer b such that $0 \leq b+1 \leq n$, $\Pr\{S \geq b+1 \mid M = M_0\} \leq \alpha$, and $\Pr\{S \geq b+1 \mid M = M_0 + \Delta\} \geq 1 - \beta$, where $0 \leq \Delta < N - M_0$; $(A3)$ k is an integer such that $2 \leq k \leq n$.

Theorem 1

Under the conditions (A1)-(A3), assume $b \leq n - k$. Then there exists a Type 1 k-stage exact sequential test such that $n_1 \geq n_2 - n_1 \geq \ldots \geq n_k - n_{k-1}$ and $n_k = n$ the significance level of the test is α and the power of the test at $M = M_0 + \Delta$ is at least $1 - \beta$.

Proof: Let $n_1 = n - k + 1$ and $n_i = n_{i-1} + 1$ for $i = 2, \ldots, k$. The test statistic at stage i is S_i. Let $b_1 = b_2 = \ldots = b_k = b$ and $a_1 = a_2 = \ldots = a_{k-1} = 0$. Since $b_1 + 1 = b + 1 \leq n - k + 1 = n_1 \leq n_i$ for $i = 1, \ldots, k$, then the sample size $(n_1, \ldots, n_k)$ and the critical values $[(a_1, b_1), \ldots, (a_{k-1}, b_{k-1}), b_k]$ become a proper Type 1 k-stage test. According to this test, H_1 is accepted if and only if at least one $S_i \geq b+1$, $1 \leq i \leq k$, which is equivalent to $S_k = S \geq b+1$, since $S_1 \leq S_2 \leq \ldots \leq S_k = S$. Thus, the power function of the test is the same as that of S.

Theorem 2

Under the conditions (A1)-(A3), assume $b \geq k-1$. Then there exists a Type 2 k-stage exact sequential test such that $n_1 \geq n_2 - n_1 \geq \ldots \geq n_k - n_{k-1}$ and $n_k = n$, the significance level of the test is α and the power of the test at $M = M_0 + \Delta$ is at least $1 - \beta$.

Proof: Let $n_1 = n - k + 1$ and $n_i = n_{i-1} + 1$ for $i = 2, \ldots, k$. The test statistic at stage i is S_i. Let $a_i = b - k + i + 1$ and $b_i = n_i$ for $i = 1, \ldots, k-1$ and $b_k = b$. Since $1 \leq a_1 \leq b_i = n_i$, $i = 1, \ldots, k-1$ and $k \leq b+1 \leq n$, then the sample size $(n_1, \ldots, n_k)$ and the critical values $[(a_1, b_1), \ldots, (a_{k-1}, b_{k-1}), b_k]$ become a proper Type 2 k-stage test. According this test H_1 is rejected if and only if at least one $S_i \leq a_i = b - k + i + 1$. That is, there exists at least one $i, 1 \leq i \leq k,$ such that $n_i - S_i \leq n_i - b + k + i = n - b$. This is equivalent to $n - S_k \leq n - b$, because $n_1 - S_i \leq n_2 - S_2 \leq \ldots \leq n_k - S_k$. Thus, the power function of the test is the same as that of S.

After Theorems 1-2 are established, following the same proof as that for Theorem 3 in Kepner and Chang [7], we can establish the following theorem.

Theorem 3

Under the conditions (A1) - (A3), assume $k - 1 \leq b \leq N - k$. Then there exists a Type 3 exact k-stage sequential test such that $n_1 \geq n_2 - n_1 \geq \ldots \geq n_k - n_{k-1}$ and $n_k = n$, the significance level of the test is α and the power of the test at $M = M_0 + \Delta$ is at least $1 - \beta$.

The implications of these theorems are two-fold. The theoretical implication is that, for a given number $k \geq 2$ satisfying some mild conditions, there exists at least one k-stage group sequential test whose maximum total sample size is bounded above by the sample size needed for the standard one-stage test to achieve the same significance level and power. The applied implication is that we can search for "optimal" exact k stage designs among those proper group sequential tests whose maximum total sizes are bounded above by the sample size needed for the standard exact one-stage test to achieve the same significance level and power.

Optimal Two-Stage Designs

In this section we focus on optimal two-stage designs, which can be extended to optimal k-stage designs. We examine the two types of optimal two-stage designs proposed in Simon [6]; one is optimal in the sense that the expected sample size under the null hypothesis is minimized, and the other is optimal in the sense that the maximum sample size is minimized. Originally they were proposed for binomial distributions. In this manuscript, we examine them for hypergeometric distributions. But there is a small difference. In Simion [6], the optimal designs are searched among all the proper two-stage designs. Hereafter, thanks to Theorems 1-3, we

only need to search for optimal designs among those proper two-stage designs whose maximum total sizes are bounded above by the sample size needed for the standard one-stage design to achieve the same significance level and power. Narrowing the search domain improves the computational efficiency.

Consider the first type of optimal two-stage designs. If there are n_1 patients in the first stage and, if necessary, $n_2 - n_1$ more patients are treated in the second stage, leading to maximum sample size to be n_2. Then the expected sample size is $EN = PET \times n_1 + (1 - PET) \times n_2$, where PET is the probability of early termination after the first stage. The decision of whether or not to terminate after the first stage depends on the type of the two-stage design (Type 1, Type 2, or Type 3) and the number of responses observed from those n_1 patients (Table 1). The expected sample size EN and the probability of early termination PET depend on the unknown parameter, M, the number of responses observed in the sample of size n_1 from the population of size N. In particular, $PET = 1 - \sum_{s=a_1}^{b_1} h(s; N, M, n_1)$. In this manuscript, as in Simon [6], we consider the optimal two-stage designs which have minimum expected sample size under the null hypothesis, $EN_0 = PET_0 \times n_1 + (1 - PET_0) \times n_2$ where $PET = 1 - \sum_{s=a_1}^{b_1} h(s; N, M_0, n_1)$. The rationale behind this is that we should expose as few patients as possible to an ineffective treatment. The second type of optimal two-stage designs is easier to describe. They are the ones that have minimum maximum sample size; that is, n_2 is minimized.

Now we are ready to examine these two types of optimal two-stage designs. First, we refresh notation for interpreting different designs in Table 1. We consider the following settings. Set significance level $\alpha = 0.05$ and Type II error rate $\beta = 0.2$. Set the population size N = 80 or 120, the proportion under the null hypothesis $p_0 = M_0/N = 0.1, 0.2, ..., 0.7$ and the treatment effect $\delta = \Delta/N = 0.15$ or 0:2. The resulting optimal designs are presented in Table 2, Table 3, Table 4 and Table 5. We have obtained these results using R codes, which can be requested by sending email to the first author.

Table 2 and Table 3 apply to a small population of size $N = 80$ and Table 4 and Table 5 apply to a small

Table 1: Notation for interpreting different designs.

Type	Design	Reject H1	Accept H1
1-Stage	(n; b)	fS bg	fS b + 1g
Type 1	(n1; b1; n2; b2)	fS1 b1g and fS2 b2g	fS1 b1 + 1g or fS1 b1g and fS2 b2 + 1g
Type 2	(n1; a1; n2; b2)	fS1 a1 1g or fS1 a1g and fS2 b2g	fS1 a1g and fS2 b2 + 1g
Type 3	(n1; a1; b1; n2; b2)	fS1 a1 1g or fa1 S1 b1g and fS2 b2g	fS1 b1 + 1g or fa1 S1 b1g and fS2 b2 + 1

Table 2: N = 80. Designs for testing H_0:p p_0 vs. H_1:p > p_0 with = 0:05 and power at least 80% at $p_1 = p_0 + 0:15$. For each setting the Rows 1-3 contain the Types 1-3 two-stage designs. The left and rights parts are for the first type and second type of optimal designs.

p0 (1stage)	2-stage minimizing EN0				2-stage minimizing n2			
	Design	EN0	n2	P ET0	Design	EN0	n2	P ET0
0.1	(24, 4; 26, 5)	25.9	26	0.05	(24, 4; 26, 5)	25.9	26	0.05
(29; 5)	(15, 2; 29, 5)	21.5	29	0.53	(15, 2; 29, 5)	21.5	29	0.53
	(15, 2, 3; 28, 5)	20.6	28	0.57	(24, 1, 4; 26, 5)	25.8	26	0.1
0.2	(21, 7; 32, 9)	31.8	32	0.02	(21, 7; 32, 9)	31.8	32	0.02
(36; 10)	(17, 4; 33, 9)	24.3	33	0.54	(17, 4; 33, 9)	24.3	33	0.54
	(17, 4, 7; 33, 9)	24.2	33	0.55	(21, 1, 7; 32, 9)	31.7	32	0.03
0.3	(33, 13; 36, 14)	35.9	36	0.04	(33, 13; 36, 14)	35.9	36	0.04
(37; 14)	(31, 16; 37, 14)	27.8	37	0.44	(31, 16; 37, 14)	27.8	37	0.44
	(21, 7, 10; 37, 14)	28	37	0.56	(31, 1, 13; 36, 14)	35.9	36	0.04
0.4	(20, 12; 37, 18)	36.8	37	0.01	(20, 12; 37, 18)	36.8	37	0.01
(39; 19)	(22, 9; 39, 19)	31.5	39	0.44	(22, 9; 39, 19)	31.5	39	0.44
	(23, 10, 13; 39, 19)	29.7	39	0.57	(20, 1, 12; 37, 18)	36.8	37	0.01
0.5	(21, 14; 39, 23)	38.6	39	0.02	(21, 14; 39, 23)	38.6	39	0.02
(39; 23)	(26, 15; 38, 22)	28.8	38	0.76	(26, 15; 38, 22)	28.8	38	0.76
	(26, 15, 21; 38, 22)	28.8	38	0.76	(26, 15, 21; 38, 22)	28.8	38	0.76
0.6	(31, 22; 36, 25)	35.8	36	0.03	(31, 22; 36, 25)	35.8	36	0.03
(38; 26)	(21, 14; 38, 26)	26.5	38	0.68	(24, 16; 35, 24)	27.2	35	0.71
	(21, 14, 17; 38, 26)	26.4	38	0.68	(24, 16, 20; 35, 24)	27.2	35	0.71
0.7	(25, 21; 33, 26)	32.9	33	0.01	(25, 21; 33, 26)	32.9	33	0.01
(33; 26)	(18, 14; 33, 26)	22.6	33	0.69	(26, 21; 32, 25)	26.7	32	0.89
	(18, 14, 16; 33, 26)	22.5	33	0.7	(26, 21, 23; 32, 25)	26.7	32	0.89

population of size $N = 120$, where δ is 0:15 for Table 2 and Table 4 is 0:20 for Table 3 and Table 5. In each table, the first column corresponds to one-stage design, the optimal two-stage designs minimizing the expected sample size under H_0 are shown on the left half of the table, and the optimal two-stage designs minimizing the maximum sample size are shown in the right half of the table. For each setting, there are three rows, which are corresponding to three types of two-stage designs (Types 1-3), respectively. The tabulated results include the optimal design (Table 1), the expected sample size

Table 3: N = 80. Designs for testing H_0:p p_0 vs. H_1:p > p_0 with = 0:05 and power at least 80% at $p_1 = p_0 + 0{:}2$. For each setting the Rows 1-3 contain the Types 1-3 two-stage designs. The left and rights parts are for the first type and second type of optimal designs.

p0	2-stage minimizing EN0				2-stage minimizing n2			
(1-stage)	Design	EN0	n2	P ET0	Design	EN0	n2	P ET0
0.1	(15, 3; 19, 4)	18.9	19	0.04	(15, 3; 19, 4)	18.9	19	0.04
(21; 4)	(11, 2; 21, 4)	14	21	0.7	(11, 2; 21, 4)	14	21	0.7
	(11, 2, 3; 21, 4)	13.9	21	0.71	(15, 1, 3; 19, 4)	18.2	19	0.21
0.2	(13, 5; 23, 7)	22.8	23	0.02	(13, 5; 23, 7)	22.8	23	0.02
(23; 7)	(17, 4; 23, 7)	19.7	23	0.54	(17, 4; 23, 7)	19.7	23	0.54
	(13, 3, 5; 23, 7)	17.9	23	0.51	(13, 3, 5; 23, 7)	17.9	23	0.51
0.3	(21, 9; 26, 11)	25.8	26	0.04	(21, 9; 26, 11)	25.8	26	0.04
(27; 11)	(14, 5; 27, 11)	19.4	27	0.59	(14, 5; 27, 11)	19.4	27	0.59
	(14, 5, 7; 27, 11)	19.1	27	0.61	(21, 1, 9; 26, 11)	25.8	26	0.04
0.4	(15, 9; 27, 14)	26.7	27	0.02	(15, 9; 27, 14)	26.7	27	0.02
(29; 15)	(16, 8; 28, 14)	19.2	28	0.74	(16, 8; 28, 14)	19.2	28	0.74
	(16, 8, 11; 28, 14)	19.1	28	0.74	(15, 1, 9; 27, 14)	26.7	27	0.02
0.5	(13, 10; 26, 16)	25.9	26	0.01	(13, 10; 26, 16)	25.9	26	0.01
(26; 16)	(13, 7; 26, 16)	19.5	26	0.5	(13, 7; 26, 16)	19.5	26	0.5
	(13, 7, 10; 26, 16)	19.4	26	0.51	(13, 7, 10; 26, 16)	19.4	26	0.51
0.6	(15, 12; 25, 18)	24.8	25	0.02	(15, 12; 25, 18)	24.8	25	0.02
(25; 18)	(16, 11; 25, 18)	18.8	25	0.69	(16, 11; 25, 18)	18.8	25	0.69
	(16, 11, 13; 25, 18)	18.7	25	0.7	(16, 11, 13; 25, 18)	18.7	25	0.7
0.7	(11, 10; 22, 18)	21.8	22	0.01	(11, 10; 22, 18)	21.8	22	0.01
(22; 18)	(12, 10; 21, 17)	14.1	21	0.77	(12, 10; 21, 17)	14.1	21	0.77
	(11, 9, 10; 22, 18)	14.1	22	0.72	(13, 11, 12; 21, 17)	14.4	21	0.83

Table 4: N = 120. Designs for testing H_0: p p_0 vs. H_1: p > p_0 with = 0: 05 and power at least 80% at $p_1 = p_0 + 0{:}15$. For each setting the Rows 1-3 contain the Types 1-3 two-stage designs. The left and rights parts are for the first type and second type of optimal designs.

p0	2-stage minimizing EN0				2-stage minimizing n2			
(1-stage)	Design	EN0	n2	P ET0	Design	EN0	n2	P ET0
	(19, 4; 29, 5)	28.8	29	0.02	(19, 4; 29, 5)	28.8	29	0.02
0.1	(22, 3; 29, 5)	24.7	29	0.62	(22, 3; 29, 5)	24.7	29	0.62
(29; 5)	(20, 3, 4; 29, 5)	22.6	29	0.71	(20, 3, 4; 29, 5)	22.6	29	0.71
	(35, 10; 38, 11)	37.9	38	0.04	(35, 10; 38, 11)	37.9	38	0.04
0.2	(23, 5; 39, 11)	31.1	39	0.49	(23, 5; 39, 11)	31.1	39	0.49
(39; 11)	(22, 5, 8; 39, 11)	29.7	39	0.55	(35, 1, 10; 38, 11)	37.9	38	0.04
	(24, 11; 44, 17)	43.6	44	0.01	(24, 11; 44, 17)	43.6	44	0.01
0.3	(26, 9; 42, 16)	31.8	42	0.64	(24, 8; 42, 16)	31.8	42	0.57
(44; 17)	(26, 9, 13; 42, 16)	31.7	42	0.64	(24, 8, 13; 42, 16)	31.8	42	0.57
	(29, 16; 45, 22)	42.7	45	0.02	(29, 16; 45, 22)	42.7	45	0.02
0.4	(31, 13; 45, 22)	37.7	45	0.52	(31, 13; 45, 22)	37.7	45	0.52
(45; 22)	(30, 13, 17; 45, 22)	36	45	0.6	(30, 13, 17; 45, 22)	36	45	0.6
	(27, 18; 44, 26)	43.8	44	0.01	(22, 16; 44, 26)	43.9	44	0.01
0.5	(24, 13; 46, 27)	33	46	0.59	(24, 13; 46, 27)	33	46	0.59
(46; 27)	(24, 13, 17; 46, 27)	32.9	46	0.6	(22, 1, 16; 44, 26)	43.9	44	0.61
	(29, 22; 42, 29)	41.8	42	0.01	(29, 22; 42, 29)	41.8	42	0.01
0.6	(21, 13; 42, 29)	32	42	0.48	(21, 13; 42, 29)	32	42	0.48
(42; 29)	(26, 17, 20; 42, 29)	31.3	42	0.67	(26, 17, 20; 42, 29)	31.3	42	0.67
	(23, 20; 38, 30)	37.9	38	0.01	(23, 20; 38, 30)	37.9	38	0.01
0.7	(24, 19; 37, 29)	26.6	37	0.8	(24, 19; 37, 29)	26.6	37	0.8
(38; 30)	(22, 17, 19; 38, 30)	26.5	38	0.72	(24, 19, 22; 37, 29)	26.6	37	0.8

under H_0, EN_0, the maximum sample size, n_2, and the early termination probability under H_0, PET_0. For some settings, there is more than one optimal design, of which the minimum maximum size is the same as the sample size in the one-stage design, then the one which has minimum expected sample size under H_0 is reported and it is indicated by an asterisk.

First and most importantly, the results in Table 2, Table 3, Table 4 and Table 5 verify the main theorems; that is, for each setting, there exists at least one two-stage design whose maximum sample size is bounded above by the sample size of the one-stage design. Because the early termination probability is strictly positive for any proper two-stage design, the corresponding expected size is always strictly smaller than the sample size of the one-stage design.

Second, in many settings, the maximum sample sizes of the two-stage designs are strictly smaller than the sample size of the corresponding one-stage design. This finding is striking, noting that the standard one-stage test is uniformly most power test.

Third, in some settings, the optimal two-stage design minimizing the expected sample size may be more attractive than the optimal two-stage design minimizing the maximum sample size. This is the case when the difference in maximum sample size is small, but the difference in expected sample sizes is large. For example, in Table 2, the third row in the session of $p_0 = 0.2$ indicates cases where the difference in maximum sample size is only one, but the difference in expected sample sizes is 7:5. For other settings, these two types of optimal exact two-stage designs are quite similar. Surprisingly, this finding is different from that in Simon [6], where it was concluded that the "minimax" designs may be more attractive.

Discussion

Optimal exact group sequential tests for a binomial proportion have been well-studied in the literature, but a corresponding study involving hypergeometric distributions is lacking. This manuscript studies some exact group sequential tests involving hypergeometric distributions, which are useful for investigating treatment effects on rare diseases.

In this manuscript, three theorems have been proved and two types of optimal two-stage designs are presented. The theorems guarantee the existence of proper exact group sequential designs whose expected sample sizes are strictly smaller than the ones from standard one-stage designs. The discussed optimal two-stage designs provide two examples of how to design optimal two stage designs. There are other criteria; for example, the optimal design minimizing the expected sample size at a given parameter, say $p_1 = p_0 + \delta$. Moreover, the tabulated results provide detailed comparisons between these two types of optimal designs.

Finally, this manuscript focuses on one-arm phase II clinical trials. One of our future projects is to study the properties of group sequential tests for comparing two hypergeometric distributions, which arise from two-arm phase II clinical trials.

Acknowledgment

The authors would like to thank two anonymous referees for their constructive comments and suggestions, which have led to a significantly improved paper.

Table 5: N = 120. Designs for testing H_0:p p_0 vs. H_1:p > p_0 with = 0:05 and power at least 80% at $p_1 = p_0$ + 0:2. For each setting the Rows 1-3 contain the Types 1-3 two-stage designs. The left and rights parts are for the first type and second type of optimal designs.

p0	2-stage minimizing EN0				2-stage minimizing n2			
(1-stage)	Design	EN0	n2	P ET0	Design	EN0	n2	P ET0
0.1	(14, 3; 20, 4)	19.8	20	0.03	(13, 3; 20, 4)	19.8	20	0.03
(21; 4)	(12, 2; 21, 4)	15.1	21	0.66	(12, 2; 21, 4)	15.1	21	0.66
	(12, 2, 3; 21, 4)	14.9	21	0.68	(13, 1, 3; 21, 4)	18.2	20	0.26
0.2	(16, 6; 26, 8)	25.8	26	0.02	(16, 6; 26, 8)	25.8	26	0.02
(26; 8)	(17, 4; 26, 8)	21.1	26	0.54	(17, 4; 26, 8)	21.1	26	0.54
	(13, 3, 5; 26, 8)	19.3	26	0.52	(13, 3, 5; 26, 8)	19.3	26	0.52
0.3	(16, 8; 29, 12)	28.8	29	0.02	(16, 8; 29, 12)	28.8	29	0.02
(29; 12)	(22, 8; 29, 12)	24.2	29	0.68	(22, 8; 29, 12)	24.2	29	0.68
	(19, 7, 9; 29, 12)	22	29	0.7	(19, 7, 9; 29, 12)	22	29	0.7
0.4	(20, 12; 29, 15)	28.9	29	0.01	(20, 12; 29, 15)	28.9	29	0.01
(31; 16)	(16, 7; 31, 16)	23.1	31	0.53	(16, 7; 31, 16)	23.1	31	0.53
	(19, 9, 12; 31, 16)	22.8	31	0.69	(20, 1, 12; 29, 15)	28.9	29	0.01
0.5	(15, 11; 29, 18)	28.8	29	0.01	(15, 11; 29, 18)	28.8	29	0.01
(29; 18)	(22, 12; 29, 18)	24.8	29	0.59	(22, 12; 29, 18)	24.8	29	0.59
	(20, 11, 14; 29, 18)	23.5	29	0.61	(20, 11, 14; 29, 18)	23.5	29	0.61
0.6	(24, 19; 28, 20)	28	28	0.01	(24, 19; 28, 20)	28	28	0.01
(28; 20)	(17, 12; 28, 20)	19.7	28	0.75	(17, 12; 28, 20)	19.7	28	0.75
	(17, 12, 14; 28, 20)	19.6	28	0.76	(17, 12, 14; 28, 20)	19.6	28	0.76
0.7	(19, 16; 23, 19)	22.9	23	0.03	(19, 16; 23, 19)	22.9	23	0.03
(23; 19)	(11, 9; 22, 18)	14.3	22	0.7	(11, 9; 22, 18)	14.3	22	0.7
	(12, 10, 11; 22, 18)	14.3	22	0.77	(12, 10, 11; 22, 18)	14.3	22	0.77

References

1. Green S, Smith A, Benedetti J, Crowley J (2012) Clinical Trials in Oncology. CRC Press.
2. Ausems M, Verbiest J, Hermans M, Kroos M, Beemer F, et al. (1999) Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counseling. Eur J Hum Genet 7: 713-716.
3. Richard S, Gardie B, Couv S, Gad S (2013) Von Hippel-Lindau: How a rare disease illuminates cancer biology. Semin Cancer Biol 23: 26-37.
4. Fleming T (1982) One-sample multiple testing procedures for phase II clinical trials. Biometrics 38: 143-151.
5. Chang M, Therneau T, Wieand HS, Cha S (1987) Designs for group sequential phase II clinical trials. Biometrics 43: 865-874.
6. Simon R (1989) Optimal two-stage designs for phase II clinical trials. Controlled Clin Trials 10: 1-10.
7. Kepner J, Chang M (2003) On the maximum total sample size of a group sequential test about binomial proportions. Statistics and Probability Letters 62: 87-92.
8. Kepner J, Chang M (2004) Samples of exact k-stage group sequential designs for phase II and pilot studies. Controlled Clin Trials 25: 326-333.