Establishing Non-Inferiority of a New Treatment in a Three-Arm Trial: Apply a Step-Down Hierarchical Model in a Papulopustular Acne Study and an Oral Prophylactic Antibiotics Study

Jung-Tzu Liu[1,2] , Chyng-Shyan Tzeng[2] and Hsiao-Hui Tsou[1,3,*]

[1]*Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan*

[2]*Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan*

[3]*Graduate Institute of Biostatistics, College of Public Health, China Medical University, Taichung, Taiwan*

Abstract: Clinical trials comparing a test treatment with an active control therapy have become very popular in drug and medical device development in the last decade. An active controlled trial without a placebo, however, exhibits some major challenges in design, analysis, and interpretation, such as the determination of the non-inferiority margin or the assumption of constancy condition. When there are no ethical concerns, the comparison of a test drug with placebo usually provides the most convincing proof of the efficacy of a new treatment. Therefore, it may be advisable to conduct a three-arm trial — including placebo, active control, and the new treatment — if it is ethically justifiable such as a papulopustular acne study and an oral prophylactic antibiotics study. In this paper, we propose a statistical methodology for a three-arm non-inferiority trial with binary outcomes. We adapt the step-down hierarchical hypotheses and give a three-step testing procedure which is more realistic in conducting a clinical trial. We derived an optimal sample size allocation rule in an ethical and reliable manner to minimize the total sample size and hence to shorten the duration of the trials. Real examples from a papulopustular acne study and an oral prophylactic antibiotics study are used to demonstrate our methodology.

Keywords: Clinical trial, binary outcomes, gold standard design, optimal sample size allocation, restricted maximum likelihood.

1. INTRODUCTION

The main purpose of a clinical trial is to demonstrate efficacy of a new treatment. Many investigators adapt an active controlled non-inferiority clinical trial without a placebo because of ethical concerns. Such a trial, however, exhibits some major challenges in design, analysis, and interpretation, such as the determination of the non-inferiority margin or the assumption of constancy condition. Because of the absence of a placebo arm, one cannot assert directly that the test treatment is superior to a placebo. Moreover, many researchers may not pay attention to verifying assay sensitivity, that is, to showing that the active control is better than the placebo in the current trial. When the placebo arm is present in the non-inferiority trial if ethically justifiable, the assay sensitivity can be concurrently verified and the issues discussed above will not be present.

Several studies have presented useful ideas for non-inferiority trial designs. Some of them suggest including a placebo arm in a non-inferiority trial when ethically justifiable (see [1−2]) and propose statistical methodologies for such design; these include Pigeot *et al*. [3], Koch and Röhmel [4], and Hauschke and Pigeot [5] for continuous outcome; Tang and Tang [6], Kieser and Friede [7], Hasler [8] for binary outcome; and Mielke *et al*. [9] for survival data, and Kombrink *et al*. [10] for censored time-to-event data. A three-arm trial including placebo, active control and test drug is referred to as gold standard design [5]. In this design, the hypotheses can be formulated more precisely, the non-inferiority of the test treatment to an active control can be verified, and the efficacy of the test treatment can be accessed directly. Tang and Tang [6] proposed sample size allocation rules for a three-arm clinical trial by using binary outcomes based on rate difference. However, they did not consider the optimal sample size determination. Kieser and Friede [7] derived approximate sample size formulas in each patient group and proposed a complete two-step test procedure. Koch and Röhmel [4] and Hauschke and Pigeot [5] suggested comparing the test treatment with a placebo in the first step. Emphasizing the importance of comparing the test treatment versus a placebo, they indicated that nothing can rescue such a trial if the superiority of an experimental over a placebo cannot be shown. Hence, we propose a testing procedure with hierarchical hypotheses based on Koch and Röhmel's

*Address correspondence to this author at the Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli County 35053, Taiwan; Tel: 886-37-246-166, ext. 36181; Fax: 886-37-586-467;
E-mail: tsouhh@nhri.org.tw

E-ISSN: 1929-6029/14 © 2014 Lifescience Global

[4] suggestion and derive an optimal sample size determination of a three-arm trial.

The outline of this paper is as follows. In Section 2, we present the models and hypotheses for a three-arm gold standard design. The null variance estimation based on two methods is proposed and an optimal sample size allocation is also given. The simulation outcomes of type I error rate, and power are displayed in Section 3. We apply the proposed method to a papulopustular acne study and an oral prophylactic antibiotics study in Section 4. Finally, we conclude with discussion in Section 5.

2. METHODOLOGY

2.1. Model and Hypotheses

In this article, we propose a statistical method for binary outcomes, such as improvement/no improvement and remission/no remission. We consider the primary clinical outcomes under a placebo, an active control, and a test treatment (X_P, X_C, and X_T), respectively, as independent and binomially distributed variables. That is, we assume that $X_k \sim B(n_k, \pi_k)$, where the success rate π_k represents unknown true response probability and n_k be the sample size, $k = P$, C, and T.

Under gold standard three-arm design, we adapted the step-down hierarchical hypotheses for binomial outcomes. In the first step, we compared a test treatment with a placebo by the following hypothesis

$$H_{01}: \pi_T \leq \pi_P. \quad (1)$$

If the superiority of an experimental over a placebo cannot be shown, nothing can rescue such a trial. Thus, it is reasonable to compare the test treatment versus a placebo in the first step. If H_{01} is rejected, we claimed that the test treatment is superior to the placebo and executed the second-step procedure.

In the second step, we compared an active control with a placebo by the following hypothesis

$$H_{02}: \pi_C \leq \pi_P. \quad (2)$$

Similarly, we claimed that the active control is superior to the placebo if H_{02} is rejected. Consequently, the assay sensitivity was established.

After both hypotheses H_{01} and H_{02} were rejected, the non-inferiority hypothesis for a test treatment versus an active control was accessed at level α with a pre-specified non-inferiority margin δ ($\delta > 0$). In other words, we wanted to ensure

$$\pi_T - \pi_C > -\delta.$$

The margin can be a function θ of difference between response probabilities π_C and π_P [3], that is $\delta = (1 - \theta)(\pi_C - \pi_P)$, where θ is a pre-specified fixed fraction of active controlled effect. Koch and Tangen [11] mentioned the reasonable region for non-inferiority test is for θ between 0.5 and 0.99. Therefore, the null hypothesis could be simply rewritten as

$$H_{03}: (\pi_T - \pi_P) / (\pi_C - \pi_P) \leq \theta \quad (3)$$

If H_{03} is successfully rejected for a given θ, we claim that the test treatment retains more than $\theta \times 100\%$ efficacy of the active control compared with the placebo. Therefore, the non-inferiority of the test treatment to the active control is declared.

2.2. Statistical Hierarchical Test Procedures

We described the step-down hierarchical hypotheses in the previous section. According to the hierarchical testing procedures (Figure **1**), the familywise error rate (FWER) could be controlled at the same level, α.

Statistical test procedures for hypotheses in (1) and (2) can be established according to conventional method [12], we now only focus on establishing the testing procedures in the third step to evaluate the non-inferiority of the test treatment to the active control. For further development, we rewrote (3) and let

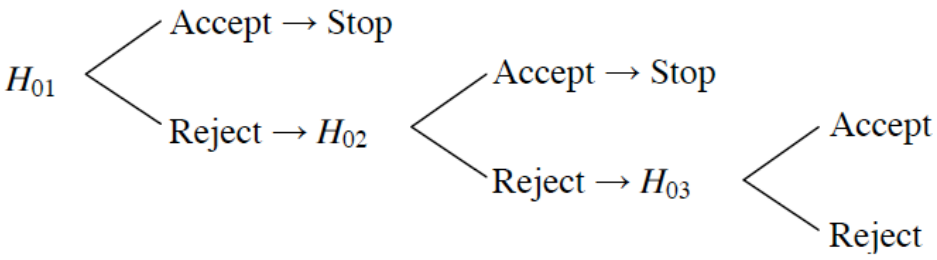

Figure 1: The partitioning hierarchical hypotheses.

$$\psi(\pi) = \pi_T - \theta\pi_C - (1-\theta)\pi_P, \quad (4)$$

which is a linear combination of $\pi = (\pi_T, \pi_C, \pi_P)$. We obtained the maximum likelihood estimator (MLE) $\hat{\pi}_k$ of π_k,

$$\hat{\pi}_k = \frac{X_k}{n_k},\ k = T, C, P.$$

Hence, $\psi(\hat{\pi}) = \hat{\pi}_T - \theta\hat{\pi}_C - (1-\theta)\hat{\pi}_P$ is a best linear unbiased estimator of $\psi(\pi)$. The test statistics T converge in distribution to a standard normal random variable by De Moivre-Laplace theorem [13]. The approximation can be used if $n_k\pi_k$ and n_k $(1-\pi_k)$ are both larger than 5 ($k = T, C, P$). The MLE $\psi(\hat{\pi})$ is asymptotically normally distributed, i.e. $(\psi(\hat{\pi}) - \psi(\pi)) \xrightarrow{D} N(0, \sigma_A^2)$, with variance

$$\sigma_A^2 = \pi_T(1-\pi_T)/n_T + \theta^2\pi_C(1-\pi_C)/n_C + (1-\theta)^2\pi_P(1-\pi_P)/n_P,$$

and the variance can be estimated by

$$\tilde{\sigma}^2 = \tilde{\pi}_T(1-\tilde{\pi}_T)/n_T + \theta^2\tilde{\pi}_C(1-\tilde{\pi}_C)/n_C + (1-\theta)^2\tilde{\pi}_P(1-\tilde{\pi}_P)/n_P$$

where $\tilde{\pi}_k$ estimates π_k under null hypothesis. We evaluate $\tilde{\pi}_k$ in $\tilde{\sigma}^2$ by using two commonly used methods under the null hypothesis (3).

Method I: The value of $\tilde{\pi}_k$ can be estimated by an observed value $\hat{\pi}_k$ [14], but it might be failed in a null hypothesis of non-zero difference between groups [15].

Method II: According to Farrington and Manning [15], the value of $\tilde{\pi}_k$ can be estimated by a restricted maximum likelihood under the null hypothesis restriction $\tilde{\pi}_T = \theta\tilde{\pi}_C + (1-\theta)\tilde{\pi}_P$. A third-degree likelihood equation of the proportion-type rates $\tilde{\pi}_k$ is a problem that can be solved by Miettinen and Nurminen [16]. The derivation is given in the Appendix A.

Therefore, we obtain the Wald statistic

$$T = \frac{\psi(\hat{\pi})}{\tilde{\sigma}} \quad (5)$$

which is asymptotically standard normally distributed for $\psi(\pi) = 0$. Thus, the null hypothesis (3) was rejected if

$$T > z_{1-\alpha} \quad (6)$$

at the one-sided significant level α, where $z_{1-\alpha}$ is the $100(1-\alpha)$ quantile of the standard normal distribution.

2.3. Power and Optimal Sample Size Determinations

In this subsection, we formulate the power function of the Wald test and determined the necessary sample size of test treatment to achieve a desired level of power $1-\beta$. According to Chow *et al*. [12], sample size formulae for Step 1 and 2 are established, respectively (see Appendix B). In Step 3, the power function of the test (6) is given by

$$1-\beta \le P(T > z_{1-\alpha}), \quad (7)$$

and the power of the above inequality (7) is approximately

$$\Phi\left(\frac{\psi(\pi) - z_{1-\alpha}\cdot\tilde{\sigma}}{\sqrt{\sigma_A^2}}\right) \quad (8)$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

Assume that an optimal sample size allocation for the test treatment group, the active control, and the placebo group can be expressed as n_T: n_C: n_P = 1: C_C: C_P. Therefore, according to (8), we determined the sample size based on the following inequality:

$$\left(z_{1-\beta}\cdot\sigma_A + z_{1-\alpha}\cdot\tilde{\sigma}\right)^2 \le \left(\psi(\pi)\right)^2$$

for observing $T \ge z_{1-\alpha}$ with a desired level of power $1-\beta$. This led to a simplification of the procedure where n_T has to be determined as the smallest value fulfilling

$$n_T \ge \left[\left(z_{1-\beta}\sqrt{\pi_T(1-\pi_T) + \theta^2\frac{\pi_C(1-\pi_C)}{C_C} + (1-\theta)^2\frac{\pi_P(1-\pi_P)}{C_P}} + z_{1-\alpha}\sqrt{\tilde{\pi}_T(1-\tilde{\pi}_T) + \theta^2\frac{\tilde{\pi}_C(1-\tilde{\pi}_C)}{C_C} + (1-\theta)^2\frac{\tilde{\pi}_P(1-\tilde{\pi}_P)}{C_P}}\right) \Big/ \psi(\pi)\right]^2 \quad (9)$$

According to Method I, the inequality (9) can be reduce to

$$n_T \ge \left(\frac{z_{1-\beta} + z_{1-\alpha}}{\psi(\pi)}\right)^2\left[\pi_T(1-\pi_T) + \frac{\theta^2\pi_C(1-\pi_C)}{C_C} + \frac{(1-\theta)^2\pi_P(1-\pi_P)}{C_P}\right] \quad (10)$$

In general, the choice of C_C and C_P may be made by clinicians or investigators at the design stage for conducting a clinical trial. Given the values of C_C and C_P, we determined the total sample size N with

$$N = n_T + n_C + n_P = n_T\left(1 + C_C + C_P\right). \quad (11)$$

To determine the minimum of the total sample size N, the optimal values for C_C and C_P are given as partial derivatives of (11) at zero. In Method I, the solutions of C_C and C_P are:

$$C_C = \theta\sqrt{\frac{\pi_C\left(1-\pi_C\right)}{\pi_T\left(1-\pi_T\right)}} \text{ and } C_P = \left(1-\theta\right)\sqrt{\frac{\pi_P\left(1-\pi_P\right)}{\pi_T\left(1-\pi_T\right)}};$$

In Method II, an iterative procedure can be used to solve C_C and C_P in (11) and minimize of the total sample size N (see Appendix A).

We explored the required total sample size N based on Method I and II for different sample size allocation rules, different combination of the design parameters (π_P, π_C, π_T) and θ, and for given α = 0.025, 1 – β = 0.8 (see Table **1**). In Table **1**, we considered four different sample size allocations (balance design, two types of unbalance designs, and our proposed optimal sample size allocation), different choices of θ, (π_P, π_C) = (0.1, 0.8), and $\pi_T = \pi_C$. We found that Method I gives smaller total sample size N than Method II. Farrington and Manning [15] pointed out that Method I, however, suffer serious drawbacks such as underestimate or overestimate the true value of the null variance under the alternative hypothesis thus leading to incorrect sample sizes. Hence, in the following discussion we focus on Method II for precise sample size. When θ = 0.1, the sample sizes of Method II for balance design, 2: 2: 1 design, 3: 2: 1 design and optimal sample size allocation are 23, 27, 27, and 16, respectively. As seen in Table **1**, the sample size of Method II obtained from optimal allocation design is always smaller than that obtained from the other sample size allocation rules. Furthermore, the total sample size increases as θ increases when other design parameters are fixed. This phenomenon is intuitively true since the requirement of the treatment effect is stronger for the larger margin θ; hence, the required total sample size is larger. In Table **2**, we set the margin at θ = 0.6 and 0.8 for four different sample size allocations. The first row is sample size of the treatment group (n_T), while the second row is the total sample size (N). For example, in the first row, sample sizes of Method II of the treatment group n_T for balance design, 2: 2: 1 design, 3: 2: 1 design, and optimal sample size allocation are 29, 31, 36, and 38, respectively. In the second row, the total sample sizes N for four sample size allocation are 85, 77, 71, and 67, respectively. The result of Table **2** is similar to Table **1**. In addition, we find that the total sample size increases as ratio of π_P/π_C increases when other design parameters are fixed. In Table **3**, the corresponding sample size based on the three steps is calculated according to Appendix B and Eq. (11). We find that the required sample sizes per group of Step 1 and Step 2 are substantially smaller than the sample size of Step 3.

In Figure **2**, we illustrate the sample size reductions for using the optimal allocation instead of a balance design, 2: 2: 1 and 3: 2: 1 designs, respectively, given π_P = 0.1, π_C = 0.8 and $\pi_T = \pi_C$. As seen in Figure **2**,

Table 1: Required Total Sample Sizes for Different Sample Size Allocation Designs at Nominal θ, Given π_P = 0.1, π_C = 0.8, $\pi_T = \pi_C$, α = 0.025, β = 0.2

θ	n_T: n_C: n_P							
	1:1:1		2:2:1		3:2:1		1:C_C:C_P	
	Method I	Method II	Method I	Method II	Method I	Method II	Method I	Method II
0.1	14	23	16	27	16	27	10	16
0.2	17	26	18	30	18	30	13	20
0.3	22	32	22	35	21	34	18	25
0.4	30	40	28	42	27	40	25	32
0.5	43	56	40	54	37	51	37	45
0.6	70	85	62	77	58	71	58	67
0.7	132	151	114	131	108	118	106	116
0.8	320	344	270	291	260	264	244	254
0.9	1396	1426	1167	1191	1145	1125	1000	1010

Table 2: Required Sample Sizes Based on Method II for the Treatment Group (First Row) and Total Sample Size (Second Row) with $\pi_T = \pi_C$, $\alpha = 0.025$, $\beta = 0.2$ for Different Sample Size Allocation Designs

θ	(π_P, π_C)	$n_T: n_C: n_P$			
		1:1:1	2:2:1	3:2:1	$1:C_C:C_P$
0.6	(0.1, 0.8)	29	31	36	38
		85	77	71	67
	(0.2, 0.7)	66	72	84	85
		196	179	167	163
	(0.3, 0.6)	199	219	261	262
		597	548	522	515
	(0.4, 0.5)	1856	2048	2462	2463
		5566	5119	4923	4877
0.8	(0.1, 0.8)	115	117	132	139
		344	291	264	254
	(0.2, 0.7)	281	286	336	340
		842	714	672	649
	(0.3, 0.6)	880	898	1078	1052
		2638	2245	2155	2072
	(0.4, 0.5)	8232	8421	10181	9730
		24694	21053	20361	19557

Table 3: Required Sample Sizes Per Group (*T*: Test Treatment; *C*: Active Control; *P*: Placebo) in Three Steps for Different Combination of Design Parameters θ, (π_P, π_C) with $\pi_T = \pi_C$, $\alpha = 0.025$, $\beta = 0.2$ According to the Proposed Optimal Sample Size Allocation Using Method II

θ	(π_P, π_C)	Step 1		Step 2		Step 3		
		T	*P*	*C*	*P*	*T*	*C*	*P*
0.6	(0.1, 0.8)	8	2	6	3	38	16	13
	(0.2, 0.7)	21	8	16	9	85	45	33
	(0.3, 0.6)	70	27	51	31	262	151	102
	(0.4, 0.5)	677	266	485	317	2463	1453	961
0.8	(0.1, 0.8)	13	1	11	2	139	92	23
	(0.2, 0.7)	36	6	30	7	340	248	61
	(0.3, 0.6)	119	23	100	23	1052	820	200
	(0.4, 0.5)	1158	227	966	236	9730	7881	1946

there are at least 20% sample size reductions when the balance design is replaced by optimal allocation. In addition, the sample size reductions are even greater than 30%, where the margins for θ are close to 0.1 and 0.9, respectively. For 3: 2: 1 design, we could save at least 2% sample size by relocating to the optimal sample size allocation.

Figure **3** presents the total sample size for $\theta = 0.5$, 0.6, 0.7, 0.8, and 0.9, with the different proportions of π_P / π_C from 0.125 to 1, given $\pi_C = 0.8$ and $\pi_T = \pi_C$. As seen in Figure **3**, the total sample size increases with increasing values of θ and active control effect π_P / π_C. The total sample size is enormous when π_P / π_C close to one and $\theta = 0.9$, which is impractical in clinical trials.

3. SIMULATION

To examine the performance of proposed optimal sample size allocation in Method II, we conducted a simulation study for the type I error rate and the simulated power. All parameter constellations were simulated with 100,000 replications.

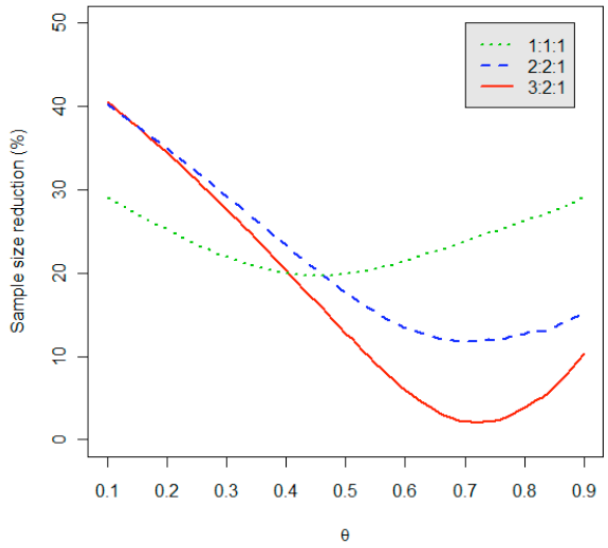

Figure 2: Reduction in total sample size when optimal allocation is used instead of balance design (green dotted line), 2: 2: 1 allocation (blue dashed line) and 3: 2: 1 (red solid line), given $\pi_P = 0.1$, $\pi_C = 0.8$, and $\pi_T = \pi_C$.

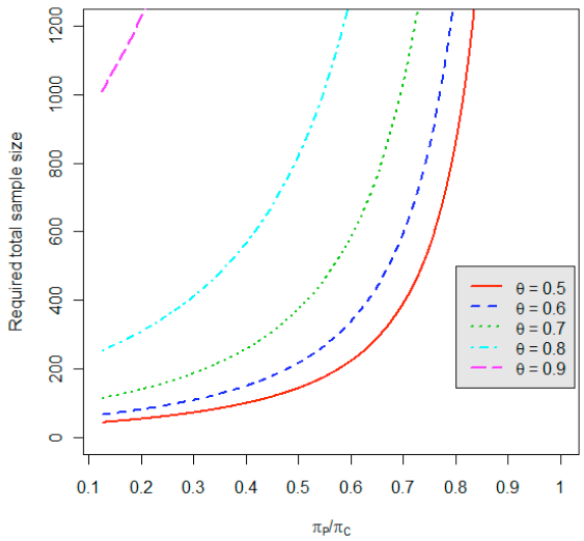

Figure 3: Required total sample size based on optimal sample size allocation design for θ = 0.5, 0.6, …, 0.9 with the different proportions of π_P / π_C from 0.125 to 1, given $\pi_C = 0.8$, and $\pi_T = \pi_C$.

3.1. Type I Error Rate

In order to assess the type I error rate, we set $\pi_T = \theta\pi_C + (1 - \theta)\pi_P$, margin $\theta = 0.8$, and considered $(\pi_P, \pi_C) = (0.1, 0.8)$, (0.2, 0.7), (0.3, 0.6), and (0.4, 0.5) for a nominal significance level of $\alpha = 0.025$. We considered allocation ratios of n_T: n_C: n_P = (a) 1: 1: 1; (b) 2: 2: 1; (c) 3: 2: 1 (d) 1: C_C: C_P with total sample size N (= $n_T + n_C + n_P$) being 60, 150, and 300. Table **4** presents the simulated type I error rate of Wald's test in four different sample size allocation plans. The type I error rate close to $\alpha = 0.025$ as N increases. For $N = 300$, all sample size allocation designs control the type I error rates quite well at the nominal $\alpha = 0.025$. We concluded that the proposed optimal sample size allocation controls type I error very well.

3.2. Power

To assess the power performance of the proposed method, we set $(\pi_T - \pi_P) / (\pi_C - \pi_P) > \theta$, $\theta = 0.8$, $\alpha = 0.025$, and considered $(\pi_P, \pi_C, \pi_T) = (0.1, 0.8, 0.7)$, (0.1, 0.8, 0.71), … , (0.1, 0.8, 0.99) with total sample size N being 300. Similarly, we considered the four allocation ratios as stated above. Figure **4** shows the power curves for the four allocation designs. Apparently, if $\theta = 0.8$, the power curves of unbalanced sample size designs are higher than those of balanced sample size design. The proposed optimal allocation design is especially more powerful than other allocation rules.

4. APPLICATION

4.1. A Papulopustular Acne Study

Papulopustular Acne is a common skin disease characterized by androgenic stimulation of sebaceous glands. Acne is a multifactorial disorder with spontaneous resolution in early adult life. Therefore, combined oral contraceptives (COCs) containing anti-androgenic progestogens are suitable candidates for acne treatment. A multinational, multicenter study was conducted as a three-arm, double-blind and randomized trial for 1326 female patients (16–45 years old) with mild to moderate papulopustular acne, which was discussed in Ernesta *et al*. [17] for the therapy of papulopustular acne of multifactorial disorder. The standard treatment was combined oral contraceptives (COCs) containing potent anti-androgen of ethinylestradiol (EE)/cyproterone acetate (CPA) drug. Ernesta *et al*. [17] showed that a new drug, EE/dienogest (DNG), is superior to a placebo and non-inferior to an active control EE/CPA. As they pointed out, there is no binding affinity between the new drug and sex hormone-binding (SHBG). Furthermore, the new drug does not compete with free testosterone for binding SHBG. Hence, the component testosterone should be decreased to make the estrogen work [17]. In Ernesta *et al*., totally, the 1326 patients were randomly allocated into the three groups, with proportion 2: 2: 1 ($n_T = 525$, $n_C = 537$, $n_P = 264$). After treatment with COCs, the improvement rates of acne were reported as $\pi_T = 91.9\%$, $\pi_C = 90.2\%$ and $\pi_P = 76.2\%$. We applied optimal sample size allocation to this example. As mentioned in Section 2.1 and

Table 4: Simulated Type I Error Rates for Tests at $\pi_T = \theta\pi_C + (1 - \theta)\pi_P$, Given $\theta = 0.8$, and a Nominal Significance Level of $\alpha = 0.025$

N	(π_C, π_P)	n_T: n_C: n_P			
		1:1:1	2:2:1	3:2:1	1:C_C:C_P
60	(0.8, 0.1)	0.0235	0.0250	0.0253	0.0288
	(0.7, 0.2)	0.0235	0.0268	0.0259	0.0278
	(0.6, 0.3)	0.0232	0.0236	0.0230	0.0280
	(0.5, 0.4)	0.0239	0.0252	0.0241	0.0244
150	(0.8, 0.1)	0.0256	0.0250	0.0261	0.0266
	(0.7, 0.2)	0.0249	0.0242	0.0248	0.0268
	(0.6, 0.3)	0.0245	0.0236	0.0244	0.0242
	(0.5, 0.4)	0.0256	0.0251	0.0248	0.0288
300	(0.8, 0.1)	0.0244	0.0242	0.0257	0.0260
	(0.7, 0.2)	0.0251	0.0254	0.0253	0.0248
	(0.6, 0.3)	0.0254	0.0243	0.0245	0.0246
	(0.5, 0.4)	0.0250	0.0241	0.0253	0.0257

observed in Figure **3**, θ is reasonable chosen between 0.5 and 0.8. Thus, given margin $\theta = 0.7$, we only need 836 patients for detect the effect size in this study with $\alpha = 0.025$ and the desired power of 80% using our method. The proposed optimal sample size allocation method reassigned patients into the three groups, with the allocation ratio n_T: n_C: n_P = 1: C_C: C_P = 1: 0.64: 0.41 (n_T = 408, n_C = 261, n_P = 167). Compared to Ernesta *et al*. [17], 97 patients were reassigned to the treatment or active control group from the placebo group utilizing the proposed optimal sample size allocation. According to our method, the results showed DNG was superior to the placebo and non-inferior to CPA at $\alpha = 0.025$ with 80% power in this study. From the economical point of view, the required total sample size was reduced from 1326 to 836. Thus, the proposed optimal sample size allocation is an economically advantageous method for saving 37% of the total sample size.

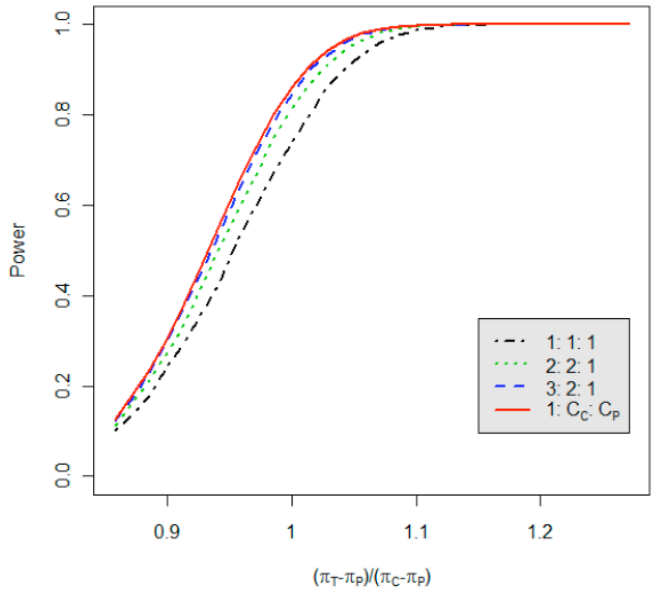

Figure 4: Simulated power for 1: 1: 1 (black dot-dash line), 2: 2: 1 (green dotted line), 3: 2: 1 (blue dashed line) and 1: C_C: C_P (red solid line), respectively for $\theta = 0.8$, $\pi_P = 0.1$, $\pi_C = 0.8$, $\pi_T = 0.7, 0.71, \ldots, 0.99$, $N = 300$, and $\alpha = 0.025$.

4.2. An Oral Prophylactic Antibiotics Study

A large prospective study with 2083 patients, who were completed flexible cystoscopy (FC) from a three-arm placebo controlled trial was conducted to examine whether oral prophylactic antibiotics (ciprofloxacin or trimethoprim) reduce the risk of bacteriuria after FC [18]. A treatment group was treated with ciprofloxacin (500 mg); and an active control group was treated with trimethoprim (200 mg). The sample sizes for the three groups were n_T = 687, n_C = 712, and n_P = 684. The sample size allocation ratio was approximately 1: 1: 1 (balanced design). After FC, the proportions of patients with a negative urine culture were π_T = 97.2%, π_C = 95.4%, π_P = 90.9%. In this example, we applied optimal sample size allocation to let the number of patients in the placebo group be as small as possible. Given margin $\theta = 0.7$, the required total sample size is 1272 for detect the effect size in this study given $\alpha = 0.025$

and the desired power of 80% by our method. The new sample size allocation ratio was n_T: n_C: n_P = 1: 0.68: 0.41 (n_T = 608, n_C = 414, n_P = 250) by optimal sample size allocation. Compared to Johnson *et al*. [18], 434 patients were reassigned to the therapeutic group using our method. As a result, the proposed method using the optimal sample size allocation design was more ethical than the balance design. We concluded that the ciprofloxacin significantly reduced the bacteriuria after FC at α = 0.025 with 80% power in this study. Using optimal sample size allocation, the required total sample size is reduced from 2083 to 1272. Thus, the proposed method saved 61.6% of total sample size compared to Johnson *et al*. [18]. We concluded that the proposed design is economically and ethically better than the balanced design.

5. DISCUSSION

For two-arm non-inferiority trials, issues such as a choice of non-inferiority margin, constancy assumption, and assay sensitivity have been debated for years, and the statistical methodology has been challenged. Given all the issues as discussed, two-arm non-inferiority trials are needed when placebo is not a choice in situation of life threatening or disease progress may be irreversible. Three-arm non-inferiority trials may be a choice in other situations when placebo is acceptable such as in the disease areas of depression, bipolar disorders, and papulopustular acne. In this article, we proposed an optimal sample size allocation design for a three-arm non-inferiority trial when it is ethically justifiable. Moreover, we use restricted maximum likelihood method to correct sample size when the null hypothesis is non-zero between groups because Method I may produce incorrect sample size under null a hypothesis of non-zero difference. The proposed method can substantially reduce the total required number of patients. Furthermore, more patients can be reassigned to the therapeutic group using the proposed design. Thus, our method is not only desirable from an ethical point of view, but also substantially save the total sample size to achieve a certain power.

Our simulation study shows that the optimal sample size allocation design controls type I error rate fairly well in nominal level α for most practical situations. In addition, the proposed design yields a power higher than the other competitive sample size allocation designs in each hypothesis testing. In conclusion, the use of the proposed design is recommended for non-inferiority three-arm trials.

In some clinical trials, more than one primary endpoint could be investigated for efficacy evaluation, which may result in significant complexity in the design, conduct, analysis, and interpretation of data. In addition, testing multiple hypotheses for multiple primary outcomes may increase the FWER. We hope to address this issue by the rationale of optimal sample size allocation in the future.

ACKNOWLEDGEMENT

This research was conducted under the Graduate Program of Biotechnology in Medicine sponsored by the National Tsing Hua University and the National Health Research Institutes. This work was supported by grants from the National Health Research Institutes, Zhunan, Taiwan (PHS-100-PP-11, PHS-101-PP-11, PHS-102-PP-11. “Designs and Evaluations in Non-inferiority Trials”).

APPENDIX A

To derive the value of $\tilde{\pi}_k$, let $L = L(\pi_T; x_T)\cdot L(\pi_C; x_C)\cdot L(\pi_P; x_P)$ be the product of four binomial probabilities under the null hypothesis restriction $\tilde{\pi}_T = \theta\tilde{\pi}_C + (1-\theta)\tilde{\pi}_P$ for Method II. Setting the partial derivative of log-likelihood with respect to π_C to be zero yields the following third-degree likelihood equation:

$$a_1x^3 + b_1x^2 + c_1x + d_1 = 0 \qquad (A.1)$$

with

$$a_1 = \theta^2(1+C_C)$$
$$b_1 = -\theta^2(1+C_C\hat{\pi}_C) - \theta(\hat{\pi}_T + C_C) + \theta(1-\theta)\pi_P(1+2C_C)$$
$$c_1 = \theta(\hat{\pi}_T + C_C\hat{\pi}_C) - \theta(1-\theta)\pi_P(1+2C_C\hat{\pi}_C) - (1-\theta)C_C\pi_P[1-(1-\theta)\pi_P]$$
$$d_1 = (1-\theta)[1-(1-\theta)\pi_P]C_C\hat{\pi}_C\pi_P.$$

Let $\tilde{\pi}_C$ be the unique solution in (0, 1) of Eq. (A.1). The solution is

$$\tilde{\pi}_C = 2u_1\cos(\omega_1) - b_1/(3a_1),$$

where

$$\omega_1 = (1/3)[\pi + \cos^{-1}(\upsilon_1/u_1^3)]$$
$$\upsilon_1 = b_1^3/(3a_1)^3 - b_1c_1/(6a_1^2) + d_1/(2a_1)$$
$$u_1 = \mathrm{sgn}(\upsilon_1)[b_1^2/(3a_1)^2 - c_1/(3a_1)]^{1/2}.$$

Setting the partial derivative of log-likelihood with respect to π_P to be zero yields the following third-degree likelihood equation:

$$a_2x^3 + b_2x^2 + c_2x + d_2 = 0 \qquad \text{(A.2)}$$

with

$$a_2 = (1-\theta)^2(1+C_P)$$
$$b_2 = -(1-\theta)^2(1+C_P\hat{\pi}_P) - (1-\theta)(\hat{\pi}_T + C_P) + \theta(1-\theta)\pi_C(1+2C_P)$$
$$c_2 = (1-\theta)(\hat{\pi}_T + C_P\hat{\pi}_P) - \theta(1-\theta)\pi_C(1+2C_P\hat{\pi}_P) - \theta C_P\pi_C[1-\theta\pi_C]$$
$$d_2 = \theta[1-\theta\pi_C]C_P\hat{\pi}_P\pi_C.$$

Let $\tilde{\pi}_P$ be the unique solution in (0, 1) of Eq. (A.2). The solution is

$$\tilde{\pi}_P = 2u_2\cos(\omega_2) - b_2/(3a_2),$$

where

$$\omega_2 = (1/3)\left[\pi + \cos^{-1}(\upsilon_2/u_2^3)\right]$$
$$\upsilon_2 = b_2^3/(3a_2)^3 - b_2c_2/(6a_2^2) + d_2/(2a_2)$$
$$u_2 = \mathrm{sgn}(\upsilon_2)\left[b_2^2/(3a_2)^2 - c_2/(3a_2)\right]^{1/2},$$

and $\tilde{\pi}_T = \theta\tilde{\pi}_C + (1-\theta)\tilde{\pi}_P$.

APPENDIX B

Sample size formulae for Steps 1 and 2 can be established based on conventional method for given C_P and C_C. According to Chow *et al*. [12], sample size formulae of test treatment ($n_{T.I}$) and the required total sample size (N_I) in Step 1 are derived as follows.

$$n_{T.I} = \left(\frac{z_{1-\beta} + z_{1-\alpha}}{\pi_T - \pi_P}\right)^2\left[\pi_T(1-\pi_T) + \frac{\pi_P(1-\pi_P)}{C_P}\right],$$

$$N_I = n_{T.I} + n_{P.I} = n_{T.I}(1+C_P).$$

In Step 2, we can establish the sample size formulae of active control ($n_{C.I}$) and required total sample size (N_{II}) in this step using the similar idea.

$$n_{C.II} = \left(\frac{z_{1-\beta} + z_{1-\alpha}}{\pi_C - \pi_P}\right)^2\left[\pi_C(1-\pi_C) + \frac{C_C\cdot\pi_P(1-\pi_P)}{C_P}\right],$$

$$N_{II} = n_{C.II} + n_{P.II} = n_{C.II}(1+C_P/C_C).$$

REFERENCES

[1] International Conference on Harmonisation. Guidance on choice of control group and related design and conduct issues in clinical trials (ICH E10), 2000. Available from URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf

[2] Committee for Medicinal Products for Human Use (CHMP). Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99, 2005. Available from URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

[3] Pigeot I, Schäfer J, Röhmel J, Hauschke D. Assessing non-inferiority in a new treatment in a three-arm clinical trial including a placebo. Stat Med 2003; 22: 883-99. http://dx.doi.org/10.1002/sim.1450

[4] Koch A, Röhmel J. Hypothesis testing in the "Gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. J Biopharm Stat 2004; 14: 315-25. http://dx.doi.org/10.1081/BIP-120037182

[5] Hauschke D, Pigeot I. Establishing efficacy of a new experimental treatment in the 'Gold Standard' design. Biom J 2005; 47: 782-6. http://dx.doi.org/10.1002/bimj.200510169

[6] Tang ML, Tang NS. Tests of noninferiority *via* rate difference for three-arm clinical trials with placebo. J Biopharm Stat 2004; 14: 337-47. http://dx.doi.org/10.1081/BIP-120037184

[7] Kieser M, Friede T. Planning and analysis of three-arm non-inferiority trials with binary endpoints. Stat Med 2007; 26: 253-73. http://dx.doi.org/10.1002/sim.2543

[8] Hasler M. Multiple comparisons to both a negative and a positive control. Pharm Stat 2012; 11: 74-81. http://dx.doi.org/10.1002/pst.503

[9] Mielke M, Munk A, Schacht A. Assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints. Stat Med 2008; 27: 5093-110. http://dx.doi.org/10.1002/sim.3348

[10] Kombrink K, Munk A, Friede T. Design and semiparametric analysis of non-inferiority trials with active and placebo control for censored time-to-event data. Stat Med 2013; 32: 3055-66. http://dx.doi.org/10.1002/sim.5769

[11] Koch GG, Tangen CM. Nonparametric analysis of covariance and its role in noninferiority clinical trials. Drug Inf J 1999; 33: 1145-59. http://dij.sagepub.com/content/33/4/1145.short

[12] Chow SC, Shao J, Wang H. Sample size calculations in clinical research. 2nd ed. Boca Raton: Taylor & Francis 2008.

[13] Bickel PJ, Doksum KA. Mathematical statistics: basic ideas and selected topics, 2nd ed. Upper Saddle River, New Jersey: Prentice Hall 2001.

[14] Dunnett CW, Gent M. Significance testing to establish equivalence between treatments with special reference to data in the form of 2 × 2 tables. Biometrics 1977; 33: 593-602. http://dx.doi.org/10.2307/2529457

[15] Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. Stat Med 1990; 9: 1447-54. http://dx.doi.org/10.1002/sim.4780091208

[16] Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med 1985; 4: 213-26. http://dx.doi.org/10.1002/sim.4780040211

[17] Ernesta PK, Schellschmidt I, Schumacher U, Gräser T. Efficacy of a combined oral contraceptive containing 0.030 mg ethinylestradiol/2 mg dienogest for the treatment of papulopustular acne in comparison with placebo and 0.035 mg ethinylestradiol/2 mg cyproterone acetate. Contraception 2009; 79: 282-9. http://dx.doi.org/10.1016/j.contraception.2008.10.010

[18] Johnson MI, Merrilees D, Robson WA, *et al*. Oral ciprofloxacin or trimethoprim reduces bacteriuria after flexible cystoscopy. BJU Int 2007; 100: 826-9. http://dx.doi.org/10.1111/j.1464-410X.2007.07093.x

Received on 31-12-2013 Accepted on 22-01-2014 Published on 31-01-2014

http://dx.doi.org/10.6000/1929-6029.2014.03.01.3

© 2014 Liu *et al*.; Licensee Lifescience Global.
This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.