University of North Carolina at Chapel Hill
The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series

Year 2009 *Paper* 12

Graphical Displays for Clarifying How Allocation Ratio Affects Total Sample Size for the Two Sample Logrank Test

Benjamin R. Saville* Yong S. Kim† Gary G. Koch‡

*Vanderbilt University School of Medicine, b.saville@vanderbilt.edu
†Telik Inc.
‡University of North Carolina at Chapel Hill

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/uncbiostat/art12

Copyright ©2009 by the authors.

Graphical Displays for Clarifying How Allocation Ratio Affects Total Sample Size for the Two Sample Logrank Test

Benjamin R. Saville, Yong S. Kim, and Gary G. Koch

Abstract

For time-to-event data, the power of the two sample logrank test for the comparison of two treatment groups can be greatly influenced by the ratio of the number of patients in each of the treatment groups. Despite the possible loss of power, unequal allocations may be of interest due to a need to collect more data on one of the groups or to considerations related to the acceptability of the treatments to patients. Investigators pursuing such designs may be interested in the cost of the unbalanced design relative to a balanced design with respect to the total number of patients required for the study. We present graphical displays to illustrate the sample size adjustment factor, or ratio of the sample size required by an unequal allocation compared to the sample size required by a balanced allocation, for various survival rates, treatment hazards ratios, and sample size allocation ratios. These graphical displays conveniently summarize information in the literature and provide a useful tool for planning sample sizes for the two sample logrank test.

Graphical Displays for Clarifying How Allocation Ratio Affects Total Sample Size for the Two Sample Logrank Test

Benjamin R. Saville[1], Yong S. Kim[2], Gary G. Koch[3]

[1] *Department of Biostatistics, Vanderbilt University School of Medicine*

S-2323 Medical Center North, 1161 21st Avenue South

Nashville, TN 37232-2158

e-mail: b.saville@vanderbilt.edu

[2] *Telik Inc., 3165 Porter Drive*

Palo Alto, CA 94304

[3] *Department of Biostatistics, University of North Carolina at Chapel Hill*

CB #7420, Chapel Hill, NC 27599

SUMMARY

For time-to-event data, the power of the two sample logrank test for the comparison of two treatment groups can be greatly influenced by the ratio of the number of patients in each of the treatment groups. Despite the possible loss of power, unequal allocations may be of interest due to a need to collect more data on one of the groups or to considerations related to the acceptability of the treatments to patients. Investigators pursuing such designs may be interested in the cost of the unbalanced design relative to a balanced design with respect to the total number of patients required for the study. We present graphical displays to illustrate the sample size adjustment factor, or ratio of the sample size required by an unequal allocation compared to the sample size required by a balanced allocation, for various survival rates, treatment hazards ratios, and sample size allocation ratios. These graphical displays conveniently summarize information in the literature and provide a useful tool for planning sample sizes for the two sample logrank test.

KEYWORDS: Logrank test; time-to-event; sample size; power; unequal allocation

Hosted by The Berkeley Electronic Press

1. Introduction

Many studies with time-to-event data use the two sample logrank test for the comparison of two treatment groups. The design of such a study should have careful planning for the sample size with respect to power and the allocation of patients to the two treatment groups. In some studies, it may be of interest to use an unequal allocation of patients to the respective treatment groups due to a need to collect more data on one of the groups or to considerations related to the acceptability of the treatments to patients. For example, if the control treatment is expected to provide minimal benefit relative to its tolerability, patients might not participate if equal numbers of them have assignment to both treatment groups. Also, investigators may want to increase the number of patients in the test treatment group in order to collect as much safety data as possible on the test treatment. However, an unequal sample size allocation can affect the power of the logrank test for assessing treatment differences.

For unbalanced designs, Pocock [1] preferred a sample size ratio of 2:1 or 3:2 for a test treatment versus a control treatment, and cautioned that a ratio greater than 3:1 could lead to considerable loss of power for the logrank test. Kalish and Harrington [2] advocated the use of balanced designs and argued that equal allocation designs are highly efficient, even with large treatment differences and large amounts of censoring. They conducted simulation studies to show that the power curve of the logrank test is very flat across different hazards ratios (2,4) and degrees of censoring (0, 25, 50, 75%). They noted that the optimal design allocates more patients to the treatment group with the lower event rate and tends to make the total number of events in the treatment groups at any point in time more balanced than if equal allocation had been used.

There are various formulas in the literature for calculating sample size for the logrank test. Hsieh [3] compared the sample size formulas given by Freedman [4] (used in Nquery and STATA 10), Schoenfeld [5], Hsieh [6], and Shuster [7], and found that Freedman's formula gives the highest power for the logrank test if the sample size ratio of the two groups is equal to the reciprocal of the hazards ratio, whereas the other formulas give highest power when sample sizes in the two groups are equal. Hseih also observed that it was difficult to locate the optimal

http://biostats.bepress.com/uncbiostat/art12

sample size ratio for the logrank test due to the flat portion of the power curves, but suggested that the the optimal sample size ratio may exist in the neighborhood from equal allocation of sample sizes to equal allocation of events. This perspective agreed with the assignment of more patients to the treatment group with the lower event rate, and hence with the tendency to balance the number of events between the two groups.

We focus on the sample size formula of Freedman [4] with the objective of evaluating the impact of unequal allocation as it relates to power and sample size. Inherent assumptions of Freedman [4] are that all patients have the same fixed time period of follow-up and that essentially no patient has premature withdrawal, although Freedman provides some informal strategies for addressing departures from this data structure. More importantly, such assumptions are reasonably realistic in practice, particularly for studies in which patients receive treatment and/or require diagnostic procedures for the event of interest throughout their durations of follow-up. Examples include studies for the prevention of unfavorable events such as myocardial infarction, stroke, or death in cardiovascular studies, skeletal fractures in osteoporosis studies, and recurrences (or exacerbations) of major impairments in studies for psychiatric or neurologic disorders (as well as related counterparts for gastrointestinal or respiratory disorders). The method of Freedman [4] is also reasonably applicable to vaccine studies with intensive follow-up for identifying new cases of a disease which is to have prevention and to studies for reducing the time to occurrence of favorable events during a treatment period, such as relief of symptoms for migraine headache or the healing of an infection or a gastrointestinal disorder.

In studies like those previously described, it may be of interest to use an unequal sample size allocation even though it may not lead to better power. If the investigators wish to maintain a specific level of power for assessing comparisons (e.g. 0.90), an unequal sample size allocation may require a greater total number of patients compared to that required for a balanced design. In these situations, it may not be clear to the investigators how the allocation ratio affects the required total number of patients relative to a balanced design, and how this relationship differs depending on the hazards ratio and event rate. In this manuscript, we present graphical displays to clarify the impact of an unequal allocation design in terms of the

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

sample size adjustment factor. This factor is defined as the ratio of the sample size required by the unequal allocation compared to the sample size required by the balanced allocation, under the same specifications for power and type I error across the two allocations. These graphical displays conveniently summarize information in the literature and provide a useful tool for planning sample sizes for the two sample logrank test for time-to-event data.

2. The effect of allocation ratio on total sample size

We first summarize the sample size formula given by Freedman [4]. Let P_1 and P_2 denote the survival rates for patients in groups 1 and 2, which typically correspond to test and control treatments, respectively. A hazards ratio (constant over time) for comparing the two groups can be expressed as

$$\theta = \frac{\log_e(P_1)}{\log_e(P_2)}. \tag{1}$$

A hazards ratio greater than one indicates more events (and lower survival rates) for patients in group 1 versus group 2, and a hazards ratio less than one indicates fewer events (and higher survival rates) for patients in group 1 versus group 2. For a study with equal sample sizes for each group, Freedman [4] indicates that the total number of events d required for $(1-\beta)$ power at the one-sided significance level α can be approximated by

$$d = (z_\alpha + z_\beta)^2 \left(\frac{1+\theta}{1-\theta}\right)^2, \tag{2}$$

in which z_α and z_β are the standard normal quantiles corresponding to the one-sided significance level α and required power $(1-\beta)$ in the logrank test. This approximation assumes that the ratio of patients at risk in the two groups before each event is equal to one and that the quantity $2\sqrt{\theta}/(\theta+1)$ is approximately equal to one. It follows that the total number of patients required in the study can be approximated by

$$N = 2d/(2 - P_1 - P_2). \tag{3}$$

For unequal allocation in the number of patients in the two groups, say $\phi : 1$ for group 1 versus group 2, the total number of events is approximated by

$$d_\phi = \frac{(z_\alpha + z_\beta)^2(1+\theta\phi)^2}{\phi(1-\theta)^2}. \tag{4}$$

http://biostats.bepress.com/uncbiostat/art12

It follows that the total number of patients required under unequal allocation can be approximated by

$$N_\phi = \frac{d_\phi(1+\phi)}{\phi(1-P_1)+(1-P_2)}. \tag{5}$$

The approximation for the unequal allocation assumes that the ratio of patients at risk in the two groups before each event is equal to ϕ and that the quantity $\sqrt{\theta}(1+\phi)/(1+\theta\phi)$ is approximately equal to one.

To simplify notation, we set $P = P_2$, and observe from equation (1) that $P_1 = P^\theta$. For $0 \leq P < 1$, the sample size adjustment factor, or ratio of total sample size for an unequal allocation relative to 1:1 allocation, can be expressed as

$$\frac{N_\phi}{N} = \frac{(1+\theta\phi)^2(1+\phi)(2-P-P^\theta)}{2\phi\{\phi(1-P^\theta)+(1-P)\}(1+\theta)^2} = f(\theta,\phi,P). \tag{6}$$

Note that the standard normal deviates z_α and z_β are not in this ratio, so that the sample size adjustment factor is only a function of θ, ϕ, and P (assuming power and type I error do not vary across the two allocations).

2.1. Graphical displays

Figure 1 shows the sample size adjustment factor relative to a 1:1 allocation for various combinations of the survival rate in group 2 (P), the hazards ratio (θ), and common choices of the allocation ratio (ϕ). We only present results for $\theta < 1$, as sample size ratios for $\theta > 1$ can be obtained by noting that such ratios are the same as the corresponding ratios with hazards ratio $1/\theta$, allocation ratio $1/\phi$ and survival rate P^θ (i.e. $f(\theta,\phi,P) = f(1/\theta, 1/\phi, P^\theta)$). From this perspective, one can note that $(1/\theta) = \log_e(P)/\log_e(P^\theta) = \log_e(P_2)/\log_e(P_1)$ with $P_1 = P^\theta$ serving as P in (6). For example, to obtain the sample size adjustment factor for $\theta = 3$ with allocation $2:1$ and a group 2 survival rate of 0.2, one can look at the value corresponding to hazards ratio $1/3$, allocation $1:2$, and survival rate $P^3 = 0.2^3$.

We are particularly interested in situations in which the ratio (N_ϕ/N) is close to one or less than one, indicating that the unequal allocation does not cost the researcher much in terms of total sample size. For allocation ratios with $\phi < 1$ (indicating more patients on the treatment with the higher event rate when $\theta < 1$), the sample size adjustment factor is greater than one

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

for all combinations of P and θ (note only $\phi = 0.5$ is shown). As the hazards ratio gets closer to one, the sample size adjustment factor decreases and draws nearer to one.

For allocation ratios with $\phi > 1$ (indicating more patients on the treatment with the lower event rate), many of the sample size adjustment factors are less than one. More specifically, for small P and small θ, sample size adjustment factors are typically much less than 1 but increase as P or θ increase, although the approximations in Freedman [4] are possibly less accurate for small P. For $\phi \leq 1/\theta$, or allocation ratios less than the reciprocal of the hazards ratio, the sample size adjustment factor is less than one for all values of P. For $\phi > 1/\theta$, or allocation ratios greater than the reciprocal of the hazards ratio, the sample size adjustment factor is less than 1 for either no values of P or only for smaller values of P. For values of P close to 1, the sample size adjustment factor is close to 1 for $\phi = 1/\theta$, indicating that the optimal allocation in terms of total sample size is somewhere between $\phi = 1$ and $\phi = 1/\theta$. This tendency seems to disagree with the claim of Hsieh [3] that the logrank test predicts highest power when the sample size allocation ratio is equal to the reciprocal of the hazards ratio. For smaller values of P (e.g. $P = 0.5$), the adjustment factor appears to be rather flat across allocation ratios for $\phi > 1$, but still may not agree with the claim of Hsieh [3]. For example, for $P = 0.7$ and $\theta = 0.4$, the sample size adjustment factor is 0.94 for $\phi = 1.5$, 0.95 for $\phi = 2$, 0.98 for $\phi = 2.5$, and > 1 for $\phi \geq 3$, suggesting that $\phi = 1.5$ is a more efficient sample size allocation than the reciprocal of the hazards ratio, or $\phi = 2.5$.

3. Discussion

As noted previously, the approximations of Freedman rely on two approximations. The first is that the ratio of patients at risk in the two groups before each event is equal to ϕ, and the second is that the quantity $\sqrt{\theta}(1+\phi)/(1+\theta\phi)$ is approximately equal to one. Freedman [4] conducted Monte-Carlo simulations studies to verify the accuracy of these sample size approximations. However, given that the second approximation appears to be a computational convenience, one can derive the sample size calculations more exactly by not invoking this approximation. In the general case, it can be shown that the sample size invoking only the first approximation

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uncbiostat/art12

can be approximated by

$$d_\phi^* = \frac{\{z_\alpha + z_\beta\sqrt{\theta}(1+\phi)/(1+\theta\phi)\}^2(1+\theta\phi)^2}{\phi(1-\theta)^2}. \tag{7}$$

It is straightforward to show that the sample size adjustment factor for an unequal allocation relative to a 1:1 allocation is given by

$$\frac{N_\phi^*}{N^*} = \left(\frac{z_\phi^*}{z^*}\right)\left(\frac{N_\phi}{N}\right), \tag{8}$$

in which $z^* = \{z_\alpha + 2z_\beta\sqrt{\theta}/(1+\theta)\}^2$ and $z_\phi^* = \{z_\alpha + z_\beta\sqrt{\theta}(1+\phi)/(1+\theta\phi)\}^2$. The modified sample size adjustment factor N_ϕ^*/N^* depends on the power and type I error and is equal to the original sample size adjustment factor times the quantity (z_ϕ^*/z^*). Values of (z_ϕ^*/z^*) greater than 1 lead to larger estimates of the sample size adjustment factor relative to (6), and values less than 1 lead to smaller estimates of the sample size adjustment factor relative to (6). Holding the one-sided type I error constant at 0.025 (or two-sided 0.05), we found that for $\phi < 1$, the quantity (z_ϕ^*/z^*) is greater than one for power$<$ .50 and less than 1 for power$>$ .50. For $\phi > 1$, the quantity (z_ϕ^*/z^*) is less than one for power$<$ .50 and greater than 1 for power$>$ .50. The magnitude of the increase or decrease is primarily driven by the ratio (z_β/z_α).

Table I gives the values of the quantity (z_ϕ^*/z^*) for various combinations of power, allocation ratio, and hazards ratios. For hazards ratios close to 1, the quantity (z_ϕ^*/z^*) is close to one, suggesting that the absence of the second approximation of Freedman does not greatly influence the sample size adjustment factor. For $\phi > 1$, a decreasing hazards ratio and increasing power are associated with larger ratios, indicating less benefit for the unequal allocation relative to the balanced allocation. For $\phi < 1$, a decreasing hazards ratio and increasing power are associated with smaller ratios, indicating less penalty for the unequal allocation relative to the balanced allocation. Figure 2 gives the modified sample size adjustment factor for $\phi = (0.5, 1.5, 2, 3)$, one sided $\alpha = 0.025$ (or two-sided $\alpha = 0.05$), and power 0.90. One can see that the benefit of the unequal allocation for $\phi > 1$ has diminished in this figure relative to the benefit shown in Figure 1. Hence, if one does not invoke the second approximation of Freedman [4], one obtains a more conservative estimate of the benefit for $\phi > 1$ (allocation ratio) relative to a balanced design. However, given that Freedman performed Monte Carlo simulations to verify the accuracy of

Hosted by The Berkeley Electronic Press

his sample size formulas using both approximations, these estimates may be an artifact of invoking the first approximation of Freedman in the absence of the second approximation, resulting in more conservative estimates of the sample size adjustment factor (see Figure 2).

As shown in Figure 1, the allocation of more patients to the treatment with the greater event rate leads to an increase in the total number of patients required relative to a 1:1 allocation. Hence, unbalanced designs can only be cost-effective in terms of total sample size when more patients are allocated to the treatment with the lower event rate, as such designs can lead to a total sample size that is smaller or not much larger than that required by the balanced design. For designs with more patients allocated to the treatment with a lower event rate, it appears that the optimal allocation is between one and the reciprocal of the hazards ratio, but this depends on the survival rates of the treatment groups.

Given that the approximation of Freedman [4] has potential limitations for high event rates (i.e. P near 0), one should mainly use these graphs for assessing the sample size adjustment factor for larger values of P, or studies with low event rates and a fixed time period for follow-up (e.g. studies of treatments to prevent rare events). However, the associations for smaller P are given in Figures 1-2 to provide further understanding with respect to the association between the adjustment factor, survival rate, hazards ratio, and the allocation ratio. Figure 1 reinforces the idea that the optimal sample size allocation for the two sample logrank test may exist in the neighborhood from equal allocation of sample sizes to equal allocation of events, as the number of events in each group would tend to be closer when the allocation ratio is equal to the reciprocal of the hazards ratio (for $\phi > 1$).

The fact that the sample size adjustment factors are generally not less than 0.9 for moderate to large P reinforces the view of Kalish [2] that the balanced design is highly efficient. However, in circumstances in which there are strong practical reasons to use an unequal allocation, Figure 1 provides a useful tool for assesssing the impact of a desired allocation ratio with respect to the cost of total sample size.

http://biostats.bepress.com/uncbiostat/art12

Acknowledgments

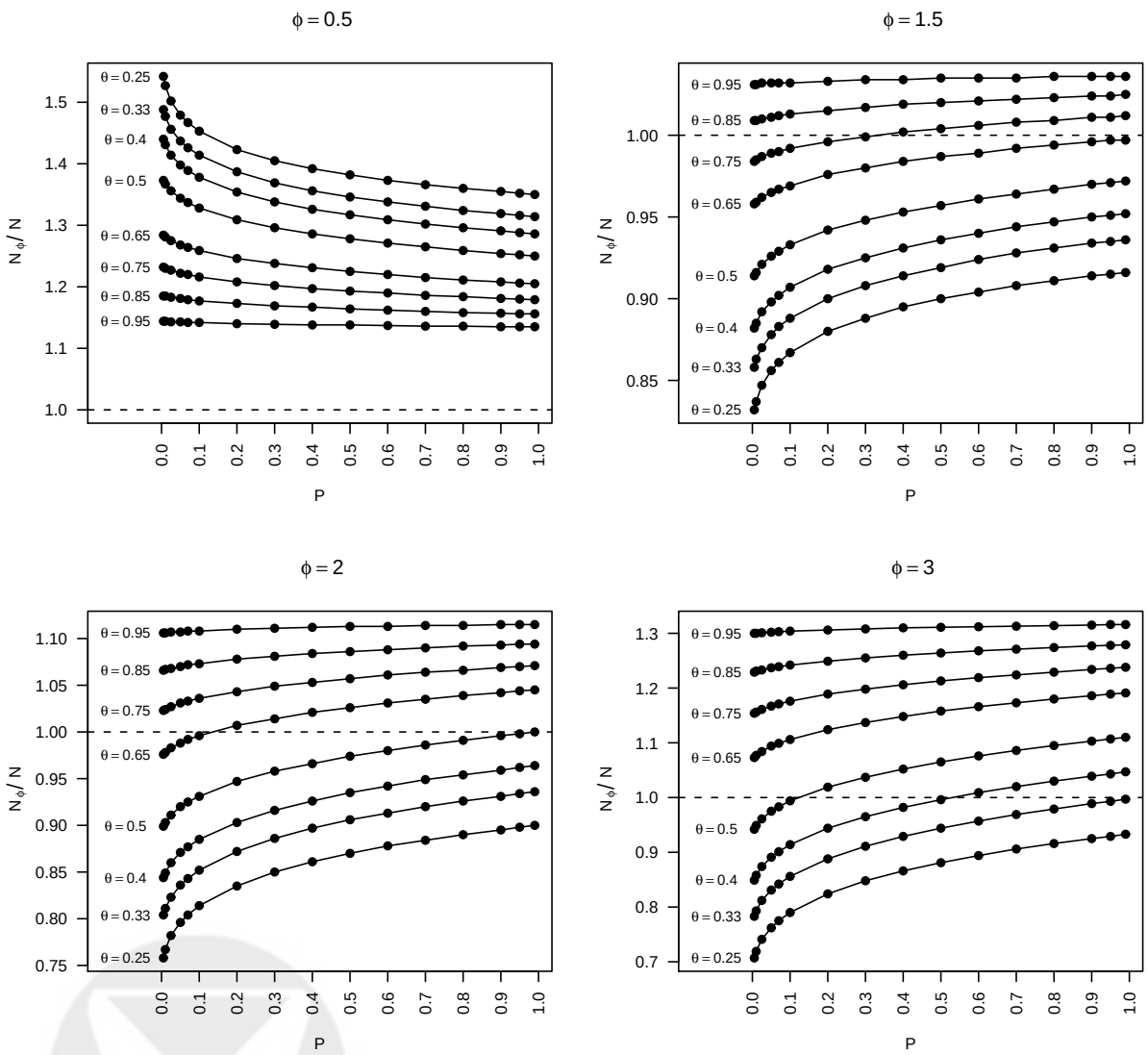
We would like to acknowledge Mike Hussey from the Department of Biostatistics at the University of North Carolina for his helpful review of the manuscript and assistance with computer software.

REFERENCES

1. Pocock SJ. *Clinical Trials*. Wiley: Chichester, UK, 1983.
2. Kalish LA, Harrington DP. Efficiency of balanced treatment allocation for survival analysis. *Biometrics* 1988; **44**:815–821.
3. Hseih FY. Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Statistics in Medicine* 1992; **11**:1091–1098.
4. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1982; **1**:121–129.
5. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 1981; **68**:316–319.
6. Hsieh FY. A simple method of sample size calculation for unequal-sample-size designs that use the logrank or *t*-test. *Statistics in Medicine* 1987; **6**:577–581.
7. Shuster J. *Handbook of Sample Size Guidlines for Clinical Trials*. CRC Press, Inc: West Palm Beach, FL, 1990.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

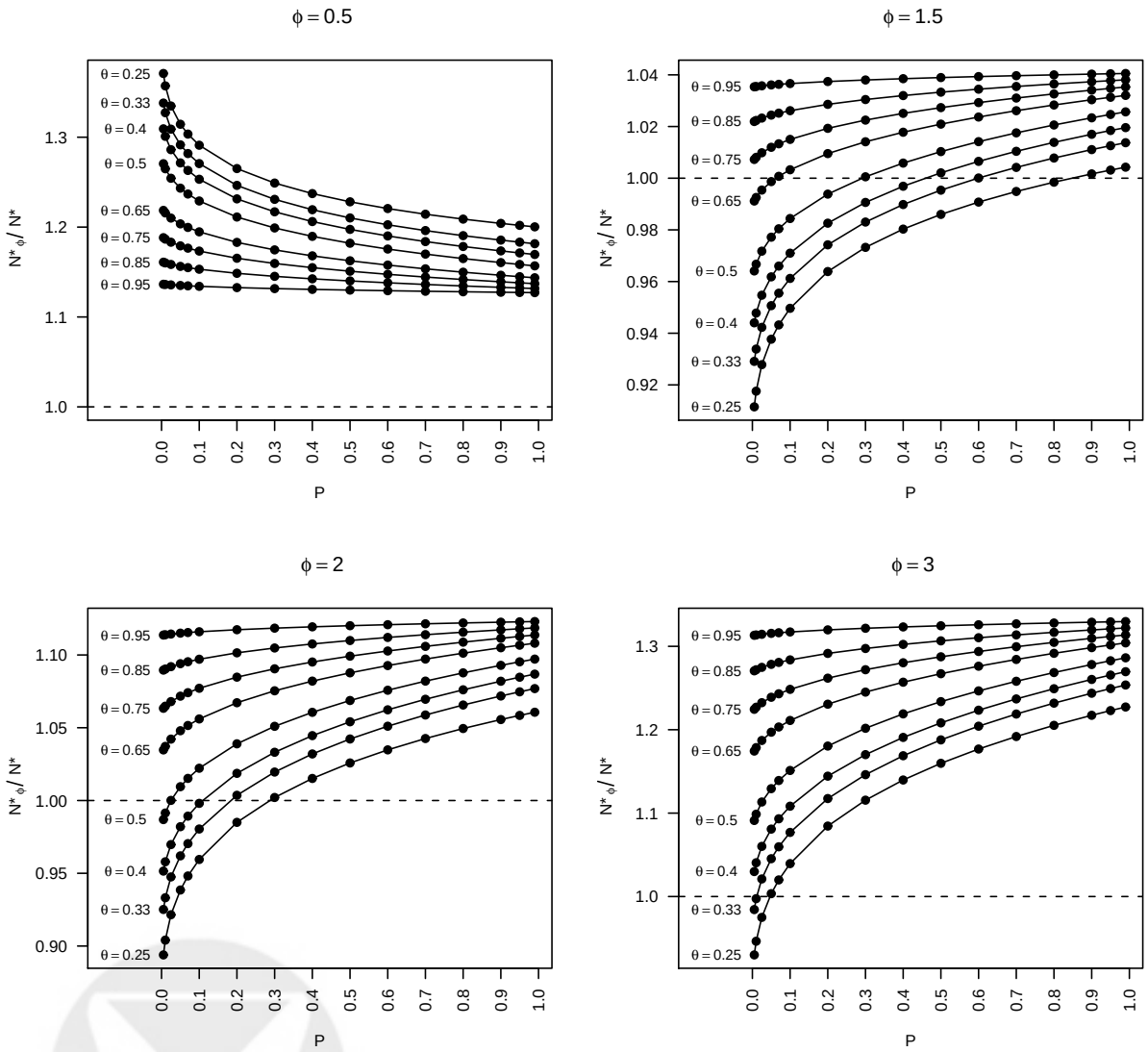


ϕ = Allocation ratio
P = Proportion without event for group 2
N ϕ / N = Sample size adjustment factor

Figure 1. Sample size adjustment factor relative to 1:1 allocation, for any α and Power

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uncbiostat/art12



ϕ = Allocation ratio
P = Proportion without event for group 2
N*ϕ / N* = Modified sample size adjustment factor

Figure 2. Modified sample size adjustment factor relative to 1:1 allocation, one-sided $\alpha = 0.025$ (or two-sided $\alpha = 0.05$), Power = 0.90

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Table I. Values of the ratio z^*_ϕ / z^*

Parameters		Values of θ							
ϕ	Power	0.25	0.33	0.44	0.50	0.66	0.75	0.85	0.95
0.5	0.8	0.92	0.92	0.93	0.94	0.96	0.97	0.98	0.99
	0.85	0.9	0.91	0.92	0.93	0.96	0.97	0.98	0.99
	0.9	0.89	0.9	0.91	0.93	0.95	0.96	0.98	0.99
	0.95	0.87	0.88	0.9	0.91	0.94	0.96	0.98	0.99
	0.99	0.84	0.86	0.87	0.9	0.93	0.95	0.97	0.99
1.5	0.8	1.07	1.06	1.05	1.04	1.03	1.02	1.01	1
	0.85	1.08	1.07	1.06	1.05	1.03	1.02	1.01	1
	0.9	1.1	1.08	1.07	1.06	1.03	1.02	1.01	1
	0.95	1.11	1.1	1.08	1.06	1.04	1.03	1.02	1
	0.99	1.14	1.12	1.1	1.08	1.05	1.03	1.02	1.01
2	0.8	1.13	1.11	1.1	1.07	1.05	1.03	1.02	1.01
	0.85	1.15	1.13	1.11	1.08	1.05	1.03	1.02	1.01
	0.9	1.18	1.15	1.13	1.1	1.06	1.04	1.02	1.01
	0.95	1.21	1.18	1.15	1.11	1.07	1.05	1.03	1.01
	0.99	1.26	1.21	1.18	1.14	1.08	1.05	1.03	1.01
3	0.8	1.23	1.19	1.16	1.12	1.07	1.05	1.03	1.01
	0.85	1.27	1.22	1.18	1.14	1.08	1.05	1.03	1.01
	0.9	1.32	1.26	1.21	1.16	1.09	1.06	1.03	1.01
	0.95	1.37	1.3	1.25	1.18	1.11	1.07	1.04	1.01
	0.99	1.46	1.37	1.3	1.22	1.13	1.08	1.05	1.01

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uncbiostat/art12