Statistical Inference for the Randomized Play the Winner Design

by

David C. Tolusso

A Thesis Submitted to
the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

Statistical Inference for the Randomized Play the Winner Design

BY

David C. Tolusso

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

MASTER OF SCIENCE

Abstract

In clinical trials with extreme outcomes, it is ethically desirable to treat as many patients as possible with the superior treatment. Adaptive designs accomplish this while still producing statistically meaningful results. One such design is known as the randomized play the winner rule (RPWR). Two asymptotic methods for constructing confidence intervals based on data from the RPWR are presented and compared by simulation. It is found that both methods perform well for small sample sizes despite being approximate methods. Some other aspects of the RPWR are examined, such as the rate of convergence of a martingale central limit theorem and some appealing properties of the allocation probabilities.

Acknowledgements

My sincerest gratitude and appreciation must go to my supervisor Dr. Xikui Wang. He first introduced me to research in statistics while I held an NSERC undergraduate student research award. If it were not for him, this thesis would not be possible. Even when my research seemed like it was not leading anywhere he kept encouraging me. His comments and suggestions were greatly appreciated.

Thanks also go to all my teachers and instructors who have taught me throughout my education. In particular I wish to thank Dr. Dennis Murphy for taking the time to teach R to the graduate students in the department. This was very helpful for the simulations required for my thesis. A thank you also goes to Dr. Dean Slonowsky for pushing me to apply for NSERC awards and all his help during the application process. I also wish to thank my graduate committee members Dr. Jin Zhang (Department of Statistics) and Dr. Lisa Lix (Department of Community Health Sciences) for their time reading my thesis.

Finally, I wish to thank those who provided me with financial assistance during my time as an undergraduate and graduate student. Thank you to Dr. Xikui Wang, Dr. Dean Slonowsky, NSERC, the Faculty of Science, the University of Manitoba and the University of Manitoba Students Union.

Contents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Whenever a new medical treatment is developed, it is always necessary to determine its effectiveness. The most common method is to conduct clinical trials. In such trials, the new treatment is compared with a standard treatment or a placebo. The object is to determine if the new treatment is more effective than methods currently in use, or if it is effective at all.

In many cases, there is a simple way to conduct a clinical trial. Take all patients recruited for the study and randomly assign half to the new treatment while the rest receive the control treatment (or placebo). The statistical analysis of such a design is straightforward since this type of randomization allows us to consider the two groups of patients as coming from two different populations with all the observations being independent of each other. Ana-

lyzing data from two independent populations is a problem which has been around for quite some time.

Often in clinical trials, we only observe whether or not a treatment was effective or not. Hence, our data consists of indicator variables (either "success" or "failure") *i.e.* we have binomial responses. There are many different ways to analyze data resulting from two independent binomial populations, see section 1.3.

Assigning half the patients to each treatment does have one important effect. If there is a difference between the two treatments, then half of the patients are receiving an inferior treatment. From an ethical standpoint, it is desirable to have as many patients on the superior treatment, but we do not know ahead of time which treatment is superior. This does not pose a big problem when the situation is not potentially life threatening. For example, if a comparison is to be made between two drugs for treating a headache, the fact that one group of patients will be receiving an inferior treatment, and therefore may have a higher occurence of headaches is a very small price to pay.

However, consider an extreme situation where the response to the treatments is either the patient survives or dies. In this case, having half the patients on a treatment which has a higher mortality rate is ethically infeasible. Alternative methods of assigning patients to the treatments are needed.

This is where adaptive designs enter. An adaptive design is such that the

patients are treated sequentially, and when a patient is ready to be assigned a treatment, this assignment is dependent on the outcomes of the previous patients. The objective is to maximize the number of patients treated successfully while still obtaining statistically meaningful results. There have been many proposed adaptive designs, some of which are deterministic, while others incorporate randomization. Obviously, from a statistical point of view, the deterministic methods are not reasonable because of the possibility of selection bias.

Many of deterministic methods can be modified to incorporate randomization in some way. This is the case with one particular adaptive design, which will be the focal point of what follows.

1.2 The Randomized Play the Winner Rule

This type of design has its roots in the deterministic play the winner rule of Zelen [39]. It is based on the idea of having an urn containing balls representing each treatment. A ball is randomly drawn and the treatment which it corresponds with is assigned to the next patient. Once a response is obtained, additional balls are added to the urn according to the specified rules of the design.

The design of Zelen was such that balls were drawn *without* replacement, and as a consequence, Wei and Durham [37] noted that this rule assigned patients in approximately equal numbers to each treatment. Wei and

1.2 The Randomized Play the Winner Rule

Durham modified Zelen's rule and created the randomized play the winner rule (RPWR).

For conducting a clinical trial using the RPWR, we have an urn as described above. Initially, the urn is composed of α balls of each type. Balls are drawn *with* replacement. Suppose the ball drawn is of type i, ($i = A$ or B). If the resulting response is a success, then β balls of type i are added to the urn. Otherwise, β balls of the other type are added to the urn.

If there were any prior knowledge about the two treatments under study, this could be incorporated into the design by letting the urn initially contain α_A balls of type A and α_B balls of type B. However, it is often the case that we assume clinical equipose at the onset of the trial, that is there is no reason to initially prefer one treatment over the other, and take $\alpha_A = \alpha_B = \alpha$.

Another aspect of the RPWR is its ability to handle delayed responses as well. If a patient is ready to be assigned a treatment but previous patients responses have not yet been observed, then a ball can still be drawn from the urn with its current composition. Incorporating delayed responses increases the complexity of the analysis, so it will be assumed from here on that the previous patients responses are all available before the current patient is to be treated.

The RPWR has ethically desirable properties compared with traditional randomization. The RPWR assigns more patients to the better treatment and the RPWR has a higher total success rate. These help contribute to the RPWR being a recommended design where outcomes are life threatening.

1.2 The Randomized Play the Winner Rule

The RPWR has been used in practice. Bartlett et al. [8] describes its use in the ECMO trial. A new treatment known as extracorporeal membrane oxygenation (ECMO) was compared with the conventional treatment for infants with severe respiratory failure. This particular trial had 12 subjects, but only 10 of which were assigned treatments via the RPWR with $\alpha = \beta = 1$. Nine patients were assigned ECMO and survived, while the one patient which received conventional therapy died.

Statistical analysis of the ECMO trial can be found in Wei [36], Begg [10] and Wei et al. [38]. There was controversy over how to properly analyze such data, particularly in the discussion following the article by Begg. It is also arguable as to how successful the ECMO trial was, in large part due to only one patient being assigned to the conventional treatment. It appears as though the ECMO trial was a case where the two treatments were too unbalanced, so a meaningful comparison could not be made. In retrospect, this trial should have been run with a larger value of α so enough patients could have been assigned to the conventional treatment, and a meaningful comparison could be made.

The RPWR was used on one other clinical trial as well. Tamura et al. [33] describes its use in comparing fluoxetine with a placebo for treating depressive disorder. This was a much more complex design. There were two strata, and in each stratum the first six patients were assigned treatments by permuted block randomization, while the following patients were randomized by the RPWR with $\alpha = \beta = 1$. The end result was that fluoxetine had no

significant effect. This fact was also noticeable in the allocation proportions. In one stratum, 23 patients were assigned to fluoxetine and 22 were given the placebo, while in the other stratum, 23 patients were assigned fluoxetine and 21 were given the placebo. When the treatments have no difference, the allocation proportions are similar, as seen in this case.

1.3 The Case of Two Independent Samples

Before proceeding to how to analyze data arising from the RPWR, inference for two independent binomial samples will be reviewed, specifically dealing with constructing confidence intervals. There are two main approaches to this problem. The first is to use "exact" techniques, which means the exact distribution of a particular test statistic. The second approach is to use asymptotic approximations. Both methods are for constructing confidence intervals for $\Delta = p_A - p_B$, where p_A and p_B are the success probabilities of the two populations.

1.3.1 "Exact" Confidence Intervals

Let T be a discrete statistic for which large values support larger values of Δ, and smaller values support smaller values of Δ. For each Δ_0, let $A(\Delta_0)$ be the acceptance region for testing $H_0 : \Delta = \Delta_0$ against $H_a : \Delta \neq \Delta_0$ at level γ. For each possible value of T, let $C(t) = \{\Delta_0 : t \in A(\Delta_0)\}$. It then follows that $C(t)$ is a $1 - \gamma$ confidence set for Δ. There are several ways to

choose $A(\Delta_0)$.

One method is known as the tail method. Essentially, it corresponds to inverting two one-sided tests, each of size $\gamma/2$. The confidence limits, Δ_L and Δ_U are found by solving

$$\mathbb{P}(T \leq t_0; \Delta_U) = \frac{\gamma}{2} \quad \text{and} \quad \mathbb{P}(T \geq t_0; \Delta_L) = \frac{\gamma}{2} \tag{1.1}$$

where t_0 is the observed value of T. This method can be quite conservative.

An alternative is to invert one two-sided test. For this method, we enter values into $A(\Delta_0)$ in order of their probability under H_0, beginning with the highest, and stopping when the total probability is at least $1 - \gamma$, and then finding $C(t_0)$. Another way of creating $A(\Delta_0)$ is to order them according to a statistic such as the likelihood ratio, Wald, or score statistics. These methods are not as conservative, but do not guarantee intervals and can be computationally difficult.

There is also the problem of nuisance parameters. Two approaches are to use a procedure conditional on the sufficient statistic of the nuisance parameter, or to redefine the p-value as the maximum p-value over the range of the nuisance parameter.

Possible choices for T are,

$$T = \hat{p}_A - \hat{p}_B \tag{1.2}$$

or

$$T = \frac{(\hat{p}_A - \hat{p}_B) - \Delta_0}{\sqrt{\frac{\tilde{p}_A(1-\tilde{p}_A)}{n_A} + \frac{\tilde{p}_B(1-\tilde{p}_B)}{n_B}}} \tag{1.3}$$

where $\hat{p}_A$ and $\hat{p}_B$ are the MLE's of p_A and p_B, $\tilde{p}_A$ and $\tilde{p}_B$ are the MLE's of p_A and p_B subject to $p_A - p_B = \Delta_0$ and n_A and n_B are the samples sizes from each population. See Agresti [1] for more details.

1.3.2 Asymptotic Intervals

The asymptotic intervals presented in this subsection are all described in detail in Beal [9]. These intervals are all based on the asymptotic normality of the MLE's of p_A and p_B. The simplest is to use the method which would be seen in many introductory statistics courses, *i.e.*

$$(\hat{p}_A - \hat{p}_B) \pm z_{\gamma/2}\sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}} \tag{1.4}$$

where $q_A = 1 - p_A$, $q_B = 1 - p_B$ and $z_{\gamma/2}$ is the upper $\gamma/2$ quantile of the standard normal distribution.

This interval can be derived by solving

$$(\Delta - \hat{\Delta})^2 = c\mathbb{V}\mathrm{ar}(\hat{\Delta}; p_A, p_B) \tag{1.5}$$

and replacing p_A and p_B by their MLE's. Different intervals can be obtained

by replacing p_A and p_B with different estimates.

When p_A and p_B are estimated by their MLE's under the restriction that $p_A - p_B = \Delta$ we have the method due to Mee, while a method due to Miettinen and Nurminen is essentially the same with a minor modification. These intervals require numerical methods to find the endpoints of the confidence interval, so computationally, these do not pose a huge advantage over the exact methods.

Two similar methods due to Beal are the Haldane and Jeffreys-Perks intervals. These are both based on a similar idea. The Jeffreys-Perks method will be described in detail in section 2.3.2. Both intervals have closed form solutions which make their computation much simpler than the other methods. Again, the full details of the five asymptotic methods named here can be found in Beal [9].

1.4 Inference for the RPWR

Before describing methods of interval estimation, some tests will be described for testing if there is a difference between the two treatments. The first such test is due to Wei [36]. He constructed a permutation test based on S_A, the total number of successes on treatment A. He described an algorithm for obtaining the exact distribution of S_A, conditional on the responses of the patients.

Wei used this procedure on the ECMO data, where the observed value of

S_A was 11.[†] It was found that the p-value was $\mathbb{P}(S_A \geq 11) = 0.051$.

This approach was criticized by Begg [10]. A main argument was that Wei's p-value included as an extreme point the outcome where all patients are assigned to ECMO, yet this outcome gives no information. The true p-value using Wei's method should be 0.038. Begg also objected to the way in which Wei's test was constructed. He compared several tests with different summary statistics and conditioning on sufficient statistics of the nuisance parameter. He found that a test using $T = \hat{p}_A - \hat{p}_B$, where $\hat{p}_A$ and $\hat{p}_B$ are the same as in the case of two independent samples (see section 2.2.3), and conditioning on S, the total number of successes, performed best.

In the discussion following Begg's paper, Royall showed how p-values ranging from 0.003 to 1 can be obtained for the ECMO data using various approaches. Wei also featured in the discussion, where he supported a test based on $T = \hat{p}_A - \hat{p}_B$.

Wei et al. [38] turned the focus towards confidence intervals. He showed how the exact distribution of the sufficient statistics (see section 2.2.3) can be obtained. Based on this distribution, a conditional interval can be constructed, however this interval does not perform very well.

A better interval was the unconditional interval, which is essentially the tail method for the RPWR. One major drawback to this method is that the calculations required are quite intensive. This method is not very simple to

[†]Wei included two more patients who were assigned ECMO and survived, however these patients were not assigned by the RPWR.

implement.

In the same article, asymptotic intervals were discussed. The asymptotic normality of the MLE's was established, and hence the usual asymptotic interval could be used. However, as in the case of two independent samples, this interval performed very poorly. As an alternative, a profile likelihood based method was derived. This method will be described in detail in section 2.3.1.

1.5 A Summary of Simulation Methods

For the different methods of constructing confidence intervals for data arising from the RPWR, we would like to know which perform best. Also, for the asymptotic methods, their actual confidence levels are different from their nominal confidence levels. However, with data from the RPWR, calculating the actual confidence level is far too complex a task. A solution to this problem is to use simulation to compare the methods.

Simulation is a method to model and analyze stochastic systems using computers. A main objective of simulation is to understand the behavior of a system without actually observing the real system. In this particular case, the system to be observed is the RPWR. A particular outcome of the RPWR can be simulated, and then confidence intervals can be constructed. This can be repeated many times and the proportion of intervals which contain the true parameter value and the average length of the intervals can be determined. These are both estimates of their theoretical counterparts.

1.5 A Summary of Simulation Methods

The most important aspect of simulation is the generation of pseudo-random numbers. A pseudorandom number generator creates a sequence of numbers between 0 and 1 which appear to be uniformly distributed. In reality, they are not truly uniform, since the next number to be generated is actually a function of one or more of the previous numbers. Eventually, a generator will go through the entire sequence and return to the beginning. The length of the entire sequence until it reaches the beginning again is known as the period of the generator. It is desirable for a generator to have a large period, be efficient and quick, as well as produce numbers which appear random.

To test pseudorandom number generators, many statistical tests can be used. In particular, classical goodness of fit tests such as the χ^2 or Kolmogorov-Smirnov tests can be employed, or more powerful methods such as those described in Zhang [40] can be used to check for one-dimensional uniformity. Other tests check for uniformity in higher dimensions. For example, if each random number is plotted against the previous random number, the points should be uniformly distributed in the unit square. This concept can be extended to higher dimensions. It is also important to test a pseudorandom number generator for independence and ensure there is no autocorrelation. This can be accomplished using the serial correlation test.

A simple pseudorandom number generator is the multiplicative linear congruential generator. For this generator, we have a sequence $x_1, x_2, \ldots$ where $x_n = ax_{n-1} \mod m$. The random numbers are given by x_i/m. The

most popular choices for a and m are $a = 7^5$ and $m = 2^{31} - 1$. This generator has a period of $2^{31} - 1$.

For the simulations in section 2.5, random numbers were generated using the default generator in R, the Mersenne Twister. This generator is currently the best for simulation purposes. It is quick and efficient, and has an enormous period of $2^{19937} - 1$. It has also been shown that the generator produces random numbers which are uniformly distributed in 623 dimensions. This generator is more complex and will not be described here. For more details, see Matsumoto and Nishimura [22].

For the RPWR, generating a binomial random variable, b is crucial. This is simple once a uniform number, u is generated. If p is the probability of a success, then $b = 1$ if $u \leq p$ and $b = 0$ otherwise.

1.6 Summary of the Thesis

Inference for the RPWR is examined. In particular, the construction of confidence intervals for the difference of success probabilities is considered.

In chapter 2, the theoretical background of the RPWR is presented. Two methods for constructing confidence intervals for the difference are explained in detail. These are the profile likelihood and Jeffreys-Perks intervals. The profile likelihood interval has been used before, while the Jeffreys-Perks interval has never been employed in the case of the RPWR. The two methods will be compared by simulation. The criteria of interest is the actual confidence

level and the average length of the intervals.

Chapter 3 discusses a theoretical aspect of the Jeffreys-Perks interval. This interval is a special case of a more general method, and the reasons for choosing this particular interval will be discussed.

Some other theoretical issues can be found in chapter 4. A martingale central limit theorem is essential for the asymptotic methods of inference to be appropriate. The rate of convergence of this limiting result will be examined, as well as other theoretical results.

The RPWR still has many aspects to be examined, and there is plenty of research still to be done. Chapter 5 will discuss some possible extensions and future research.

Chapter 2

Theoretical Background and Computational Results

2.1 Introduction

In this chapter, some of the theoretical background for the RPWR will be discussed. After introducing all the necessary notation, some additional results will be determined such as the limiting behavior of the RPWR as well as the likelihood function and sufficient statistics.

Two methods for constructing confidence intervals for data arising from the RPWR will also be described in detail. Both are asymptotic methods, however both have the advantage of showing signs of having good properties for small sample sizes.

As an example, the data from the ECMO trial will be considered. Con-

fidence intervals will be constructed using both methods.

These two methods will also be compared by simulation. The criteria of interest is the actual coverage probability and the average length of the resulting intervals.

2.2 Theoretical Background

2.2.1 Notation

Our two treatments will be called "A" and "B". We have n patients in total. Define the following indicator variables:

$$T_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ patient receives treatment A} \\ 0, & \text{if the } i^{\text{th}} \text{ patient receives treatment B} \end{cases} \tag{2.1}$$

$$X_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ patient's response is a success} \\ 0, & \text{if the } i^{\text{th}} \text{ patient's response is a failure} \end{cases} \tag{2.2}$$

The success probabilities will be denoted by p_A and p_B for treatments A and B respectively. Let $q_A = 1 - p_A$ and $q_B = 1 - p_B$. Note that

$$p_A = \mathbb{P}(X_i = 1 | T_i = 1) \tag{2.3}$$

$$p_B = \mathbb{P}(X_i = 1 | T_i = 0) \tag{2.4}$$

2.2 Theoretical Background

The parameters of interest are:

$$\text{Difference:} \quad \Delta = p_A - p_B \tag{2.5}$$

$$\text{Odds Ratio:} \quad \theta = \frac{p_A q_B}{q_A p_B} \tag{2.6}$$

It will also be useful to define the number of patients on each treatment, the total number of successes and the number of successes on each treatment,

$$N_A = \sum_{i=1}^{n} T_i, \quad N_B = n - N_A, \tag{2.7}$$

$$S = \sum_{i=1}^{n} X_i, \quad S_A = \sum_{i=1}^{n} T_i X_i, \quad S_B = S - S_A. \tag{2.8}$$

Particular realizations of these variables will be written in lower case.

Initially, the urn contains α balls of each type and after each response, we add β balls of the appropriate type. An important quantity is the probability of patient i receiving treatment A, which shall be denoted by p_i. It is straightforward to verify that this probability is given by

$$p_i = \mathbb{P}(T_i = 1 | T_1, \ldots, T_{i-1}, X_1, \ldots, X_{i-1}) \tag{2.9}$$

$$= \frac{\alpha + \beta \left(2 \sum_{j=1}^{i-1} T_j X_j + (i-1) - \sum_{j=1}^{i-1} T_j - \sum_{j=1}^{i-1} X_j \right)}{2\alpha + (i-1)\beta} \tag{2.10}$$

since the numerator is the number of balls of type A and the denominator is the total number of balls after the first $i-1$ responses.

2.2.2 Limiting Results

The limiting behavior of N_A and p_i provides more insight into the problem. In order to obtain limiting results, it should be noted that the RPWR is a special case of the generalized Pólya urn model (see Rosenberger [28]). Following Athreya and Karlin [5], this model can be embedded into a continuous time Markov branching process, from which the desired limiting results can be obtained.

Define the addition matrix, $\mathbf{M}$, such that its ij^{th} entry is the expected number of balls of type j to be added to the urn after a ball of type i was drawn, $i, j = A, B$. This matrix has an interpretation in the context of branching processes from which the limiting results follow (see Athreya and Karlin [4]). For the RPWR,

$$\mathbf{M} = \beta \begin{pmatrix} p_A & q_A \\ q_B & p_B \end{pmatrix} \tag{2.11}$$

The left eigenvector[‡] associated with the largest eigenvalue of $\mathbf{M}$ determines the limits of interest. In this case, the eigenvalues of $\mathbf{M}$ are β and $\beta(p_A + p_B - 1)$, hence the largest eigenvalue is β. The left eigenvector associated

[‡]A vector $\mathbf{v}$ associated with an eigenvalue λ satisfying $\mathbf{v}^{\mathrm{T}}\mathbf{M} = \mathbf{v}^{\mathrm{T}}\lambda$

with β is

$$\mathbf{v} = (v_A, v_B)^{\mathrm{T}} \tag{2.12}$$

$$= \left(\frac{q_B}{q_A + q_B}, \frac{q_A}{q_A + q_B} \right)^{\mathrm{T}} \tag{2.13}$$

In Rosenberger and Sriram [31], it is shown that

$$\lim_{n \to \infty} \frac{N_A}{n} = v_A = \frac{q_B}{q_A + q_B} \qquad \text{a.s.} \tag{2.14}$$

and

$$\lim_{n \to \infty} p_i = v_A = \frac{q_B}{q_A + q_B} \qquad \text{a.s.} \tag{2.15}$$

A consequence of these results is that the asymptotic proportion of successes is given by

$$\frac{p_A q_B + p_B q_A}{q_A + q_B} \tag{2.16}$$

If we had randomly assigned half the patients to A and the rest to B, the success rate would be $\frac{1}{2}(p_A + p_B)$. If we assume $p_A > p_B$, then

$$\frac{q_B}{q_A + q_B} > \frac{1}{2} \qquad \text{and} \qquad \frac{q_A}{q_A + q_B} < \frac{1}{2} \tag{2.17}$$

hence more weight is placed on the more successful treatment, and the total success rate of the RPWR is higher.

2.2.3 Likelihood Function and Sufficient Statistics

The likelihood function also plays an important role, so it will be derived. Let $\mathscr{L}_n(p_A, p_B)$ be the likelihood function based on n patients. Then, following Rosenberger et al. [30],

$$\begin{aligned}
\mathscr{L}_n(p_A, p_B) &= \mathbb{P}(T_1, \ldots, T_n, X_1, \ldots, X_n) \\
&= \mathbb{P}(X_n | T_1, \ldots, T_n, X_1, \ldots, X_{n-1}) \mathbb{P}(T_1, \ldots, T_n, X_1, \ldots, X_{n-1}) \\
&= p_A^{T_n X_n} q_A^{T_n(1-X_n)} p_B^{(1-T_n)X_n} q_B^{(1-T_n)(1-X_n)} \\
&\qquad \times \mathbb{P}(T_n | T_1, \ldots, T_{n-1}, X_1, \ldots, X_{n-1}) \\
&\qquad\qquad \times \mathbb{P}(T_1, \ldots, T_{n-1}, X_1, \ldots, X_{n-1}) \\
&= p_A^{T_n X_n} q_A^{T_n(1-X_n)} p_B^{(1-T_n)X_n} q_B^{(1-T_n)(1-X_n)} \\
&\qquad\qquad \times p_n^{T_n}(1-p_n)^{1-T_n} \mathscr{L}_{n-1}(p_A, p_B)
\end{aligned} \tag{2.18}$$

and after unwinding the recursion,

$$\mathscr{L}_n(p_A, p_B) = p_A^{S_A} q_A^{N_A - S_A} p_B^{S_B} q_B^{N_B - S_B} \prod_{i=1}^{n} p_i^{T_i}(1-p_i)^{1-T_i} \tag{2.19}$$

hence, the likelihood function is proportional to the likelihood function in the usual case of two independent samples. The likelihood function can also be expressed in terms of Δ (or θ) and p_B. Hence, we have a main parameter of interest (Δ or θ) and a nuisance parameter (p_B).

Since the likelihood function corresponds exactly to the probability dis-

tribution of $T_1, \ldots, T_n$ and $X_1, \ldots, X_n$, the sufficient statistics can be determined using the factorization criterion. Clearly, N_A, N_B, S_A and S_B are jointly sufficient. However, N_B is a function of N_A. Hence, N_A, S_A and S_B are jointly minimal sufficient statistics.

2.3 Methods

Two asymptotic methods for analyzing data arising from a RPWR design will be compared by simulation. The first is the profile likelihood method (Wei et al. [38]), and the second is known as the Jeffreys-Perks method (Beal [9]). Both methods are based on maximum likelihood estimators and related ideas.

2.3.1 Profile Likelihood Method

The likelihood function for this problem is proportional to the usual likelihood function from the case of two independent samples. The MLE's are well known to be

$$\hat{p}_A = \frac{S_A}{N_A} \qquad \text{and} \qquad \hat{p}_B = \frac{S_B}{N_B} \tag{2.20}$$

In the case of two independent samples, it is known that $\hat{p}_A$ and $\hat{p}_B$ are asymptotically normally distributed, and confidence intervals can easily be constructed. For the RPWR design, there is a corresponding result. Using

martingale limit theory and the Cramér-Wold theorem, it can be shown that

$$\frac{\sqrt{N_A}(\hat{p}_A - p_A)}{\sqrt{p_A q_A}} \quad \text{and} \quad \frac{\sqrt{N_B}(\hat{p}_B - p_B)}{\sqrt{p_B q_B}} \tag{2.21}$$

are asymptotically independent and each follow a standard normal distribution.

We can use this result to construct a confidence interval based on the likelihood ratio statistic. However, we have a nuisance parameter. To overcome this, we use what is known as the profile likelihood, which essentially means the nuisance parameter is replaced by its restricted MLE (*i.e.* the MLE assuming the parameter of interest is known).

Suppose the parameter of interest is the difference, Δ. We can write the log-likelihood in terms of Δ and p_B. Let $\ell(\Delta, p_B) = \log \mathscr{L}_n(\Delta, p_B)$. Then the profile likelihood for Δ is

$$\ell^*(\Delta) = \max_{p_B \in \mathscr{D}} \ell(\Delta, p_B) \tag{2.22}$$

$$= \ell\left(\Delta, p_B^*(\Delta)\right) \tag{2.23}$$

where $p_B^*(\Delta)$ is the restricted MLE of p_B and

$$\mathscr{D} = \begin{cases} (0, 1-\Delta) & \text{if } \Delta \geq 0 \\ (-\Delta, 1) & \text{if } \Delta < 0 \end{cases} \tag{2.24}$$

Obtaining $p_B^*(\Delta)$ involves solving a cubic equation. A closed form is given in Miettinen and Nurminen [24].

Since asymptotic normality still holds, the $100(1-\gamma)\%$ confidence region is given by

$$\left\{\Delta : 2\left[\ell(\hat{\Delta}, \hat{p_B}) - \ell^*(\Delta)\right] \leq \chi_1^2(\gamma)\right\} \tag{2.25}$$

where $\chi_1^2(\gamma)$ is the χ_1^2 quantile with upper tail probability equal to γ.

2.3.2 Jeffreys-Perks Method

This method is a specific case of an idea discussed in Beal [9]. For this method, it is useful to define $a = p_A + p_B$, $u = \frac{1}{4}\left[\frac{1}{n_A} + \frac{1}{n_B}\right]$ and $v = \frac{1}{4}\left[\frac{1}{n_A} - \frac{1}{n_B}\right]$. Since the MLE's of p_A and p_B are still asymptotically normal, we have that

$$\left\{\Delta : (\Delta - \hat{\Delta})^2 \leq c\mathbb{V}\mathrm{ar}(\hat{\Delta}; \tilde{a}, \tilde{\Delta})\right\} \tag{2.26}$$

is a $100(1-\gamma)\%$ confidence region for Δ, where $c = \chi_1^2(\gamma)$ and $\tilde{a}$ and $\tilde{\Delta}$ are expressions for a and Δ.

It turns out that

$$\mathbb{V}\mathrm{ar}(\hat{\Delta}; a, \Delta) = u\left[(2-a)a - \Delta^2\right] + 2v(1-a)\Delta \tag{2.27}$$

which is quadratic in Δ, so equality holds at two points in (2.26), which are

the solutions to

$$(\Delta - \hat{\Delta})^2 = c\mathbb{V}\mathrm{ar}(\hat{\Delta}; \tilde{a}, \tilde{\Delta}) \tag{2.28}$$

These two solutions form the endpoints of the confidence interval.

Beal suggested taking $\tilde{\Delta} = \Delta$ and $\tilde{a}$ to be the Bayes estimator of a with a prior density proportional to $(p_A q_A p_B q_B)^{\kappa}$, *i.e.*

$$\begin{aligned}\tilde{a}(\kappa) = \frac{n_A}{n_A + 2(\kappa+1)}\hat{p}_A + \frac{\kappa+1}{n_A + 2(\kappa+1)} \\ + \frac{n_B}{n_B + 2(\kappa+1)}\hat{p}_B + \frac{\kappa+1}{n_B + 2(\kappa+1)}\end{aligned} \tag{2.29}$$

or,

$$\tilde{a}(\kappa) = \frac{s_A + (\kappa+1)}{n_A + 2(\kappa+1)} + \frac{s_B + (\kappa+1)}{n_B + 2(\kappa+1)} \tag{2.30}$$

From (2.29) we see how $\tilde{a}(\kappa)$ can be obtained by modifying the MLE's and (2.30) shows how $\tilde{a}(\kappa)$ can be interpreted as an estimate of a with an additional $\kappa + 1$ successes and failures on each treatment. For the Jeffreys-Perks interval, we take $\kappa = -\frac{1}{2}$.

The solutions to (2.28) in this case are given by,

$$\frac{\hat{\Delta} + cv(1-\tilde{a})}{1+cu} \pm \frac{\sqrt{c\left(\mathbb{V}\mathrm{ar}(\hat{\Delta}; \tilde{a}, \hat{\Delta}) + cu^2(2-\tilde{a})\tilde{a} + cv^2(1-\tilde{a})^2\right)}}{1+cu} \tag{2.31}$$

2.4 Example: The ECMO Trial

The ECMO trial employed the RPWR using $\alpha = 1$ and $\beta = 1$. The first ten patients were randomly assigned to ECMO (treatment A) or conventional treatment (treatment B) using this rule. The first patient received ECMO and survived. The second patient received the conventional treatment and died. The remaining patients all received ECMO and survived. In our notation, we have $t_i = x_i = 1$ for $i = 1, 3, \ldots, 10$ and $t_2 = x_2 = 0$. This gives $n_A = 9, s_A = 9, s_B = 0$.

The confidence intervals for the difference are given in table 2.1. We can easily see that in this case, the profile likelihood intervals are shorter. In the next section, simulations will assess whether or not this is always the case, or if this is a rare exception.

Confidence Level	Profile Likelihood	Jeffreys-Perks
0.90	(0.258, 1.000)	(0.140, 1.000)
0.95	(0.146, 1.000)	(−0.010, 1.000)
0.99	(−0.007, 1.000)	(−0.236, 1.000)

Table 2.1: Confidence intervals for Δ using the ECMO data.

2.5 Simulation Results

The profile based confidence interval and the Jeffreys-Perks interval were compared by simulation. The values for the parameters p_A and p_B were 0.1, 0.3, 0.5, 0.7 and 0.9. The simulations were done for sample sizes (n) of

5, 10 and 25, and nominal confidence levels $(1-\gamma)$ of 0.90, 0.95 and 0.99. The measures of interest are the actual confidence level (measured by the observed coverage probability) and the length of the intervals.

For each combination of p_A, p_B, n and $1-\gamma$, 10,000 outcomes of the RPWR with $\alpha = 1$ and $\beta = 1$ were generated. Note that p_A is always at least as large as p_B. Results for cases where p_B is larger than p_A can be determined by symmetry. The actual confidence level was the proportion of intervals which captured the true value of Δ. The average length of the intervals corresponding to the simulated outcomes was used as a measure of the length of the intervals. All simulations were done in R, version 1.8.1. The code which was used can be found in the appendix.

The results of the simulations are displayed in tables 2.2 – 2.5 and figures 2.1 – 2.10.

			Confidence Level		Length		
p_B	p_A	n	Profile Likelihood	Jeffreys–Perks	Profile Likelihood	Jeffreys–Perks	Ratio (PL/JP)
0.1	0.1	5	0.9776	0.9960	0.9739	0.9678	1.0063
		10	0.9685	0.9872	0.6434	0.6549	0.9824
		25	0.8503	0.9676	0.3950	0.3967	0.9955
	0.3	5	0.9506	0.9571	1.0321	1.0254	1.0065
		10	0.8879	0.9534	0.7381	0.7402	0.9972
		25	0.8857	0.9221	0.4815	0.4804	1.0023
	0.5	5	0.8434	0.8981	1.0536	1.0644	0.9898
		10	0.9030	0.9123	0.7703	0.7834	0.9832
		25	0.8869	0.9050	0.5029	0.5089	0.9882
	0.7	5	0.9104	0.9003	1.0298	1.0784	0.9549
		10	0.9211	0.9058	0.7630	0.8035	0.9496
		25	0.9142	0.9250	0.4926	0.5156	0.9554
	0.9	5	0.9247	0.9087	0.9629	1.0678	0.9018
		10	0.9424	0.9287	0.7229	0.8064	0.8964
		25	0.9522	0.9433	0.4946	0.5536	0.8933
0.3	0.3	5	0.8940	0.9755	1.0935	1.0774	1.0150
		10	0.8806	0.9128	0.8312	0.8159	1.0188
		25	0.8758	0.8857	0.5711	0.5588	1.0220
	0.5	5	0.8894	0.9083	1.1263	1.1166	1.0086
		10	0.8435	0.9198	0.8685	0.8581	1.0121
		25	0.8704	0.8960	0.5998	0.5904	1.0160
	0.7	5	0.8420	0.9146	1.1191	1.1361	0.9850
		10	0.8771	0.9196	0.8604	0.8723	0.9864
		25	0.8666	0.8893	0.5956	0.5988	0.9948
	0.9	5	0.9363	0.9250	1.0680	1.1343	0.9416
		10	0.9081	0.9344	0.8096	0.8701	0.9305
		25	0.8766	0.9194	0.5871	0.6246	0.9399
0.5	0.5	5	0.8423	0.9490	1.1706	1.1588	1.0101
		10	0.8551	0.8807	0.9142	0.8978	1.0183
		25	0.8823	0.8836	0.6328	0.6214	1.0183
	0.7	5	0.8430	0.8909	1.1890	1.1908	0.9985
		10	0.8225	0.8927	0.9123	0.9115	1.0009
		25	0.8767	0.8935	0.6306	0.6259	1.0009
	0.9	5	0.8623	0.9305	1.1551	1.1906	0.9701
		10	0.8650	0.9226	0.8674	0.9068	0.9566
		25	0.8625	0.8798	0.6169	0.6419	0.9611
0.7	0.7	5	0.8538	0.9203	1.2161	1.2184	0.9981
		10	0.8609	0.8905	0.9257	0.9269	0.9987
		25	0.8730	0.8943	0.6212	0.6164	1.0079
	0.9	5	0.8581	0.9362	1.2200	1.2383	0.9852
		10	0.8282	0.9180	0.8869	0.9172	0.9669
		25	0.8665	0.8946	0.5918	0.6081	0.9731
0.9	0.9	5	0.9379	0.9460	1.2631	1.2781	0.9882
		10	0.9083	0.9445	0.8975	0.9358	0.9591
		25	0.8663	0.9489	0.5335	0.5591	0.9541

Table 2.2: Confidence levels and lengths (90% Nominal level)

			Confidence Level		Length		
p_B	p_A	n	Profile Likelihood	Jeffreys–Perks	Profile Likelihood	Jeffreys–Perks	Ratio (PL/JP)
0.1	0.1	5	0.9951	0.9978	1.1843	1.0735	1.1033
		10	0.9888	0.9976	0.8058	0.7494	1.0752
		10	0.9577	0.9929	0.4887	0.4641	1.0529
	0.3	5	0.9615	0.9920	1.2330	1.1446	1.0772
		10	0.9583	0.9833	0.8954	0.8471	1.0571
		25	0.9350	0.9558	0.5813	0.5627	1.0331
	0.5	5	0.9836	0.9836	1.2370	1.1895	1.0340
		10	0.9489	0.9415	0.9270	0.9064	1.0227
		25	0.9565	0.9546	0.6044	0.5995	1.0081
	0.7	5	0.9692	0.9301	1.1984	1.2114	0.9893
		10	0.9642	0.9556	0.9050	0.9310	0.9721
		25	0.9666	0.9579	0.5935	0.6112	0.9711
	0.9	5	0.9777	0.9216	1.0934	1.1896	0.9192
		10	0.9824	0.9483	0.8413	0.9325	0.9021
		25	0.9741	0.9669	0.5923	0.6569	0.9017
0.3	0.3	5	0.9768	0.9849	1.2819	1.2050	1.0668
		10	0.9230	0.9770	0.9891	0.9347	1.0582
		25	0.9272	0.9490	0.6793	0.6532	1.0400
	0.5	5	0.9260	0.9717	1.3029	1.2411	1.0498
		10	0.9381	0.9528	1.0243	0.9830	1.0420
		25	0.9280	0.9464	0.7101	0.6906	1.0283
	0.7	5	0.9740	0.9740	1.2761	1.2577	1.0147
		10	0.9576	0.9489	1.0050	0.9995	1.0055
		25	0.9315	0.9540	0.7011	0.6996	1.0021
	0.9	5	0.9738	0.9481	1.2121	1.2551	0.9657
		10	0.9604	0.9584	0.9349	0.9930	0.9414
		25	0.9562	0.9688	0.6894	0.7341	0.9391
0.5	0.5	5	0.9473	0.9739	1.3408	1.2816	1.0461
		10	0.8938	0.9560	1.0645	1.0222	1.0414
		25	0.9379	0.9444	0.7459	0.7256	1.0279
	0.7	5	0.9185	0.9664	1.3426	1.3015	1.0316
		10	0.9184	0.9485	1.0577	1.0330	1.0240
		25	0.9308	0.9392	0.7400	0.7286	1.0156
	0.9	5	0.9723	0.9723	1.2964	1.2980	0.9988
		10	0.9561	0.9577	1.0038	1.0305	0.9741
		25	0.9325	0.9433	0.7182	0.7455	0.9634
0.7	0.7	5	0.9203	0.9758	1.3711	1.3273	1.0330
		10	0.9130	0.9588	1.0720	1.0455	1.0253
		25	0.9288	0.9489	0.7312	0.7163	1.0208
	0.9	5	0.9490	0.9856	1.3669	1.3362	1.0230
		10	0.9313	0.9680	1.0353	1.0342	1.0010
		25	0.9268	0.9463	0.6982	0.7079	0.9862
0.9	0.9	5	0.9462	0.9949	1.4074	1.3603	1.0347
		10	0.9584	0.9869	1.0290	1.0228	1.0061
		25	0.9275	0.9822	0.6329	0.6388	0.9908

Table 2.3: Confidence levels and lengths (95% Nominal level)

			Confidence Level		Length		
p_B	p_A	n	Profile Likelihood	Jeffreys–Perks	Profile Likelihood	Jeffreys–Perks	Ratio (PL/JP)
0.1	0.1	5	0.9981	1.0000	1.5258	1.2344	1.2361
		10	0.9995	1.0000	1.1179	0.9055	1.2345
		25	0.9976	0.9998	0.6833	0.5851	1.1677
	0.3	5	0.9990	0.9994	1.5431	1.3171	1.1715
		10	0.9993	0.9995	1.1872	1.0251	1.1581
		25	0.9866	0.9855	0.7813	0.7123	1.0969
	0.5	5	0.9998	0.9998	1.5238	1.3713	1.1112
		10	0.9917	0.9818	1.1996	1.1028	1.0877
		25	0.9938	0.9868	0.8012	0.7631	1.0499
	0.7	5	0.9952	0.9745	1.4647	1.4019	1.0448
		10	0.9915	0.9746	1.1554	1.1409	1.0127
		25	0.9934	0.9844	0.7884	0.7902	0.9978
	0.9	5	0.9928	0.9718	1.3544	1.3964	0.9699
		10	0.9945	0.9748	1.0572	1.1515	0.9181
		25	0.9966	0.9899	0.7764	0.8557	0.9073
0.3	0.3	5	0.9844	1.0000	1.5727	1.3752	1.1436
		10	0.9971	0.9989	1.2669	1.1259	1.1252
		25	0.9843	0.9944	0.8860	0.8226	1.0771
	0.5	5	0.9948	0.9968	1.5709	1.4179	1.1079
		10	0.9933	0.9957	1.2894	1.1835	1.0895
		25	0.9846	0.9914	0.9147	0.8679	1.0539
	0.7	5	0.9974	0.9974	1.5368	1.4438	1.0644
		10	0.9867	0.9891	1.2550	1.2077	1.0391
		25	0.9916	0.9932	0.9002	0.8839	1.0184
	0.9	5	0.9923	0.9810	1.4540	1.4355	1.0129
		10	0.9936	0.9840	1.1517	1.1960	0.9630
		25	0.9958	0.9895	0.8599	0.9154	0.9394
0.5	0.5	5	0.9714	1.0000	1.5973	1.4571	1.0962
		10	0.9907	0.9977	1.3281	1.2280	1.0816
		25	0.9827	0.9889	0.9541	0.9100	1.0484
	0.7	5	0.9883	0.9958	1.5874	1.4710	1.0791
		10	0.9904	0.9957	1.3112	1.2361	1.0608
		25	0.9845	0.9913	0.9441	0.9131	1.0339
	0.9	5	0.9975	0.9975	1.5365	1.4632	1.0501
		10	0.9905	0.9915	1.2248	1.2121	1.0105
		25	0.9907	0.9899	0.9009	0.9244	0.9745
0.7	0.7	5	0.9776	1.0000	1.6141	1.4908	1.0827
		10	0.9872	0.9974	1.3227	1.2360	1.0701
		25	0.9843	0.9911	0.9369	0.8944	1.0474
	0.9	5	0.9902	0.9983	1.5983	1.4775	1.0818
		10	0.9896	0.9981	1.2737	1.2046	1.0574
		25	0.9836	0.9902	0.8837	0.8653	1.0213
0.9	0.9	5	0.9946	1.0000	1.6409	1.4885	1.1024
		10	0.9967	0.9999	1.2799	1.1725	1.0916
		25	0.9881	0.9974	0.8314	0.7828	1.0622

Table 2.4: Confidence levels and lengths (99% Nominal level)

$1-\gamma$	Δ	n	Profile Likelihood	Jeffreys–Perks
0.90	0.0	5	0.60	0.00
		10	0.60	0.40
		25	1.00	0.60
	0.2	5	0.75	0.25
		10	1.00	0.25
		25	1.00	0.75
	0.4	5	1.00	0.33
		10	0.67	0.00
		25	1.00	0.67
	0.6	5	0.00	0.00
		10	0.00	0.00
		25	0.50	0.00
	0.8	5	0.00	0.00
		10	0.00	0.00
		25	0.00	0.00
0.95	0.0	5	0.60	0.00
		10	0.60	0.00
		25	0.80	0.60
	0.2	5	0.75	0.00
		10	0.75	0.25
		25	1.00	0.75
	0.4	5	0.00	0.00
		10	0.33	0.67
		25	0.67	0.33
	0.6	5	0.00	1.00
		10	0.00	0.00
		25	0.00	0.00
	0.6	5	0.00	1.00
		10	0.00	1.00
		25	0.00	0.00
0.99	0.0	5	0.60	0.00
		10	0.20	0.00
		25	0.80	0.20
	0.2	5	0.25	0.00
		10	0.25	0.00
		25	1.00	0.25
	0.4	5	0.00	0.00
		10	0.33	0.67
		25	0.00	0.67
	0.6	5	0.00	1.00
		10	0.00	1.00
		25	0.00	1.00
	0.8	5	0.00	1.00
		10	0.00	1.00
		25	0.00	1.00

Table 2.5: Proportion of intervals that are anti–conservative

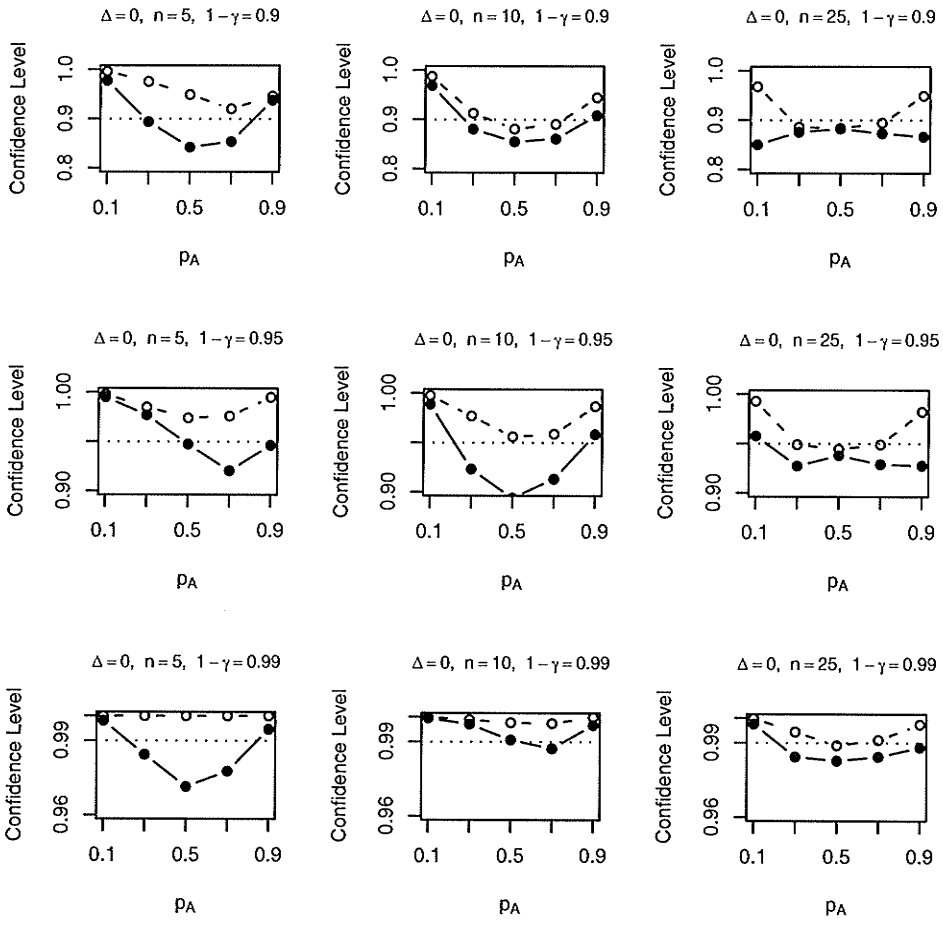

Figure 2.1: Observed confidence level of the profile likelihood (solid) and Jeffreys-Perks (dashed) methods when $\Delta = 0$

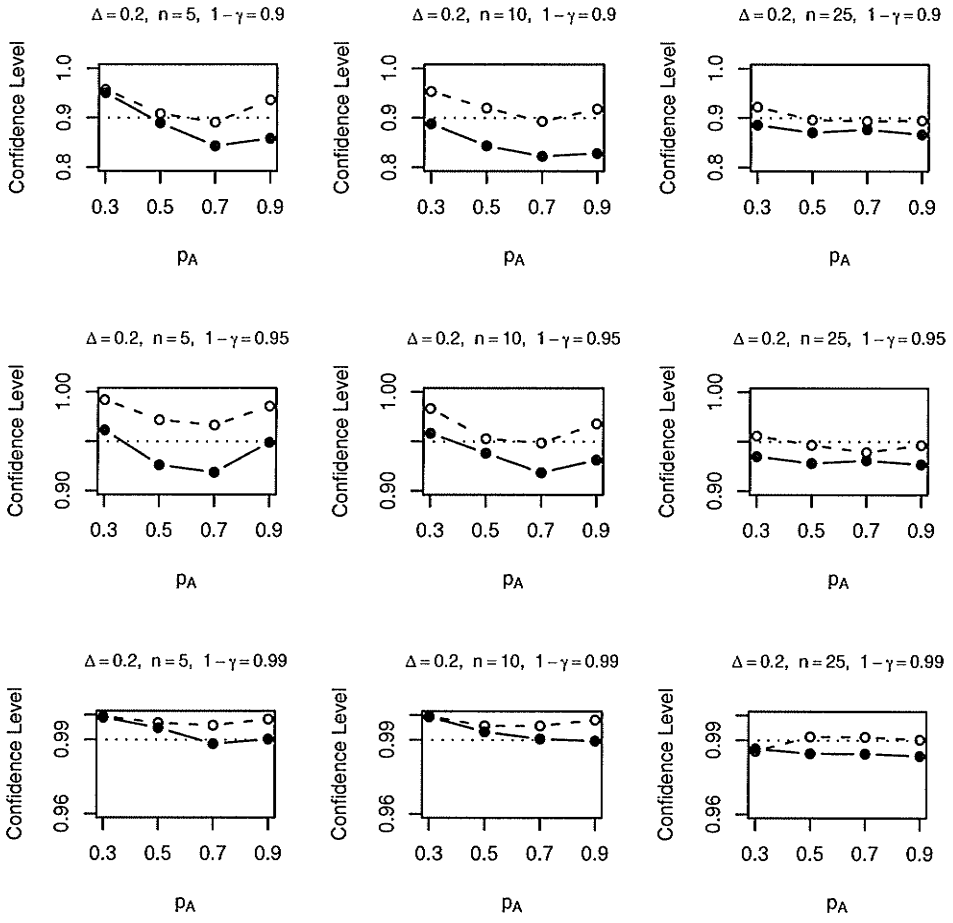

Figure 2.2: Observed confidence level of the profile likelihood (solid) and Jeffreys-Perks (dashed) methods when $\Delta = 0.2$

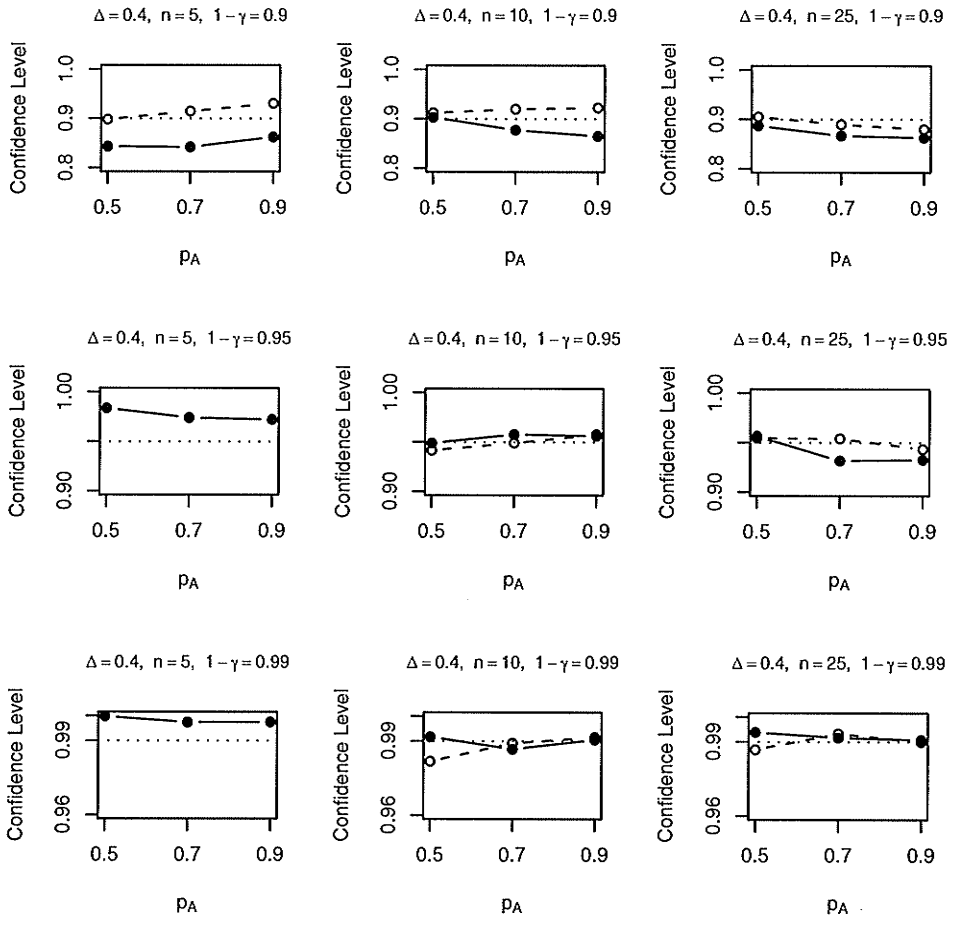

Figure 2.3: Observed confidence level of the profile likelihood (solid) and Jeffreys-Perks (dashed) methods when $\Delta = 0.4$

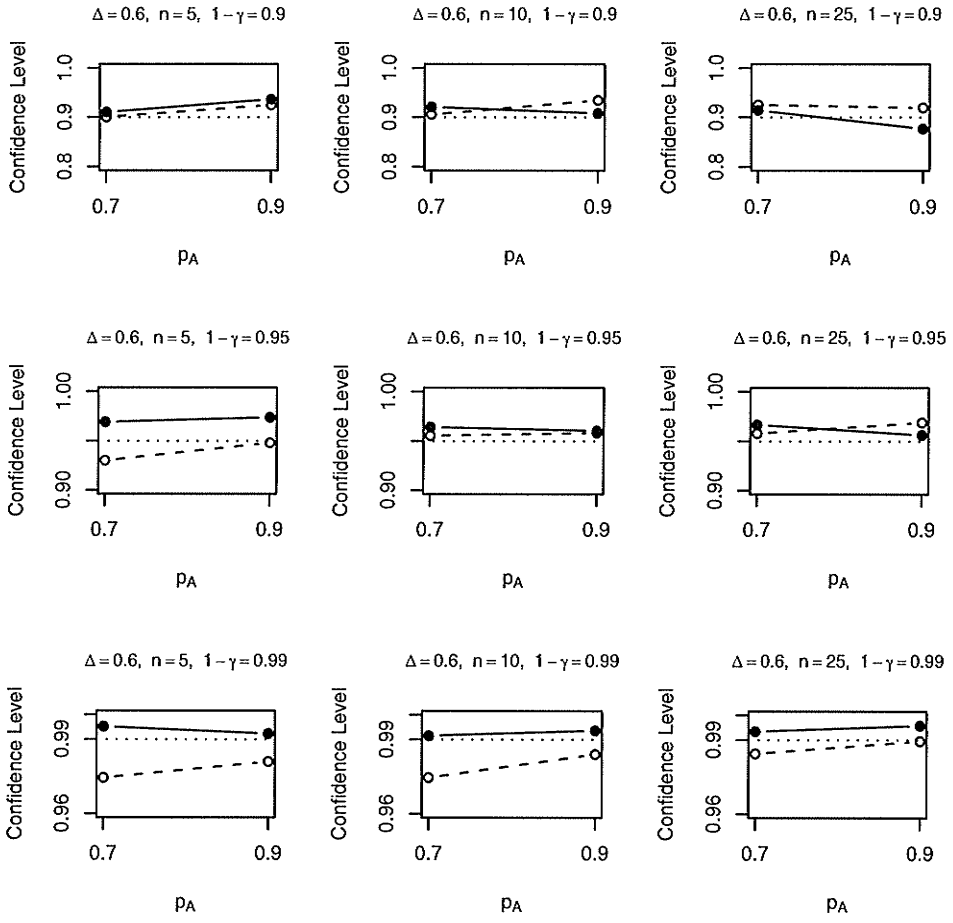

Figure 2.4: Observed confidence level of the profile likelihood (solid) and Jeffreys-Perks (dashed) methods when $\Delta = 0.6$

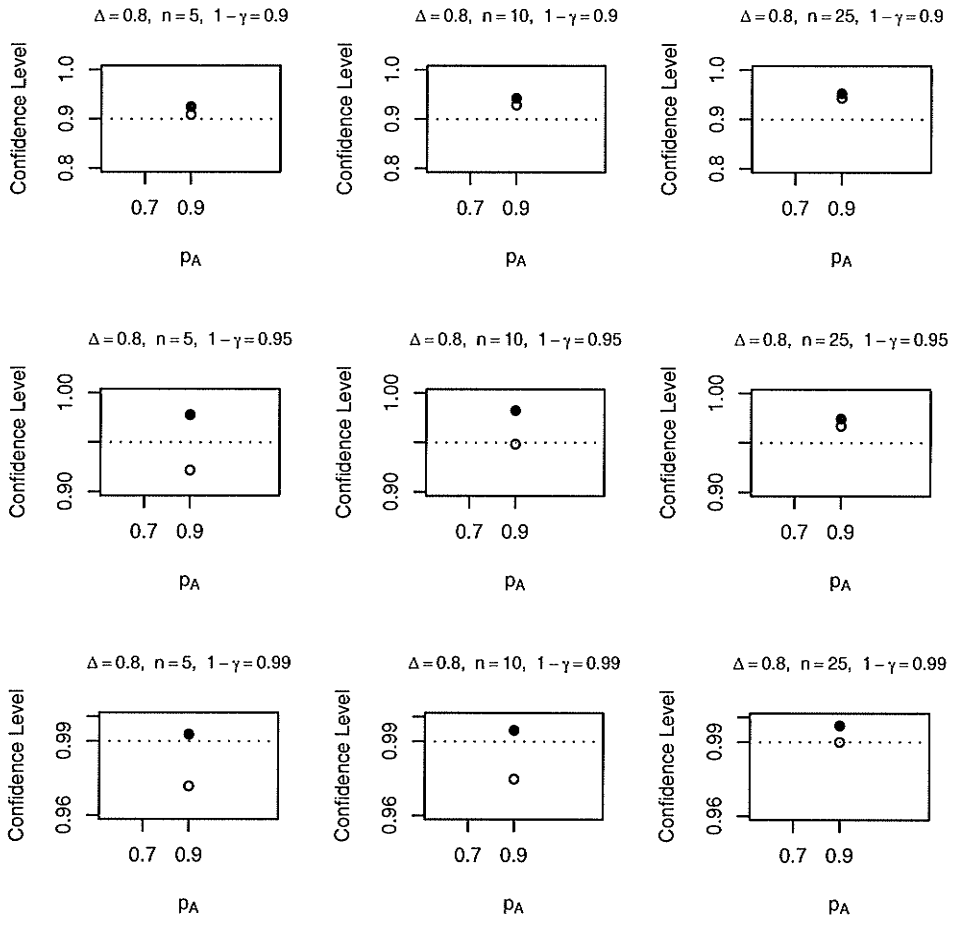

Figure 2.5: Observed confidence level of the profile likelihood (solid) and Jeffreys-Perks (dashed) methods when $\Delta = 0.8$

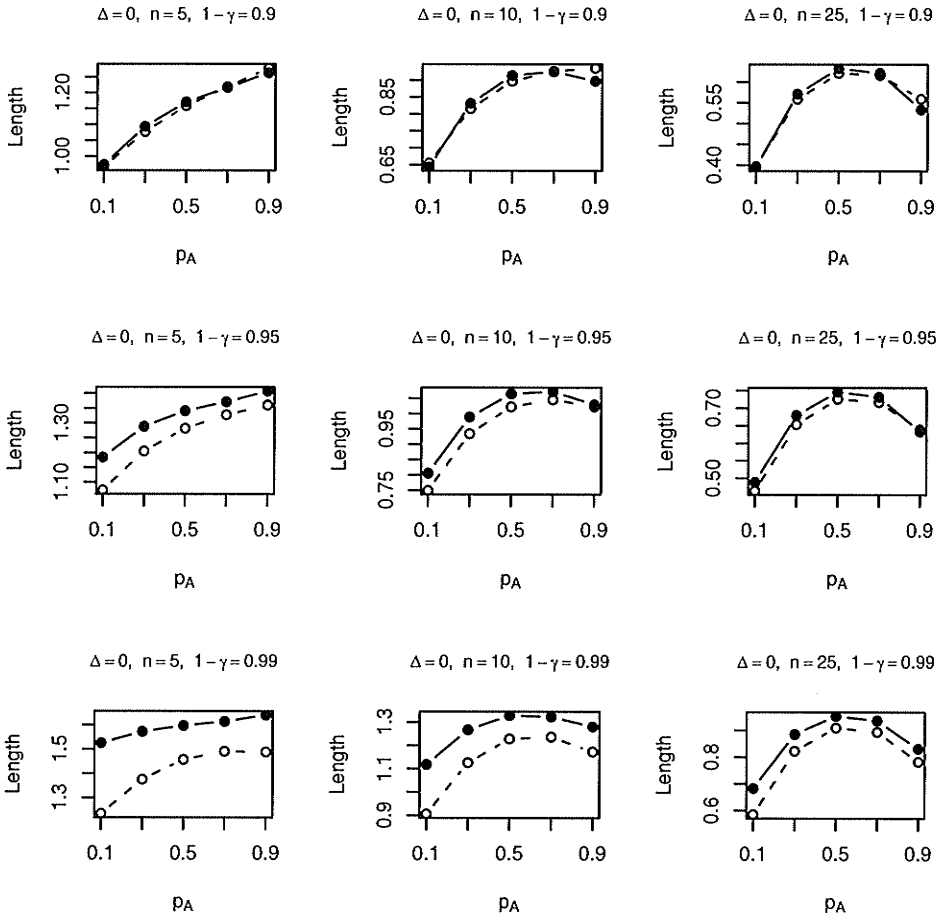

Figure 2.6: Average length of the profile likelihood (solid) and Jeffreys-Perks (dashed) confidence intervals when $\Delta = 0$

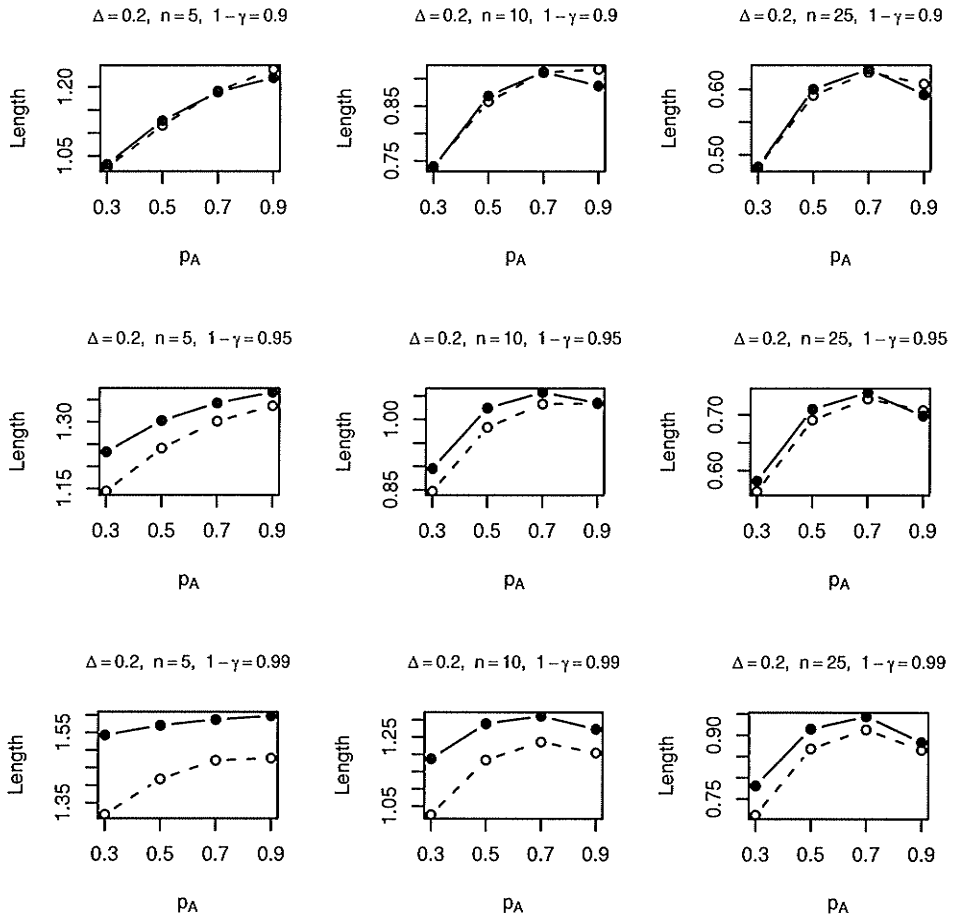

Figure 2.7: Average length of the profile likelihood (solid) and Jeffreys-Perks (dashed) confidence intervals when $\Delta = 0.2$

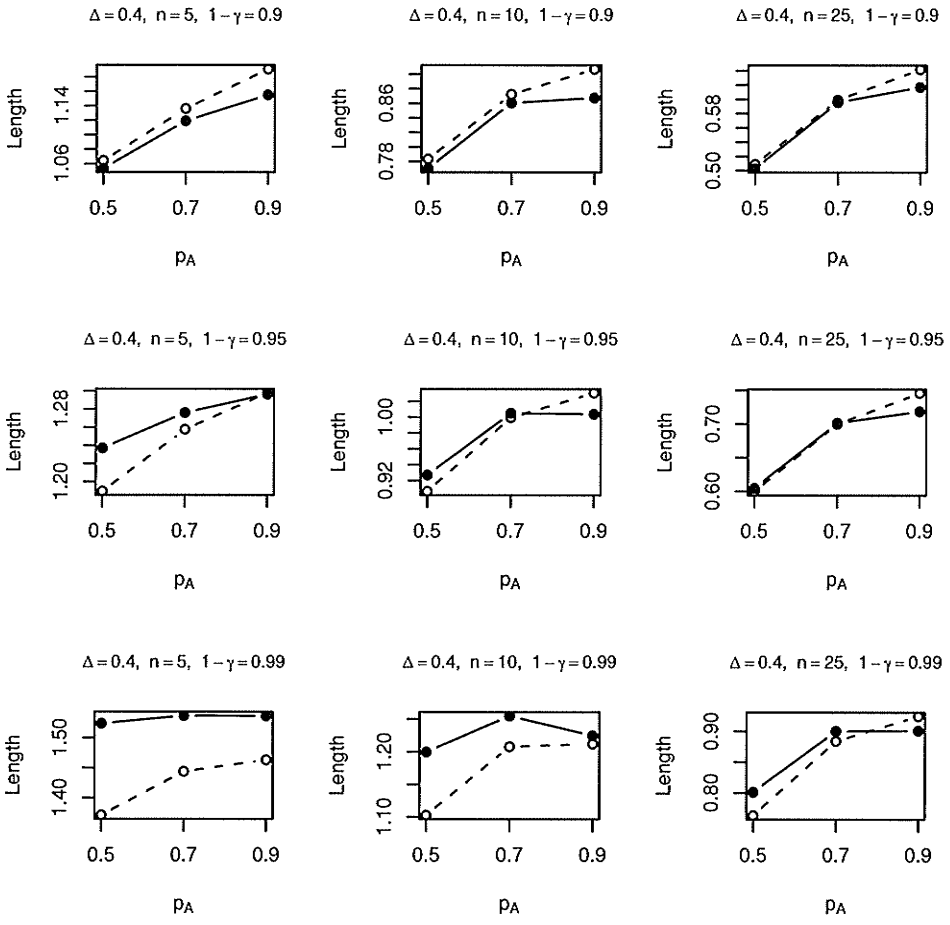

Figure 2.8: Average length of the profile likelihood (solid) and Jeffreys-Perks (dashed) confidence intervals when $\Delta = 0.4$

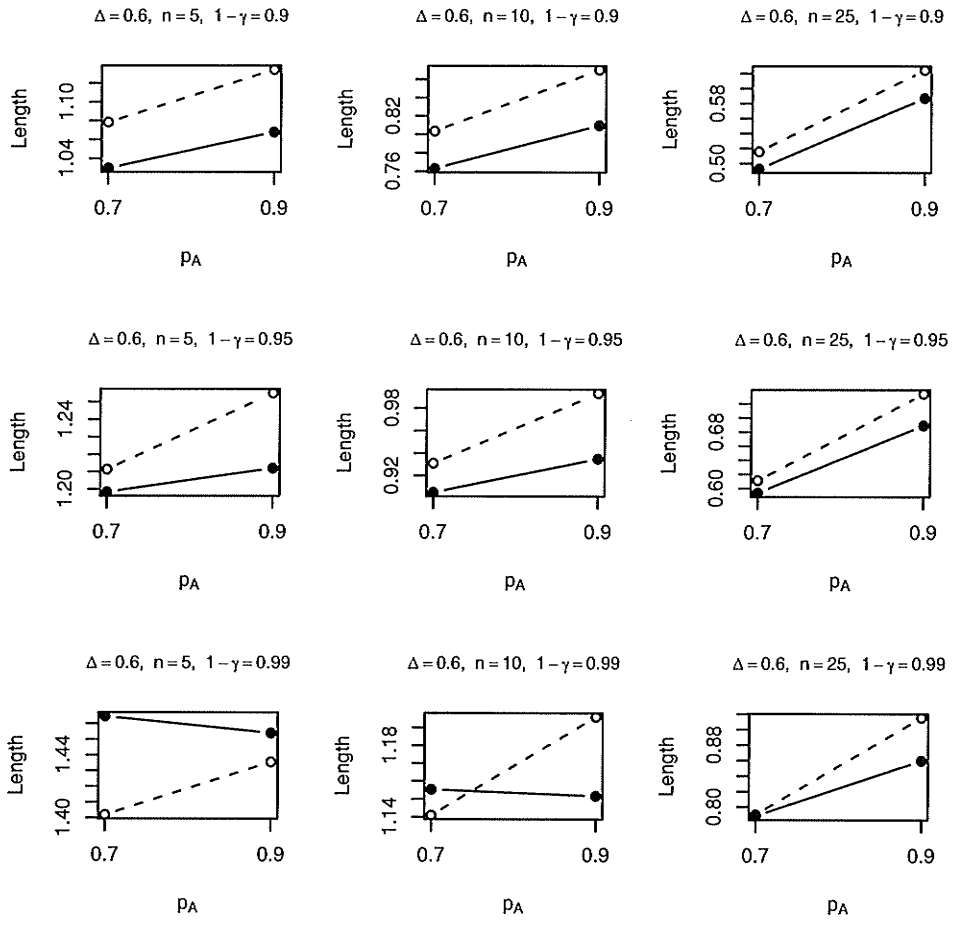

Figure 2.9: Average length of the profile likelihood (solid) and Jeffreys-Perks (dashed) confidence intervals when $\Delta = 0.6$

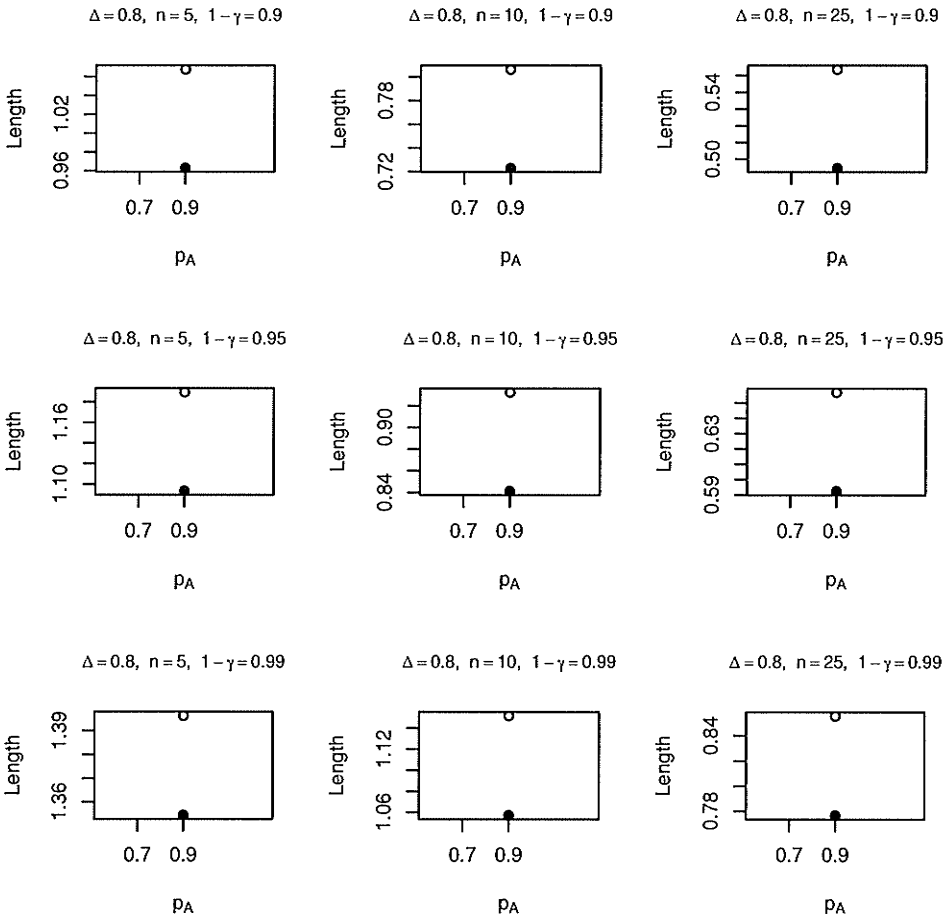

Figure 2.10: Average length of the profile likelihood (solid) and Jeffreys-Perks (dashed) confidence intervals when $\Delta = 0.8$

2.6 Observations

By examining figures 2.1 – 2.10 there are a number of things to note. One main observation is that there appears to be differences depending on the magnitude of Δ.

When Δ is 0, 0.2 or 0.4, the Jeffreys-Perks method always has a higher actual confidence level. The profile likelihood method is often anti-conservative, while the Jeffreys-Perks method is usually conservative. As the sample size increases, the coverage probability of both methods becomes closer to nominal.

When Δ is 0.6 or 0.8, the situation is slightly different. For the most part, the coverage probability of both methods is similar, except when the nominal level is 0.99. In this case, the Jeffreys-Perks method tends to be anti-conservative. But again, both methods get closer to nominal as the sample size increases.

As for length, when Δ is 0, 0.2 or 0.4, the Jeffreys-Perks method is almost always shorter. There is quite a difference at the 0.99 level while at the 0.9 level, the difference is minimal. As the sample size increases, the lengths of both methods become similar.

When Δ is 0.6 or 0.8, the profile likelihood method becomes shorter, and again, a larger sample size causes both methods to be similar in length.

It appears that for small values of Δ (say less than 0.5 in absolute value) the Jeffreys-Perks method is superior since it results in shorter intervals,

while at the same time being conservative. When Δ is large (say greater than 0.5 in absolute value) the profile based method is superior since it is shorter and has an actual confidence level closer to nominal.

These results seem to suggest that if we believe Δ is small, then the Jeffreys-Perks method is best. Should we believe Δ is large, the profile likelihood method is best. However, if we have a large sample size, the magnitude of Δ is not as important, since both methods are similar. In this case, the Jeffreys-Perks method is preferred since it is simpler computationally. The Jeffreys-Perks interval can be calculated using explicit formulae, while the profile likelihood interval requires numerical methods to determine the endpoints of the interval.

Another point of note is that sample sizes of 5, 10 or 25 can be considered to be "small", while the performance of the intervals is still quite good. This seems to suggest the central limit theorem which both results depend on has a quick rate of convergence.

One last point to note is that similar intervals can be constructed for the odds ratio. However, simulation results suggested that the interval constructed along the lines of the Jeffreys-Perks method performed quite poorly, while a profile likelihood based interval for θ performs reasonably well (see Wei et al. [38]).

2.7 Summary

Some of the theoretical background of the RPWR has been presented. The limiting results have shown how the RPWR can increase the total number of successes. The likelihood function is the same as in the case of two independent samples, which makes calculating MLE's simple.

Two methods of constructing confidence intervals for the difference were introduced. Both methods are asymptotic and can be used in the usual case of two independent samples because of the limiting result in (2.21). As an example, the ECMO data was used to demonstrate the two methods.

Simulation results suggest there are differing results depending on the magnitude of Δ. Small Δ favours the Jeffreys-Perks method, while large Δ favours the profile likelihood method. When the sample size increases, the difference between the two methods becomes smaller, and in this case, the Jeffreys-Perks method is best for computational reasons.

Chapter 3

Further Details of the Jeffreys-Perks Method

3.1 Introduction

Some aspects of the Jeffreys-Perks interval will be examined. In particular, support for estimating a by $\tilde{a}(\kappa)$ will be presented, as well as justifying the choice of $\kappa = -\frac{1}{2}$. This support is based on some difficulties which can arise when estimating a probability. A suitable estimate for probabilities will be discussed as well as how it carries over to the RPWR case.

3.2 Estimating Probabilities

There are some difficulties in estimating the probability of an event, especially if the sample size is small. These difficulties exist for both the two sample independent case and for the RPWR, since the likelihood functions are proportional. The RPWR is to be used with small sample sizes, so this problem is particularly relevant.

First, consider estimating a single probability, p, from a sample of size n. Let S be the total number of successes. The simplest estimator of p is the ratio of successes to the number trials, $\hat{p} = S/n$. This also happens to be the MLE. At first glance, it would seem reasonable to use this estimate. After all, it is the observed proportion of successes. However, this is only a useful estimate if n is large. When n is small, this estimator has problems.

Consider the case where n is small, and $S = 0$. In this case $\hat{p} = 0$. This is a strong statement. Essentially, it says that observing a success is impossible. But if we have seen no successes in a small number of trials, this is by no means an indication that a success is impossible. Alternative estimators of p are clearly desirable.

An entire class of estimators can be derived by taking a Bayesian approach to estimating p. A crucial step in this approach is the selection of the prior density, $f(p)$. An obvious first choice is $f(p) = 1, 0 \leq p \leq 1$. This approach also coincides with an approach used by Laplace, known as Laplace's law of succession. The result is that the estimate of p is given by $(S + 1)/(n + 2)$.

This also arises from the Bayes postulate, which suggests assuming the prior probabilities should be equal if they are unknown.

A uniform prior is actually a special case of a more general prior. The beta distribution provides more flexibility. It is of the form $f(p) \propto p^{\kappa_1}(1-p)^{\kappa_2}$, $0 \leq p \leq 1$, $\kappa_1, \kappa_2 \geq -1$. Under squared error loss, we obtain as our estimate of p the posterior mean,

$$\tilde{p} = \frac{S + \kappa_1 + 1}{n + \kappa_1 + \kappa_2 + 2} \tag{3.1}$$

Often, it can be assumed that successes and failures convey the same amount of information. In this case, we take $\kappa_1 = \kappa_2 = \kappa$. Choosing $\kappa = 0$ gives the uniform prior. If $\kappa = -1$ the prior is improper. This prior leads to the MLE, so this choice for κ is no good.

Jeffreys and Perks both objected to the use of a uniform prior because of its connection with the Bayes postulate. They wished to find a prior which was more theoretically pleasing. Both created invariance theories for which the appropriate choice of κ would assign the same probability to a region of the parameter space even if the parameter is shifted. The result of Jeffreys and Perks is the same for the binomial case, leading to $\kappa = -\frac{1}{2}$, which is like a compromise between the MLE and the Bayes postulate. See Good [15] for even more on this topic. Jeffreys rule can also be interpreted in a rather interesting way. To use Jeffreys rule, simply take a prior proportional to the square root of the determinant of the information matrix (see Kass [20]).

Remember that this is true mainly for estimating probabilities from small samples. When the sample size is large, the MLE does not suffer from the aforementioned problem. However, the Bayesian estimators with a beta prior (of which the MLE is a special case) are all asymptotically equivalent. Hence, for large sample sizes it really doesn't matter which particular estimator of this form is chosen.

3.3 Estimating a in the RPWR

To use the method introduced by Beal, an estimate of $a = p_A + p_B$ is needed. Since this is a sum of two probabilities, the problems discussed in the previous section will come into play. The invariance rule of Jeffreys will be used to overcome this. To employ this rule, Fisher's information is required.

In the case of two independent samples, it is well known that the information matrix is given by

$$\mathcal{I} = \begin{pmatrix} \frac{n_A}{p_A q_A} & 0 \\ 0 & \frac{n_B}{p_B q_B} \end{pmatrix} \tag{3.2}$$

It then follows that the determinant of $\mathcal{I}$ is proportional to $(p_A q_A p_B q_B)^{-1}$, and Jeffreys' rule implies the prior should be proportional to $(p_A q_A p_B q_B)^{-\frac{1}{2}}$. This justifies the choice of $\kappa = -\frac{1}{2}$ for two independent samples.

To find the information matrix for the RPWR (see Rosenberger and Sri-

ram [31]), the derivative of the log-likelihood is required. Recall from (2.19) that

$$\mathscr{L}_n(p_A, p_B) = p_A^{S_A} q_A^{N_A - S_A} p_B^{S_B} q_B^{N_B - S_B} \prod_{i=1}^{n} p_i^{T_i} (1 - p_i)^{1 - T_i} \tag{3.3}$$

hence, the log-likelihood is,

$$\begin{aligned} \ell(p_A, p_B) = S_A \log p_A + (N_A - S_A) \log q_A \\ + S_B \log p_B + (N_B - S_B) \log q_B \\ + \log \left(\prod_{i=1}^{n} p_i^{T_i} (1 - p_i)^{1 - T_i} \right) \end{aligned} \tag{3.4}$$

The first derivatives of the log-likelihood are given by

$$\frac{\partial \ell}{\partial p_A} = \frac{S_A}{p_A} - \frac{(N_A - S_A)}{q_A} \tag{3.5}$$

$$= \frac{S_A - N_A p_A}{p_A q_A} \tag{3.6}$$

$$= \frac{\sum_{i=1}^{n} (X_i - p_A) T_i}{p_A q_A} \tag{3.7}$$

and

$$\frac{\partial \ell}{\partial p_B} = \frac{S_B}{p_B} - \frac{(N_B - S_B)}{q_B} \tag{3.8}$$

$$= \frac{S_B - N_B p_B}{p_B q_B} \tag{3.9}$$

$$= \frac{\sum_{i=1}^{n} (X_i - p_B)(1 - T_i)}{p_B q_B} \tag{3.10}$$

from which it is clear that

$$\frac{\partial^2 \ell}{\partial p_B p_A} = \frac{\partial^2 \ell}{\partial p_A p_B} = 0 \tag{3.11}$$

As for the diagonal elements of the information matrix, we have,

$$\mathbb{E}\left[\frac{\partial \ell}{\partial p_A}\right]^2 = \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - p_A)T_i}{p_A q_A}\right]^2 \tag{3.12}$$

$$= \frac{\sum_{i=1}^n \mathbb{E}\left[(X_i - p_A)^2 T_i\right]}{p_A^2 q_A^2} \tag{3.13}$$

where the cross products are zero in (3.13) since $(X_i - p_A)(X_j - p_A)T_i T_j$, $i \neq j$ is nonzero only when $T_i = T_j = 1$, in which case successive conditioning leads to $\mathbb{E}\left[(X_i - p_A)(X_j - p_A)T_i T_j | T_i = 1, T_j = 1\right] = 0$. Also,

$$\mathbb{E}\left[(X_i - p_A)^2 T_i\right] = \mathbb{E}\left[(X_i - 2X_i p_A + p_A^2)T_i\right] \tag{3.14}$$

$$= (1 - 2p_A)\mathbb{E}\left[X_i T_i\right] + p_A^2 \mathbb{E}\left[T_i\right] \tag{3.15}$$

$$= (1 - 2p_A)p_A \mathbb{E}\left[T_i\right] + p_A^2 \mathbb{E}\left[T_i\right] \tag{3.16}$$

$$= p_A q_A \mathbb{E}\left[T_i\right] \tag{3.17}$$

where $\mathbb{E}[X_i T_i] = p_A \mathbb{E}[T_i]$ since

$$\begin{aligned}
\mathbb{E}[X_i T_i] &= \mathbb{E}[\mathbb{E}[X_i T_i | X_1, \ldots, X_{i-1}, T_1, \ldots, T_{i-1}]] \\
&= \mathbb{E}[\mathbb{P}(X_i = 1, T_i = 1 | X_1, \ldots, X_{i-1}, T_1, \ldots, T_{i-1})] \\
&= \mathbb{E}[\mathbb{P}(X_i = 1 | X_1, \ldots, X_{i-1}, T_1, \ldots, T_{i-1}, T_i = 1) \\
&\qquad \times \mathbb{P}(T_i = 1 | X_1, \ldots, X_{i-1}, T_1, \ldots, T_{i-1})] \\
&= p_A \mathbb{E}[\mathbb{E}[T_i | X_1, \ldots, X_{i-1}, T_1, \ldots, T_{i-1}]] \\
&= p_A \mathbb{E}[T_i]
\end{aligned} \tag{3.18}$$

It now follows that

$$\mathbb{E}\left[\frac{\partial \ell}{\partial p_A}\right]^2 = \frac{\sum_{i=1}^{n} p_A q_A \mathbb{E}[T_i]}{p_A^2 q_A^2} \tag{3.19}$$

$$= \frac{\mathbb{E}[N_A]}{p_A q_A} \tag{3.20}$$

Similarly, it can be shown that

$$\mathbb{E}\left[\frac{\partial \ell}{\partial p_B}\right]^2 = \frac{\mathbb{E}[N_B]}{p_B q_B} \tag{3.21}$$

Expressions for $\mathbb{E}[N_A]$ and $\mathbb{E}[N_B]$ can be derived by obtaining a recursive relationship. The solution, as given in Rosenberger and Sriram [31] is quite complicated, and involves both p_A and p_B. This makes it very difficult to take a prior distribution according to Jeffreys' rule.

A simple solution is to estimate $\mathbb{E}[N_A]$ and $\mathbb{E}[N_B]$ by the observed values

of N_A and N_B. In doing so, the problem reduces to the case of two independent samples, and once again, the choice of $\kappa = -\frac{1}{2}$ seems quite reasonable.

It should also be noted that this choice of κ is the same as if p_A and p_B were assumed to have independent beta prior distributions, where κ was chosen to be $-\frac{1}{2}$ for both. This interpretation is also appealing, since it results from choosing the priors for p_A and p_B separately according to Jeffreys' Rule, then combining them by assuming independence of the priors.

3.4 Summary

This chapter examined the Jeffreys-Perks interval. The choice of $\kappa = -\frac{1}{2}$ was investigated. It was found that this was a good overall choice for the problem.

Estimating the probability of an event is always a tricky problem, especially with small samples. The MLE is not a very sensible estimator in such a case. Lapalce's rule of succession led to another estimator, but this estimator has been criticized, mainly for its connection with the Bayes postulate. The invariance theories of Jeffreys and Perks lead to an estimator which is a compromise between the two other estimators, and this estimator works better. Jeffreys' invariance rule can be easily stated.

The problem of estimating probabilities carries over into the RPWR for estimating a. Jeffreys' rule requires the information matrix, and for the RPWR it is the same as in the case of two independent samples, except

3.4 Summary

N_A and N_B are random, and are replaced by their expectations. As an approximation, $\mathbb{E}[N_A]$ and $\mathbb{E}[N_B]$ can be replaced by N_A and N_B, which justifies why $\kappa = -\frac{1}{2}$ is a good choice for the RPWR as well.

Chapter 4

Further Properties of the RPWR

4.1 Introduction

Some interesting results regarding the RPWR will be examined. One of the previously mentioned ideas is the quick convergence to normality which allows the use of the two methods for constructing confidence intervals. The theory behind this convergence is discussed as well as some ideas relating to the rate of convergence. Simulation results are also presented which demonstrate how quick the convergence is.

Some additional results regarding the allocation probabilities of the two treatments are also presented, including a comparison with some previously established properties for a deterministic version of the play the winner rule.

4.2 Rate of Convergence

The asymptotic normality of $\hat{p}_A$ and $\hat{p}_B$ was stated in section 2.3.1. In section 2.6 it was noted how there appears to be a quick rate of convergence. The details of this limiting result will now be examined. It is of interest to see what affects the rate of convergence.

4.2.1 Asymptotic Normality of the MLE's

A summary of main results in Wei et al. [38] is as follows. Let

$$Z_{ni} = n^{-\frac{1}{2}} \left[c_1 (X_i - p_A) T_i + c_2 (X_i - p_B)(1 - T_i) \right] \tag{4.1}$$

for $i = 1, \ldots, n$, $n \geq 1$ and constants c_1 and c_2 (to be used with the Cramér-Wold theorem). Then,

$$\sum_{i=1}^{n} Z_{ni} = c_1 n^{-\frac{1}{2}} (S_A - N_A p_A) + c_2 n^{-\frac{1}{2}} (S_B - N_B p_B) \tag{4.2}$$

is a martingale and $\{Z_{ni} : i = 1, \ldots, n, n \geq 1\}$ forms a martingale difference array. Theorem 3.2 of Hall and Heyde [16] can be used to show

$$\sum_{i=1}^{n} Z_{ni} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \eta^2) \quad \text{as } n \to \infty \tag{4.3}$$

where

$$\eta^2 = \frac{q_A q_B (c_1^2 p_A + c_2^2 p_B)}{q_A + q_B} \tag{4.4}$$

Hence, we have that every linear combination of the variables $n^{-\frac{1}{2}}(S_A - N_A p_A)$ and $n^{-\frac{1}{2}}(S_B - N_B p_B)$ converges to the corresponding linear combination of W_1 and W_2, where W_1 and W_2 are jointly bivariate normal with,

$$\mathbb{E}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbb{C}\text{ov}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \frac{q_A q_B}{q_A + q_B}\begin{pmatrix} p_A & 0 \\ 0 & p_B \end{pmatrix} \tag{4.5}$$

It now follows from the Cramér-Wold theorem that

$$n^{-\frac{1}{2}}(S_A - N_A p_A) \xrightarrow{\text{d}} W_1 \tag{4.6}$$

$$\text{and} \qquad n^{-\frac{1}{2}}(S_B - N_B p_B) \xrightarrow{\text{d}} W_2 \tag{4.7}$$

Recall from section 2.2.2 that

$$\lim_{n\to\infty} \frac{N_A}{n} = \frac{q_B}{q_A + q_B} \qquad \text{a.s.} \tag{4.8}$$

$$\lim_{n\to\infty} \frac{N_B}{n} = \frac{q_A}{q_A + q_B} \qquad \text{a.s.} \tag{4.9}$$

Applying Slutsky's theorem leads to the conclusion that $N_A^{\frac{1}{2}}(\hat{p}_A - p_A)$ and $N_B^{\frac{1}{2}}(\hat{p}_B - p_B)$ are asymptotically independent and jointly normally distributed with means of 0 and variances of $p_A q_A$ and $p_B q_B$ respectively.

4.2.2 Factors Which Affect the Rate of Convergence of the Central Limit Theorem

To examine the rate of convergence in the martingale central limit theorem the following notation will be useful. In particular, the conditional and unconditional variances are important. Let $\mathscr{F}_{ni}$ be the σ-algebra generated by $Z_{n1}, \ldots, Z_{ni}$. Define

$$\sigma_{ni}^2 = \mathbb{E}\left[Z_{ni}^2 | \mathscr{F}_{n,i-1}\right] \tag{4.10}$$

$$\bar{\sigma}_{ni}^2 = \mathbb{E}\left[Z_{ni}^2\right] \tag{4.11}$$

$$s_n^2 = \sum_{i=1}^{n} \bar{\sigma}_{ni}^2 \tag{4.12}$$

$$V_n^2 = \sum_{i=1}^{n} \frac{\sigma_{ni}^2}{s_n^2} \tag{4.13}$$

Bolthausen [11] derived several results on the rate of convergence based on these quantities. In particular the rates at which $|V_n^2 - 1|$ and $|\sigma_{ni}^2 - \bar{\sigma}_{ni}^2|$ go to zero affect the rate of convergence. Rinott and Rotar [26] also found that these quantities and some other subtle aspects of the dependency structure are the determinants of an upper bound for the rate of convergence.

For the RPWR,

$$Z_{ni}^2 = n^{-1}\left[c_1^2(X_i - p_A)^2 T_i^2 + c_2^2(X_i - p_B)^2(1 - T_i)^2 + c_1 c_2(X_i - p_A)(X_i - p_B)T_i(1 - T_i)\right] \quad (4.14)$$

$$= n^{-1}\left[c_1^2(X_i^2 - 2X_i p_A + p_A^2)T_i + c_2^2(X_i^2 - 2X_i p_B + p_B^2)(1 - T_i)\right] \quad (4.15)$$

$$= n^{-1}\left[c_1^2 T_i(X_i(1 - 2p_A) + p_A^2) + c_2^2(1 - T_i)(X_i(1 - 2p_B) + p_B^2)\right] \quad (4.16)$$

By definition, $\mathbb{E}\left[T_i|\mathscr{F}_{n,i-1}\right] = p_i$ and in a manner similar to equation (3.18) it can be shown that $\mathbb{E}\left[T_i X_i|\mathscr{F}_{n,i-1}\right] = p_i p_A$ and $\mathbb{E}\left[(1 - T_i)X_i|\mathscr{F}_{n,i-1}\right] = (1 - p_i)p_B$, hence

$$\sigma_{ni}^2 = n^{-1}\left[c_1^2(\mathbb{E}\left[T_i X_i|\mathscr{F}_{n,i-1}\right](1 - 2p_A) + \mathbb{E}\left[T_i|\mathscr{F}_{n,i-1}\right]p_A^2) + c_2^2(\mathbb{E}\left[(1 - T_i)X_i|\mathscr{F}_{n,i-1}\right](1 - 2p_B) + \mathbb{E}\left[(1 - T_i)|\mathscr{F}_{n,i-1}\right]p_B^2)\right] \quad (4.17)$$

$$= n^{-1}\left[c_1^2(p_i p_A(1 - 2p_A) + p_i p_A^2) + c_2^2((1 - p_i)p_B(1 - 2p_B) + (1 - p_i)p_B^2)\right] \quad (4.18)$$

$$= n^{-1}\left[c_1^2 p_i p_A(1 - p_A) + c_2^2(1 - p_i)p_B(1 - p_B)\right] \quad (4.19)$$

and

$$\bar{\sigma}_{ni}^2 = n^{-1}\left[c_1^2\mathbb{E}\left[p_i\right]p_A(1 - p_A) + c_2^2(1 - \mathbb{E}\left[p_i\right])p_B(1 - p_B)\right] \quad (4.20)$$

It now follows that,

$$|V_n^2 - 1| = \frac{1}{s_n^2}\left|\sum_{i=1}^{n}\sigma_{ni}^2 - \sum_{i=1}^{n}\bar{\sigma}_{ni}^2\right| \tag{4.21}$$

$$= \frac{1}{s_n^2}\left|\sum_{i=1}^{n}(\sigma_{ni}^2 - \bar{\sigma}_{ni}^2)\right| \tag{4.22}$$

$$= \frac{1}{s_n^2}\left|\sum_{i=1}^{n}\left(\frac{1}{n}\left[c_1^2 p_i p_A(1-p_A) + c_2^2(1-p_i)p_B(1-p_B)\right]\right.\right.$$
$$\left.\left. -\frac{1}{n}\left[c_1^2\mathbb{E}[p_i]\,p_A(1-p_A) + c_2^2(1-\mathbb{E}[p_i])p_B(1-p_B)\right]\right)\right| \tag{4.23}$$

$$= \frac{1}{s_n^2}\left|\sum_{i=1}^{n}\frac{1}{n}\left(p_i(c_1^2 p_A(1-p_A - c_2^2 p_B(1-p_B)) + c_2^2 p_B(1-p_B)\right.\right.$$
$$\left.\left. -\mathbb{E}[p_i]\left(c_1^2 p_A(1-p_A) - c_2^2 p_B(1-p_B)\right) - c_2^2 p_B(1-p_B)\right)\right| \tag{4.24}$$

$$= \frac{1}{ns_n^2}\left|c_1^2 p_A(1-p_A) - c_2^2 p_B(1-p_B)\right|\left|\sum_{i=1}^{n}(p_i - \mathbb{E}[p_i])\right| \tag{4.25}$$

i.e. $\left|\sum_{i=1}^{n}(p_i - \mathbb{E}[p_i])\right|$ is an important factor in $|V_n^2 - 1|$. Also,

$$\left|\sigma_{ni}^2 - \bar{\sigma}_{ni}^2\right| = \left|\frac{1}{n}\left[c_1^2 p_A(1-p_A)(p_i - \mathbb{E}[p_i]) + c_2^2 p_B(1-p_B)(\mathbb{E}[p_i] - p_i)\right]\right| \tag{4.26}$$

$$= \frac{1}{n}\left|c_1^2 p_A(1-p_A) - c_2^2 p_B(1-p_B)\right|\left|p_i - \mathbb{E}[p_i]\right| \tag{4.27}$$

i.e. $|p_i - \mathbb{E}[p_i]|$ is an important factor in $|\sigma_{ni}^2 - \bar{\sigma}_{ni}^2|$.

In order to examine the quantities in (4.25) and (4.27) expressions for p_i

and $\mathbb{E}[p_i]$ can be obtained recursively as

$$p_{i+1} = \frac{(2\alpha + \beta(i-1))p_i + \beta(T_i X_i + (1-T_i)(1-X_i))}{2\alpha + i\beta} \tag{4.28}$$

$$= \frac{2\alpha + \beta(i-1)}{2\alpha + i\beta} p_i + \frac{\beta(T_i X_i + (1-T_i)(1-X_i))}{2\alpha + i\beta} \tag{4.29}$$

and

$$\mathbb{E}[p_{i+1}] = \frac{2\alpha + \beta(i-1)}{2\alpha + i\beta}\mathbb{E}[p_i] + \frac{\beta(\mathbb{E}[T_i X_i] + \mathbb{E}[(1-T_i)(1-X_i)])}{2\alpha + i\beta} \tag{4.30}$$

$$= \frac{2\alpha + \beta(i-1)}{2\alpha + i\beta}\mathbb{E}[p_i] + \frac{\beta(p_A \mathbb{E}[p_i] + q_B(1 - \mathbb{E}[p_i])}{2\alpha + i\beta} \tag{4.31}$$

$$= \frac{2\alpha + \beta(p_A - q_B + (i-1))}{2\alpha + i\beta}\mathbb{E}[p_i] + \frac{\beta q_B}{2\alpha + i\beta} \tag{4.32}$$

Attempts to find a bound on $|p_i - \mathbb{E}[p_i]|$ using these recursive relationships have so far been unsuccessful. It would appear that finding useful bounds using only elementary techniques may not be possible. Tighter bounds may be potentially obtained using more sophisticated martingale techniques.

It should also be noted that more than just the central limit theorem is being used to prove the asymptotic normality of the MLE's. The last step in section 4.2.1 involves the convergence of $\frac{N_A}{n}$ and $\frac{N_B}{n}$ to v_A and v_B respectively. This convergence has been examined by Rosenberger [28]. Also, Rosenberger and Sriram [31] show how simulations suggest the convergence is very quick unless p_A or p_B is very large.

To help visualize the situation, further simulations were conducted. For the three sample sizes considered in chapter 2 ($n = 5, 10, 25$) and various

values of p_A and p_B, 10,000 outcomes of the RPWR were generated. For each outcome, the statistic

$$z = \frac{(\hat{p}_A - \hat{p}_B) - (p_A - p_B)}{\sqrt{\frac{p_A q_A}{N_A} + \frac{p_B q_B}{N_B}}} \tag{4.33}$$

was calculated. Histograms of this statistic for the given values of n, p_A and p_B are shown in figures 4.1 - 4.3 along with superimposed standard normal curves.

These simulations help visualize the quick convergence which was suspected back in section 2.6. When $n = 5$, the histograms are not exactly normal, but the deviation from normality is not too extreme. As n increases, the situation improves dramatically. For $n = 10$, most of the histograms are quite close to the normal curves, and when $n = 25$ almost all histograms appear to be normal.

The statistic z uses the true values of p_A and p_B, so in practice these must somehow be estimated. The best way around this problem in the case of two independent samples is to use the Jeffreys-Perks method. As previously discussed, this is also a good way to proceed for the RPWR. The Jeffreys-Perks idea has proved to be useful in a number of situations involving binominal responses, as seen in Piegorsch [25].

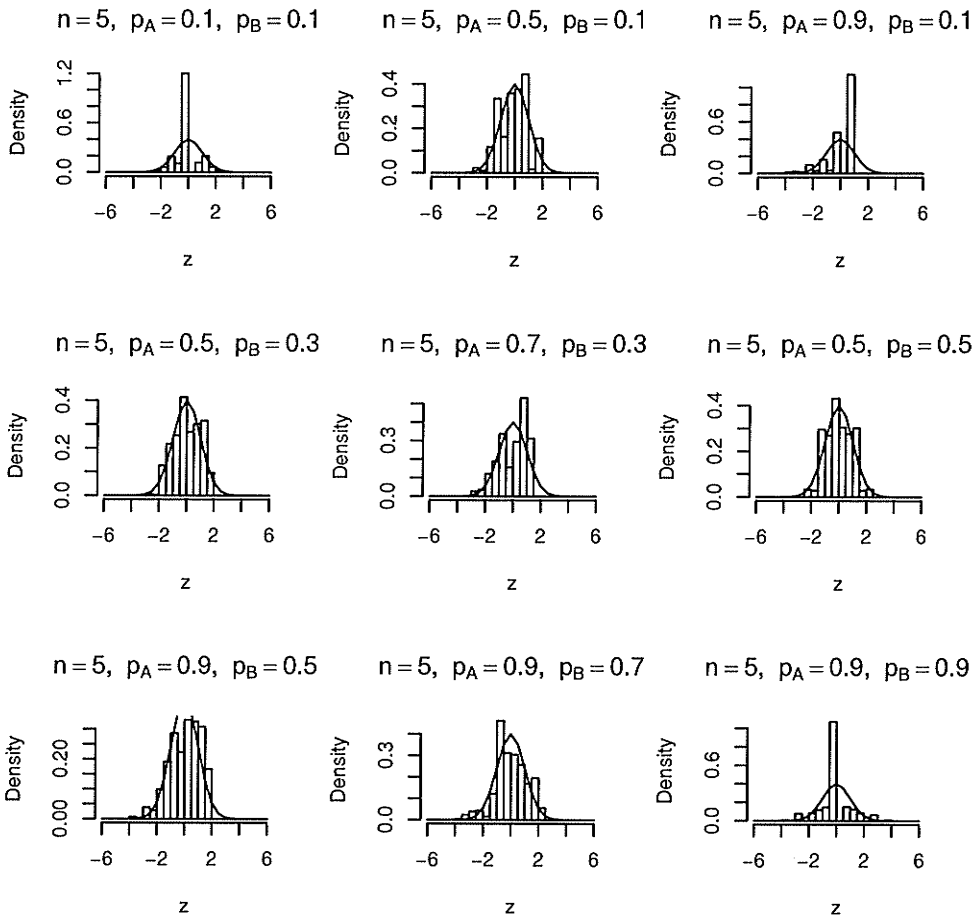

Figure 4.1: Histogram of z for $n = 5$

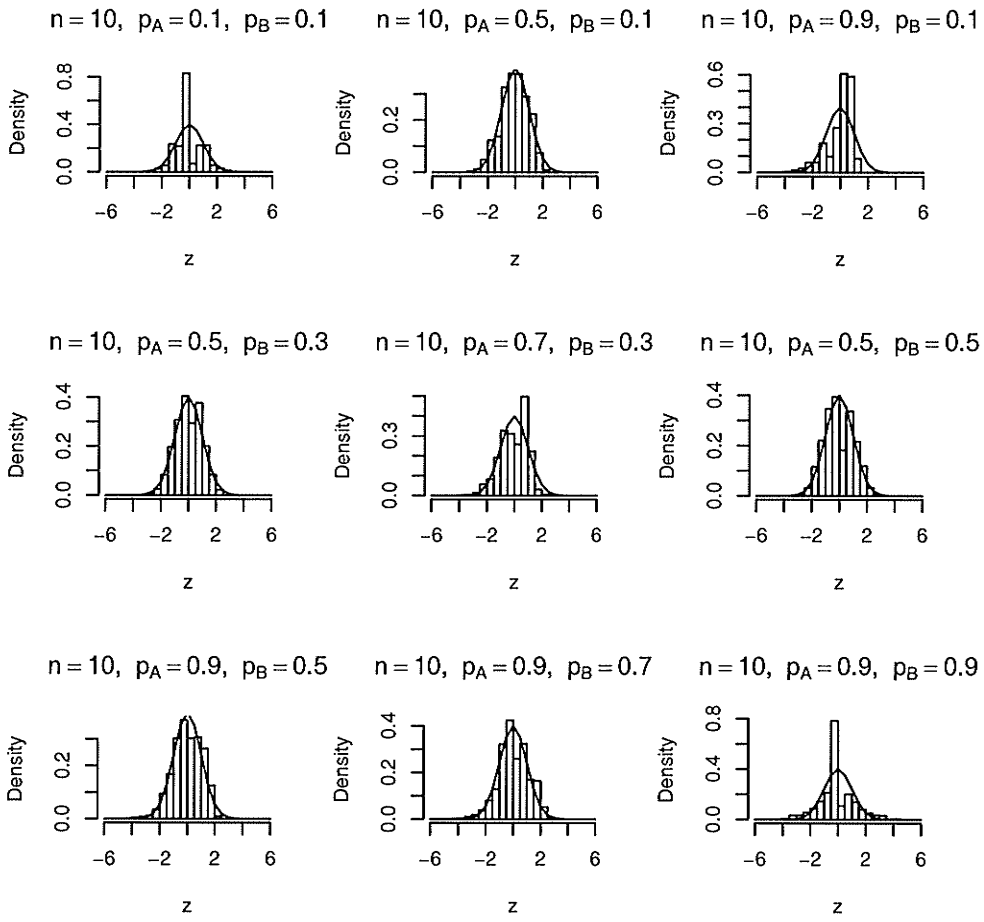

Figure 4.2: Histogram of z for $n = 10$

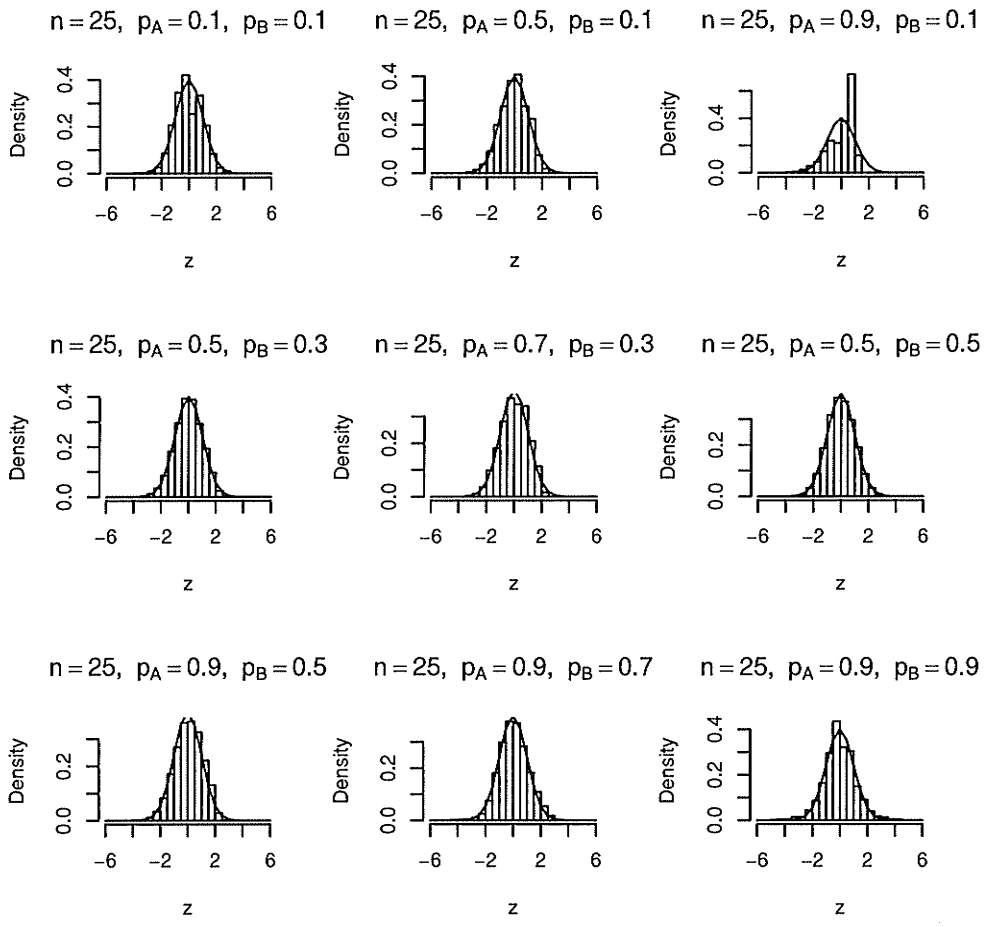

Figure 4.3: Histogram of z for $n = 25$

4.3 Properties of the Allocation Probabilities

The RPWR has desirable properties. Some have already been seen. For example, in section 2.2.2 it was shown that the asymptotic success rate of the RPWR is

$$\frac{p_A q_B + p_B q_A}{q_A + q_B} \tag{4.34}$$

which is larger than $\frac{1}{2}(p_A + p_B)$, hence the success rate of the RPWR is higher than the success rate for 50-50 randomization. There are other desirable properties relating to the allocation probabilities. These properties are also present for the deterministic play the winner rule (DPWR).

To employ the DPWR, the first patient is randomly assigned treatment A or B with equal probability. The next patient's treatment is completely determined by the previous patient's response. If the response is a success, then the next patient receives the same treatment as the previous patient. If the response is a failure, then the next patient receives the other treatment. In Wang and Pullman [35] the DPWR is used as an example of how adaptive designs provide desirable results. This paper also contained many results relating p_{i+1}, the probability of the next patient receiving treatment A after the previous i responses are known. Some of these properties are,

(a) $p_{i+1} > \frac{1}{2}$, $i \geq 1$, if $p_A > p_B$

$p_{i+1} < \frac{1}{2}$, $i \geq 1$, if $p_A < p_B$

$p_{i+1} = \frac{1}{2}$, $i \geq 1$, if $p_A = p_B$

(b) If $p_A + p_B > 1$, then

p_{i+1} is increasing in i when $p_A > p_B$

p_{i+1} is decreasing in i when $p_A < p_B$

p_{i+1} is constant in i when $p_A = p_B$

(c) $p = \lim_{i\to\infty} p_{i+1} = \frac{q_B}{q_A+q_B}$

(d) $p > \frac{1}{2}$, if $p_A > p_B$

$p < \frac{1}{2}$, if $p_A < p_B$

$p = \frac{1}{2}$, if $p_A = p_B$

(e) $\lim_{p_A\to 1} p = 1$

$\lim_{\Delta\to 0} p = \frac{1}{2}$

$\lim_{p_B\to 1} p = 0$

Properties (a) and (b) show how the probability of being assigned the superior treatment is more than 50% when using the DPWR. Properties (c), (d) and (e) show the limiting behavior of the DPWR. In particular, the limiting probability of using the superior treatment is greater than 50%. Also, if one treatment has a probability of success very close to 1, then the limiting probability of using that treatment will be close to 1, while if the two treatments are very similar (*i.e.* Δ is close to 0) then the limiting probability for each treatment is about 50%.

It would be of interest to know if these properties (or similar ones) hold for the RPWR. Due to the randomization involved with the RPWR, the above

properties do not hold for p_{i+1}. However, if the randomness is eliminated by taking the expected value, then all the properties hold for $\mathbb{E}[p_{i+1}]$.

Proposition 1. For the RPWR, the following properties hold.

(a) $\mathbb{E}[p_{i+1}] > \frac{1}{2}$, $i \geq 1$, if $p_A > p_B$

$\mathbb{E}[p_{i+1}] < \frac{1}{2}$, $i \geq 1$, if $p_A < p_B$

$\mathbb{E}[p_{i+1}] = \frac{1}{2}$, $i \geq 1$, if $p_A = p_B$

(b) $\mathbb{E}[p_{i+1}]$ is increasing in i when $p_A > p_B$

$\mathbb{E}[p_{i+1}]$ is decreasing in i when $p_A < p_B$

$\mathbb{E}[p_{i+1}]$ is constant in i when $p_A = p_B$

(c) $p = \lim_{i\to\infty} \mathbb{E}[p_{i+1}] = \mathbb{E}[\lim_{i\to\infty} p_{i+1}] = \frac{q_B}{q_A+q_B}$

(d) $p > \frac{1}{2}$, if $p_A > p_B$

$p < \frac{1}{2}$, if $p_A < p_B$

$p = \frac{1}{2}$, if $p_A = p_B$

(e) $\lim_{p_A\to 1} p = 1$

$\lim_{\Delta\to 0} p = \frac{1}{2}$

$\lim_{p_B\to 1} p = 0$

Proof. For parts (a) and (b) assume $p_A > p_B$. The case of $p_A < p_B$ is similar, while the case of $p_A = p_B$ is trivial. Part (b) will be proved first. From (4.32),

$$\mathbb{E}[p_{i+1}] = \frac{2\alpha + \beta(p_A - q_B + (i-1))}{2\alpha + i\beta}\mathbb{E}[p_i] + \frac{\beta q_B}{2\alpha + i\beta} \tag{4.35}$$

$$= B_{i+1}\mathbb{E}[p_i] + A_{i+1} \tag{4.36}$$

Also,

$$\frac{A_{i+1}}{1-B_{i+1}} = \frac{\frac{\beta q_B}{2\alpha+i\beta}}{\frac{\beta(q_A+q_B)}{2\alpha+i\beta}} \tag{4.37}$$

$$= \frac{q_B}{q_A+q_B} \tag{4.38}$$

$$= v_A \tag{4.39}$$

This is equivalent to saying $B_{i+1}v_A + A_{i+1} = v_A$. Since $p_A > p_B$, this implies $v_A > \frac{1}{2}$, and since $\mathbb{E}[p_1] = \frac{1}{2}$, $\mathbb{E}[p_1] < v_A$. Induction will be used to show $\mathbb{E}[p_{i+1}] < v_A$ for all $i \geq 0$.

Assume $\mathbb{E}[p_i] < v_A$ for some $i \geq 1$. Then

$$\mathbb{E}[p_{i+1}] = B_{i+1}\mathbb{E}[p_i] + A_{i+1} \tag{4.40}$$

$$< B_{i+1}v_A + A_{i+1} \tag{4.41}$$

$$= v_A \tag{4.42}$$

and $\mathbb{E}[p_{i+1}] < v_A$ for all $i \geq 0$.

From this,

$$\mathbb{E}[p_i] < \frac{A_{i+1}}{1-B_{i+1}} \tag{4.43}$$

$$\Leftrightarrow \quad \mathbb{E}[p_i] - B_{i+1}\mathbb{E}[p_i] < A_{i+1} \tag{4.44}$$

$$\Leftrightarrow \quad A_{i+1} + B_{i+1}\mathbb{E}[p_i] > \mathbb{E}[p_i] \tag{4.45}$$

$$\Leftrightarrow \quad \mathbb{E}[p_{i+1}] > \mathbb{E}[p_i] \tag{4.46}$$

for all $i \geq 1$.

Part (a) immediately follows from (b). Part (c) follows from the results of section 2.2.2 and switching of the limit and expectation is permitted by the dominated convergence theorem since $|p_{i+1}| \leq 1$ a.s. for all i. Parts (d) and (e) are the same as in the DPWR and are easy to verify. □

Interpretation of these results is similar to before. Parts (c), (d) and (e) are similar to the DPWR, so their interpretation is simple. They relate to the limiting behavior of the RPWR. Parts (a) and (b) again have the same interpretation as for the DPWR, except it is after taking the expected value. They relate to the fact that the superior treatment has a higher expected probability of selection, and this expected probability increases monotonically. It should also be noted that part (b) for the RPWR does not have the restriction of $p_A + p_B > 1$, while part (b) for the DPWR does have this restriction.

4.4 Summary

In this chapter, some interesting properties of the RPWR were investigated. The convergence to normality was examined, as well as its rate of convergence. Some key quantities were found which have an impact on this rate. Simulations also visualized and helped support the presence of quick convergence.

The allocation probabilities were also discussed. It was found that the

expected value of these probabilities hold some desirable properties. It was shown that the better treatment has a higher expected probability of selection, and this holds in the limit as well. These results are quite similar to those for the deterministic play the winner rule.

Chapter 5

Conclusion

5.1 What Has Been Achieved

The randomized play the winner rule is a good choice for a clinical trial where the patients responses are extreme and dichotomous. It has a significant advantage over traditional randomization since as the trial goes on, patients have a higher chance of receiving the superior treatment. Also, the RPWR has a higher asymptotic success rate.

While the design has these desirable features, one major drawback was how to analyze data arising from this design. Previous methods were met with controversy, or were extremely difficult from a computational point of view. The 'exact' method due to Wei et al. [38] certainly cannot be easily understood by the practitioner.

It is for this reason that other methods of analysis were required. The

profile likelihood method had been used before and worked quite well. However, while it is computationally simpler than the exact method, it is still a little complex. The Jeffreys-Perks method has been used in the case of two independent samples, and is quite good. Using some limiting results, use of the Jeffreys-Perks method for the RPWR can be justified. It turns out that it is still quite good in this case, while being much simpler computationally.

The profile likelihood method and the Jeffreys-Perks method were compared by simulation. The criteria was the actual coverage probability and the average length. It was found that the Jeffreys-Perks method performed best for small values of the difference, Δ, while the profile likelihood method was better for large Δ. As the sample size increases, the two methods becomes very similar. Since the Jeffreys-Perks method is easier to implement and is easier for the practitioner to comprehend it is a very valuable tool for analyzing data from a RPWR design.

The Jeffreys-Perks method also has some issues relating to the estimation of probabilities. The MLE is not very good for estimating a probability when the sample size is small. Bayesian ideas lead to a class of possible estimators. The estimator used in the Jeffreys-Perks method is the best out of this class for the case of two independent samples. For the RPWR an argument was presented where with reasonable approximations this results still holds.

The rate of convergence of the martingale central limit theorem was also of interest, since the two asymptotic methods performed well even for small sample sizes. Some key quantities relating to this convergence were derived,

as well as simulations which give a good visual of the quick convergence.

Some additional properties of the RPWR were also derived. It was shown that the allocation probabilities have many properties similar to the deterministic play the winner rule. These properties all show how the RPWR gives patients a greater chance of receiving the superior treatment.

5.2 Future Research

Adaptive designs and the randomized play the winner rule are still fairly new, so there are a number of open questions. One area of exploration relates to the Jeffreys-Perks method. It was shown how $\kappa = -\frac{1}{2}$ is a good choice for two independent samples, and for the RPWR this was justified by replacing $\mathbb{E}[N_A]$ by N_A and $\mathbb{E}[N_B]$ by N_B. Perhaps there is a better way of choosing the prior using the expressions for $\mathbb{E}[N_A]$ and $\mathbb{E}[N_B]$. There may be other approximations which lead to better intervals.

Another unanswered question deals with the rate of convergence. This was explored, but never fully answered. Elementary techniques have so far led to bounds that were not at all useful. It would appear that more sophisticated techniques, most likely based on martingale theory, are need to solve this problem.

Only asymptotic methods were considered because of their simplicity. The existing exact method is quite complicated. There may be other ways of constructing exact methods which may be better than the existing ones.

Or, maybe modifications could be made to make an exact method computationally more feasible, while still maintaining good results. This is another area open for exploration.

Choice of the design parameters, α and β, is another area of research. The simulations were all done with $\alpha = \beta = 1$, mainly since this is what has been already used in practice. But intuitively, it seems like a good idea to have a larger value of α so initially there can be patients on both treatments. For different values of α there may be better values of β. How to choose the optimal values of these parameters is still undetermined.

This was all done under the assumption of immediate responses. Delayed responses can be incorporated into the RPWR. How these methods carry over to such a case is yet to be answered. Steps in this direction have been taken by Bai et al. [6].

The randomized play the winner rule and adaptive designs in general provide a number of interesting research possibilities. Answering these questions is crucial for realizing the full ethical and statistical benefits to clinical trials.

References

[1] AGRESTI, A. (2003). Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research* **12** 3–21.

[2] AGRESTI, A. and CAFFO, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician* **54** 280–288.

[3] AGRESTI, A. and MIN, Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* **57** 963–971.

[4] ATHREYA, K. and KARLIN, S. (1967). Limit theorems for the split times of branching processes. *Journal of Mathematics and Mechanics* **17** 257–277.

[5] ATHREYA, K. and KARLIN, S. (1968). Embedding of urn schemes into continuous time markov branching processes and related limit theorems. *Annals of Mathematical Statistics* **39** 1801–1817.

[6] BAI, Z., HU, F. and ROSENBERGER, W. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *The Annals of Statistics* **30** 122–139.

[7] BANDYOPADHYAY, U. and BISWAS, A. (2003). Interval estimation in adaptive allocations using point estimates. *Calcutta Statistical Association Bulletin* **54** 31–44.

[8] BARTLETT, R., ROLOFF, D., CORNELL, R., ANDREWS, A., DILLON, P. and ZWISCHENBERGER, J. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* **76** 479–487.

[9] BEAL, S. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics* **43** 941–950.

References

[10] BEGG, C. (1990). On inferences from wei's biased coin design for clinical trials, with discussion. *Biometrika* **77** 467–484.

[11] BOLTHAUSEN, E. (1982). Exact convergence rates in some martingale central limit theorems. *Annals of Probability* **10** 672–688.

[12] CHAN, I. and ZHANG, Z. (1999). Test-based exact confidence intervals for the difference of two binomial proportions. *Biometrics* **55** 1202–1209.

[13] CHEN, X. (2002). A quasi-exact method for the confidence intervals of the difference of two independent binomial proportions in small sample cases. *Statistics in Medicine* **21** 943–956.

[14] FAREWELL, V., VIVEROS, R. and SPROTT, D. (1993). Statistical consequences of an adaptive treatment allocation in a clinical trial. *The Canadian Journal of Statistics* **21** 21–27.

[15] GOOD, I. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* The MIT Press, Cambridge, MA.

[16] HALL, P. and HEYDE, C. (1980). *Martingale Limit Theory and Applications.* Academic Press, New York.

[17] IBRAGIMOV, I. (1963). A central limit theorem for a class of dependent random variables. *Theory of Probability and its Applications* **8** 83–89.

[18] JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **186** 453–461.

[19] JEFFREYS, H. (1961). *Theory of Probability.* Clarendon Press, Oxford.

[20] KASS, R. (1989). The geometry of asymptotic inference. *Statistical Science* **4** 188–219.

[21] KATO, Y. (1978). Rates of convergence in central limit theorem for martingale differences. *Bulletin of Mathematical Statistics* **18** 1–8.

[22] MATSUMOTO, M. and NISHIMURA, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation* **8** 3–30.

[23] MATTHEWS, P. and ROSENBERGER, W. (1997). Variance in randomized play-the-winner clinical trials. *Statistics & Probability Letters* **35** 233–240.

References

[24] MIETTINEN, O. and NURMINEN, M. (1985). Comparative analysis of two rates. *Statistics in Medicine* **4** 213–226.

[25] PIEGORSCH, W. (1991). Multiple comparisons for analyzing dichotomous response. *Biometrics* **47** 45–52.

[26] RINOTT, Y. and ROTAR, V. (1998). Some bounds on the rate of convergence in the CLT for martingles. I. *Theory of Probability and its Applications* **43** 604–619.

[27] ROSENBERGER, W. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *The Annals of Statistics* **21** 2098–2107.

[28] ROSENBERGER, W. (1996). New directions in adaptive designs. *Statistical Science* **11** 137–149.

[29] ROSENBERGER, W. (1999). Randomized play-the-winner clinical trials: Review and recommendations. *Controlled Clinical Trials* **20** 328–342.

[30] ROSENBERGER, W., FLOURNOY, N. and DURHAM, S. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference* **60** 69–76.

[31] ROSENBERGER, W. and SRIRAM, T. (1997). Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference* **59** 309–319.

[32] SANTNER, T. and SNELL, M. (1980). Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 contingency tables. *Journal of the American Statistical Association* **75** 386–394.

[33] TAMURA, R., FARIES, D., ANDERSEN, J. and HEILIGENSTEIN, J. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association* **89** 768–776.

[34] TROENDLE, J. and FRANK, J. (2001). Unbiased confidence intervals for the odds ratio of two independent binomial samples with application to case-control data. *Biometrics* **57** 484–489.

[35] WANG, X. and PULLMAN, D. (2001). Play-the-winner rule and adaptive designs of clinical trials. *International Journal of Mathematics and Mathematical Sciences* **27** 229–236.

References

[36] WEI, L. (1988). Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika* **75** 603–606.

[37] WEI, L. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association* **73** 840–843.

[38] WEI, L., SMYTHE, R., LIN, D. and PARK, T. (1990). Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association* **85** 156–162.

[39] ZELEN, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of the American Statistical Association* **64** 131–146.

[40] ZHANG, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society Series B* **64** 281–294.

Appendix

Computer Code

Here are the R functions and code fragments used for the simulations and computing the confidence intervals.

A very important function is `rpwrsim()` which returns a list with elements `t` and `x` where `t` is the vector of the t_i's and `x` is the vector of the x_i's for a sample outcome from the RPWR with sample size `n`, success probabilities `pa` and `pb` and with the urn dynamics determined by `alpha` and `beta` (both defaulting to 1).

```
rpwrsim<-function(n,pa,pb,alpha=1,beta=1){
  a<-alpha
  b<-alpha
  r<-runif(2*n)
  c<-0
  x<-NULL
  t<-NULL
  for (i in 1:n) {
    c<-c+1
    if (r[c]<a/(a+b)) {
      t<-c(t,1)
      c<-c+1
```

```
      if (r[c]<pa) {
        a<-a+beta
        x<-c(x,1)
      }
      else {
        b<-b+beta
        x<-c(x,0)
      }
    }
    else {
      t<-c(t,0)
      c<-c+1
      if (r[c]<pb) {
        b<-b+beta
        x<-c(x,1)
      }
      else {
        a<-a+beta
        x<-c(x,0)
      }
    }
  }
  list(treatments=t,responses=x)
}
```

The function `profll.ci()` computes the profile likelihood based confidence interval. It takes as arguments a vector of t_i's, a vector of x_i's and optionally the significance level (which defaults to 0.95).

```
profll.ci<-function(t,x,sig=0.95) {
  #Log-Likelihood
  ll<-function(pb,delta,na,sa,nb,sb){
    log((delta+pb)^sa*(1-delta-pb)^(na-sa)
                              *pb^sb*(1-pb)^(nb-sb))
  }
  #Profile Log-Likelihood
```

```
profll<-function(delta,na,sa,nb,sb){
  n<-na+nb
  s<-sa+sb
  L.3<-n
  L.2<-(2*n-na)*delta-n-s
  L.1<-(delta*(n-na)-n-2*(s-sa))*delta+s
  L.0<-(s-sa)*delta*(1-delta)
  q<-(L.2/(3*L.3))^3-L.1*L.2/(6*L.3^2)+L.0/(2*L.3)
  p<-sign(q)*sqrt((L.2/(3*L.3))^2-L.1/(3*L.3))
  if (q==0) p<-sqrt((L.2/(3*L.3))^2-L.1/(3*L.3))
  if (q/p^3>1) q<-p^3
  a<-(pi+acos(q/p^3))/3
  pb.star<-2*p*cos(a)-L.2/(3*L.3)
  ll(pb.star,delta,na=na,sa=sa,nb=nb,sb=sb)
}
#Equation to be solved for upper and lower limits of the CI
ci.eq<-function(delta,na,sa,nb,sb,sig){
  2*(ll(sb/nb,sa/na-sb/nb,na,sa,nb,sb)
                  -profll(delta,na,sa,nb,sb))-qchisq(sig,1)
}
na<-sum(t)
sa<-sum(x*t)
nb<-length(t)-na
sb<-sum(x)-sa
if (na==0 || nb==0) return(c(-1,1))
else {
  rdelta<-seq(-0.9999,0.9999,length=200)
  testval<-sapply(rdelta,ci.eq,na=na,sa=sa,
                                        nb=nb,sb=sb,sig=sig)
  if (ci.eq(-0.9999,na,sa,nb,sb,sig)<0 &&
              ci.eq(0.9999,na,sa,nb,sb,sig)<0){
    low<--1
    up<-1
  }
  if (ci.eq(-0.9999,na,sa,nb,sb,sig)<0){
    low<--1
    up<-uniroot(ci.eq,
                c(max(rdelta[testval<0],na.rm=T),0.9999),
```

```
                na=na,sa=sa,nb=nb,sb=sb,sig=sig)$root
    }
    if (ci.eq(0.9999,na,sa,nb,sb,sig)<0){
      up<-1
      low<-uniroot(ci.eq,
                   c(-0.9999,min(rdelta[testval<0],na.rm=T)),
                   na=na,sa=sa,nb=nb,sb=sb,sig=sig)$root
    }
    if (ci.eq(-0.9999,na,sa,nb,sb,sig)>=0 &&
                ci.eq(0.9999,na,sa,nb,sb,sig)>=0){
      low<-uniroot(ci.eq,
                   c(-0.9999,min(rdelta[testval<0],na.rm=T)),
                   na=na,sa=sa,nb=nb,sb=sb,sig=sig)$root
      up<-uniroot(ci.eq,
                  c(max(rdelta[testval<0],na.rm=T),0.9999),
                  na=na,sa=sa,nb=nb,sb=sb,sig=sig)$root
    }
    return(c(low,up))
  }
}
```

The function `jp.ci()` computes the Jeffreys-Perks confidence interval. Its arguments are the same as those for `profll.ci()`.

```
jp.ci<-function(t,x,a=-1/2,sig=0.95) {
  n<-length(t)
  na<-sum(t)
  nb<-n-na
  sa<-sum(t*x)
  sb<-sum(x)-sa
  if (na==0 || nb==0) return(c(-1,1))
  else {
    pa.hat<-sa/na
    pb.hat<-sb/nb
    c<-qchisq(sig,1)
    a.hat<-na/(na+2*(a+1))*pa.hat + (a+1)/(na+2*(a+1))
             + nb/(nb+2*(a+1))*pb.hat + (a+1)/(nb+2*(a+1))
```

```
    b.hat<-pa.hat-pb.hat
    u<-(1/na+1/nb)/4
    v<-(1/na-1/nb)/4
    V<-u*((2-a.hat)*a.hat-b.hat^2)+2*v*(1-a.hat)*b.hat
    est<-(b.hat+c*v*(1-a.hat))/(1+c*u)
    moe<-sqrt(c*(V+c*u^2*(2-a.hat)*a.hat
                              +c*v^2*(1-a.hat)^2))/(1+c*u)
    low<-est-moe
    up<-est+moe
    return(c(max(-1,low),min(1,up)))
  }
}
```

The following code was used to obtain the simulation results. In the end, it creates a data frame named `sim.results` which contains the simulation results.

```
r<-10000
sim.results<-NULL
correct<-function(ci,pa,pb) {
  delta<-pa-pb
  c<-0
  if (ci[1]<delta && ci[2]>delta) c<-1
  return(c)
}
for (sig in c(0.9,0.95,0.99)){
  for (n in c(5,10,25)){
    for (pb in seq(0.1,0.9,by=0.2)){
      conf.prof<-NULL
      conf.jp<-NULL
      for (pa in seq(pb,0.9,by=0.2)){
        c.prof<-0
        c.jp<-0
        l.prof<-0
        l.jp<-0
        for (j in 1:r) {
```

```
          rpwrlist<-rpwrsim(n,pa,pb)
          ci.prof<-profll.ci(rpwrlist$treatments,
                             rpwrlist$responses,sig)
          ci.jp<-jp.ci(rpwrlist$treatments,
                             rpwrlist$responses,-1/2,sig)
          c.prof<-c.prof+correct(ci.prof,pa,pb)
          c.jp<-c.jp+correct(ci.jp,pa,pb)
          l.prof<-l.prof+ci.prof[2]-ci.prof[1]
          l.jp<-l.jp+ci.jp[2]-ci.jp[1]
        }
        conf.prof<-c.prof/r
        conf.jp<-c.jp/r
        length.prof<-l.prof/r
        length.jp<-l.jp/r
        sim.results<-rbind(sim.results,c(sig,n,pa,pb,
                 conf.prof,conf.jp,length.prof,length.jp))
      }
    }
  }
}
sim.results<-data.frame(sim.results)
names(sim.results)<-c("sig","n","pa","pb","conf.prof",
                      "conf.jp","length.prof","length.jp")
```

To generate the results for the histograms in section 4.2.2, the following code was run. It results in a vector named `stat` which contains all the generated outcomes for given values of `n`, `pa` and `pb`.

```
r<-10000
stat<-NULL
for (j in 1:r) {
  rpwrlist<-rpwrsim(n,pa,pb)
  t<-rpwrlist$treatments
  x<-rpwrlist$responses
  na<-sum(t)
  sa<-sum(x*t)
```

```
  s<-sum(x)
  nb<-n-na
  sb<-s-sa
  pa.hat<-sa/na
  pb.hat<-sb/nb
  if (na!=0 & nb!=0) {
    nextstat<-((pa.hat-pb.hat)-(pa-pb))
                  /sqrt(pa*(1-pa)/na+pb*(1-pb)/nb)
    stat<-c(stat,nextstat)
  }
}
```