The Randomized Play the Winner Rule in Adaptive Clinical Trials

By

Heather J. Prior

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

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

THE RANDOMIZED PLAY THE WINNER RULE IN ADAPTIVE CLINICAL TRIALS.

BY

Heather Jane Prior

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

Master of Science

Abstract

The randomized play the winner rule design, although popular in the statistical literature since its introduction by Wei and Durham in 1978, has been rarely used in medical trials. One well-known instance where the design was used in an adaptive clinical trial was in the controversial ECMO trials. The ECMO trials randomized their extremely ill infant patients into one of two treatments. The controversy surrounding the trials resulted from the ethical concerns of randomizing patients with life-threatening illnesses. The randomized play the winner rule attempts to distinguish these ethical issues by making a compromise between information gathering and immediate payoff to the patient. The primary goal of the design is to maximize the number of patients who receive the superior treatment without sacrificing the power of the test so that a valid conclusion can still be obtained. Through simulation studies, the expected proportion of patients allocated to the superior treatment is compared for various randomized and adaptive designs. Furthermore, the monotonicity of the randomized play the winner design is studied through modifications to the parameters in the classical randomized play the winner rule design. Finally, comparisons with real clinical trials will attempt to improve upon previous results.

Acknowledgements

I would first like to thank Dr. Xikui Wang of the University of Manitoba for his wisdom, patience, encouragement and support throughout the development of this thesis. Without his helpful comments and motivation, this paper would not have been possible. I am also grateful for the generous research grants I received from Dr. Wang. I wish to thank the faculty and staff of the Department of Statistics for planting the seed of my interest in statistics and nurturing it throughout my education. I would also like to thank committee members Dr. Dean Slonowsky of the Department of Statistics at the University of Manitoba and Dr. Lisa Lix of the Department of Community Health Services at the University of Manitoba.

To my family and friends, thanks for being there; your love and support is appreciated more than you know. Last, but not least, I would like to thank my father for always instilling in me that as long as I strive to do my best, there is no limit to what I can achieve.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background Information

The science of medicine is one of the greatest accomplishments of mankind. Medical knowledge has expanded faster in the past century than it ever has before. The study of medicine, which pursues to prolong peoples' lives as well as reduce their pain and suffering, is both rewarding and endless. Medical research plays an integral part in improving the knowledge base that doctors have to help cure their patients.

Before there was medical research or clinical trials, there was the scientific method. The scientific method originated in the seventeenth century and was developed by great scientists including Galileo, Francis Bacon and Sir Isaac Newton [15]. The process of the scientific method involves making observations about a process or item under study without any prior assumptions. Randomization was incorporated into the scientific method by Sir Ronald A. Fisher in the early twentieth century. Fisher was conducting agricultural experiments when he had the innovative idea of allocating different conditions to plots of land "deliberately at random" [12], and thus, the first randomized experiment was conducted.

Around the same time that Fisher was carrying out randomized comparative experiments, Karl Pearson, considered another founder of statistics, was developing the

concept of a control [17]. He reasoned that in order to study the full effects of an anti-typhiod inoculation, only half of those who wished to receive the inoculation should actually receive it, while the other half should receive a placebo. Furthermore, Pearson felt that for the control to be most effective, the two groups should be as similar as possible in terms of age, sex and other covariates that may bias the trial.

Today, randomized experiments involving controls to reduce bias are used in many disciplines such as chemistry, biology and psychology. In medical clinical trials, randomization is considered a necessity. Almost all medical experiments must incorporate some sort of randomization if their results are to be considered valid. In most cases, medical trials use equal randomization, also called 50-50 randomization, where an equal number of patients receive each of the treatments and the treatment groups are most likely balanced with respect to the patients' characteristics. This is often considered the only way to allocate patients to treatments in order to obtain a valid statistical conclusion, and in most cases it is the best way.

When patients are extremely ill, conducting a medical experiment using equal allocation can be considered unethical. If one experimental treatment or pharmaceutical holds promise of performing better than the conventional treatment, then preventing half of the deathly-ill patients under study from receiving the new treatment is indeed distressing to them. Some patients will take their chances in hopes of receiving the new treatment. However, just because the treatment is new and experimental does not mean that it is superior. In a randomized clinical trial, the superior treatment is not determined until the end of the study and in some cases there may be insufficient evidence to detect a difference among treatments. So, the fifty per cent of patients hoping to receive a

treatment better than the conventional one are not only disappointed, but may also have been randomized to receive an inferior treatment.

For very ill patients enrolled in clinical trials, there are more humane and ethical alternatives to equal randomization. Adaptive clinical trials incorporate both randomization and information from the trial to identify the superior treatment under study and to maximize the proportion of patients who receive that treatment. Adaptive designs are a more ethical alternative to randomized clinical trials, particularly equal randomization, because they allocate more than fifty per cent of patients to the superior treatment. For critical patients, this may mean more lives saved.

There have been cases in the history of clinical trials that would have warranted use of adaptive designs over randomized clinical trials simply because of the possibility of more lives saved. The use of equal randomization was unethical. One such instance is the UK ECMO trial of 1993-1995 [25]. Extracorporeal membrane oxygenation (ECMO) was tested in a clinical trial for treatment of persistent pulmonary hypertension in newborns. Due to the extreme severity of the illness, randomly assigning the young patients to either ECMO or the conventional treatment limited the proportion of babies who could receive the ECMO treatment to only half. An adaptive design would have been a more ethical choice for the study. A similar case was in the Antiviral Zidovudine Treatment (AZT) trial of 1991-1994 [5]. The pharmaceutical AZT was given to Human Immunodeficiency Virus (HIV) positive pregnant women to prevent the spreading of the virus to their unborn child. Again, a completely randomized design was used in the trial when an adaptive design would have been more appropriate and could have possibly prevented more HIV positive babies.

Although adaptive designs are not suitable for all clinical trials, in cases where the usual 50-50 randomization is unethical, adaptive designs are an attractive alternative. They are currently not widely used in the medical community, but hopefully will become as acceptable as concepts such as the scientific method, the randomized trial and the control group are in present day.

1.2 Summary

There are both advantages and disadvantages to adopting an adaptive design instead of a randomized design in a clinical trial. Obviously, any ethical worry is diminished when an adaptive design is used, especially when patients are critically ill. Since more patients are allocated to the superior treatment, more have the benefit of that treatment and a better chance of improving their health. Randomized clinical trials, however, are the most powerful choice for detecting a difference between treatments. Certainly, they have achieved hallowed status [3] among experimental designs in the field of medicine due to their ability to draw a statistically significant conclusion and to reduce bias.

In cases where patients are quite ill, the objective of using a randomized design with a powerful statistical conclusion plays second to maximizing the number of patients who receive the superior treatment in the trial. Adaptive clinical trials are ideal in such cases because they can be thought of as a compromise between the goal of gathering information for future patients and the goal of immediate payoff to current patients. Thus, adaptive designs partially satisfy both collective ethics and individual ethics. Chapter two describes in detail the ethical concerns of randomization in medical trials and how various adaptive designs overcome these concerns. Also, historic cases of

unethical randomization including the UK ECMO trial and the AZT trial will be discussed.

The main advantage of adaptive designs is that they maximize the number of patients allocated to the superior treatment without significant loss of power. This is a key idea that will be demonstrated in this thesis through simulation studies. In particular, Wei and Durham's Randomized Play the Winner Rule (RPW) [30] will be compared with 50-50 randomization and a sequential maximum likelihood procedure specified by Neyman Allocation [18]. Chapter three discusses the basics of random number simulation, and then uses simulation results to compare the aforementioned designs and explore the monotonic properties of the randomized play the winner design. Then, simulations of real data will be compared with actual experiments to see if the use of adaptive design methods could have improved the number of patients who received the superior treatment without noticeable loss of power. Finally, some recommendations are made as to when it is appropriate to use certain adaptive designs.

The goal of this thesis is to show that a randomized design is not always the best choice for a clinical trial, and to demonstrate this by way of simulation studies. It will be shown that adaptive clinical trials, and the randomized play the winner rule in particular, are simple to integrate into a medical trial and carry out, and provide results nearly as powerful as randomized clinical trials without overt ethical question.

Chapter 2

Randomized Trials and Adaptive Designs

2.1 Randomization: Ethics and Practice

To conduct an effective experiment is no menial task; it takes planned, purposeful and methodological research. To conduct an effective medical experiment, however, is something quite more challenging. The medical community, including physicians and statisticians, believe that for clinical trial to be valid, it must include randomization. Sir Ronald A. Fisher, often named "the father of statistics," developed the concept of the comparative randomized experiment [12]. Fisher believed that randomization was essential to an experiment. Without randomization, he reasoned, an experiment would be reduced to an observational study. A randomized experiment improves on an observational study due to the fact that at the end of a successful randomized trial, it may be possible to establish a cause and effect relationship, while at the end of an observational study, only an association can be concluded.

Randomization reduces the possibility of systematic bias in a trial. Without randomization or comparison of treatments, experimental results can be negatively affected by the experimental design, the selection of patients or the placebo effect. Bias, or the systematic favouritism toward one outcome, may ensue. In a medical trial, there are usually two groups of patients under study, the treatment group and the control group.

Randomly placing subjects into one of the two groups reduces the probability that the response of one group may dominate over the other due to chance, or due to confounding factors. It also helps in balancing out unknown covariates, or factors that have the possibility to bias final results. Once treatment groups are balanced, analysis can be conducted without being concerned about systematic bias. Furthermore, using randomization in clinical trials simplifies the method of analysis and also guarantees the validity of the conclusion. Probabilities used in the random assignment of treatment groups can also be used in the comparison of the treatments. Another feature of randomization is that these probabilities remain constant throughout the entire trial. Once a comparison is made, and if there is a statistically significant difference between the two treatments, a cause and effect relationship can be established. Thus, randomization is necessary in medical trials to reduce bias and make the leap from association to causation.

For obvious reasons the medical community trusts randomized clinical trials as the gold standard of medical experimentation [3]. Randomized trials have been used for decades, and thus are usually positively received and well understood. Many clinicians are quite sceptical of new treatments or pharmaceuticals developed under unconventional means, especially if randomization was not used. Randomized clinical trials do however have their weaknesses. If the trial is a double-blind experiment, where neither the physician nor the patient knows which treatment they are to receive, then both parties give up control of the treatment of the patient. Moreover, if the allocation is 1:1, physicians may be apprehensive about relinquishing control of treatment, or they may decline to put certain patients in the trial for fear of having them receive the placebo. In these cases the clinical trial cannot be considered truly random since the doctor has

shown favouritism with his patients. There are other cases where doctors only enrol severely ill patients into trials whom he feels have nothing to lose by experimenting with a new treatment. Despite these disadvantages, randomized clinical trials remain the undisputed primary choice for experimentation in the medical field.

The outcomes of medical trials have the potential to affect many human lives. Not only is there a drive to perform the trial well in the traditional sense, there is also a drive to perform the trial well in an ethical sense. There are many ethical issues associated with medical experiments. The main ethical dilemma involves respecting the conflicting issues of individual ethics and collective ethics [4]. Both the wishes of the individual patient and future patients must be incorporated into the trial. Thus, to address the conflict a compromise must be made between what is best for the individual, and what is best for society.

It is often difficult to fully satisfy both individual and collective ethics simultaneously. And, often those who participate in a medical trial do not agree as to which side of the ethical dilemma should prevail. Obviously, patients in a trial care more for their well being than any patients after a trial. Moreover, patients trust that their caregivers will also do what is best for their individual needs. Physicians, who are bound by the Hippocratic Principle, "To do no harm," [32] must consider the interests of their patient before the interests of society. This concept has been termed the personal care principle [8]. Putting the needs of their patient above all else often makes it difficult for a doctor to participate in a clinical trial. They are torn between the care of their current patients and finding a better method of caring for future patients. Clinicians may wish to discover new, more effective ways of treating illness, but to do this means that their current patients may have to be randomized in a clinical trial. Randomization comes with

the price of forsaking the expectation that the doctor will always give his patient what he believes is the best treatment. In other words, the physician must forgo the personal care principle.

The only case where the personal care principle is not violated is if the physician is truly undecided between a current treatment and a new, experimental treatment. Here, the physician is in a state of equipoise [7], or mental suspense. He has no prior knowledge that one treatment is superior over the other. If equipoise exists between two treatments, then a doctor's fears of violating his patient's trust can be subdued, provided that the patients involved in the trial have been fully informed of the risks associated with the trial. Also, sick patients who volunteer for the trial may have nothing to lose; they will either receive the new treatment, which has not been proven to be any better or worse than the current one, or they will receive the standard treatment, which would have been the case anyway if they had not registered in the trial. Informed consent [16] is essential if the personal care principle is to remain significant. In cases where the physician has a pre-determined state of mind as to which treatment will most benefit their patients, true equipoise does not exist and a randomized clinical trial is unethical.

Those who design and carry out the randomized clinical trial, usually statisticians, are mostly concerned with having a well-balanced experiment, resulting in a high power and a valid conclusion. It is important to note that a trial need not necessarily have 1:1 allocation to have a high power. Often it can be difficult to get volunteers to a trial if there is only a fifty per cent chance of receiving the new treatment. If more subjects are placed in the treatment group, say using a 3:1 allocation, then more patients may receive the new, better treatment. Only the use of a balanced design will maximize the power of the test. A significant conclusion can still be reached without using 1:1 allocation if the

probabilities of selection are adjusted at the beginning of the trial. Statisticians strive for a significant conclusion of the trial since it could lead to new treatments. These new treatments may aid in curing future patients and while this directly benefits society as a whole, does little for the patients who were in the trial.

Ethicists in most cases are involved with a medical trial to some extent. They may be torn as to which category deserves more weight ethically, the individual or the society. It is impossible to say for certain which side deserves more consideration: collective ethics or individual ethics [14]. One factor that may make the choice between collective ethics and individual ethics easier is the severity of the disease under study. The more severe the disease, the more urgent the need is to care for the individual subject before the group. Yet, the ethical dilemma still comes into play since a severe disease beckons for a faster, more humane cure for all of society to benefit from. Some ethicists believe that if a true state of equipoise exists, then a physician is acting within the best interests of both current and future patients. One can further argue that given a state of equipoise, satisfying the collective need of society to find a cure for a disease also satisfies the individual need of future patients to have the best treatment. Only after a clinical trial has been carried out can a physician make a fully informed decision as to the best treatment for his patient.

2.2 Historical Cases: Unethical Randomization

There are cases where 1:1 allocation in a randomized clinical trial was unethical, but was used nonetheless. The classic example is of the British ECMO (Extracorporeal membrane oxygenation) trial [25]. ECMO is currently used to treat newborn babies inflicted with persistent pulmonary hypertension (PPHN), a condition where the baby has

decreased blood flow through the lungs, resulting in insufficient oxygenation of the blood. Before the onset of ECMO, the conventional therapy for PPHN was intensive ventilatory support. This treatment has a very low survival rate for newborns with severe forms of the disease. ECMO had been used to treat adults with acute respiratory failure, but failed to show improvements over conventional treatments. In the late 1970's ECMO began to be used as an experimental treatment on infants with PPHN. Some early studies showed promising results of the ECMO treatment on newborns, but these were observational studies since no control group was used as a comparison, only historical data [2, 11, 31].

In 1985, Dr. Robert Bartlett and his colleagues at the University of Michigan performed an adaptive clinical trial investigating the benefits of ECMO over the conventional treatment [1]. Their study incorporated the randomized consent design developed by Zelen [34] in 1979. Treatment assignments were based on the randomized urn designs proposed by Wei and Durham in 1978 [26]. This study resulted in eleven babies receiving ECMO and only one receiving the conventional treatment. Of these patients there was only one death, the baby who received the conventional treatment. This resulted in a 0% failure rate for ECMO and a 100% failure rate for the conventional treatment. Some claimed that this was good evidence that the survival rate for patients treated with ECMO was significantly higher than patients treated with the conventional medical therapy. Many, however, were not convinced because the study did not have balanced sample sizes. Obviously, it is difficult to compare the results of one patient to a group. Other small sample studies that were done include a 2-stage sequential trial in 1989 by O'Rourke *et al.* [13] and a randomized trial in 1994 by Gross *et al.* [9]. Both studies should an approximate 40% increase in the failure rate whenever the conventional

treatment was compared with ECMO. Many in the medical and statistical communities were still unconvinced that ECMO was a statistically significant improvement other the usual therapies for PPHN.

In 1993, a large-scale study of ECMO was undertaken by physicians in the UK [25]. They believed that performing a randomized medical trial was the only way to finally prove the treatment's superiority. The reason the trial was not only controversial, but also unethical was 50-50 randomization was used in the trial. Due to the severity of the disease, and the presence of evidence that one treatment may be inferior, randomizing infants in a trial brings up ethical concerns. The doctors involved were basically giving half of the newborns a better chance of survival. The results of the trial were as follows: 30 out of 93 (32.3%) babies who received ECMO died, and 54 out of 92 (58.7%) babies who received the conventional treatment died. The findings of the clinical trial proved once and for all that ECMO was indeed a superior treatment for newborn babies afflicted with PPHN. Unfortunately, 84 babies died in this trial. If another method of clinical trials had been adopted other than randomization, maybe some lives would have been saved with the same end result. Clearly, many newborns died unnecessarily.

Given the stress placed on the parents of a sick newborn, the physicians thought it would be best to use randomized informed consent [27]. This is another case of unethical randomization. All involved doubtlessly believed that they were acting in the best interests of the patients, and their parents. By using randomized informed consent, the physicians would only inform the parents of the newborn selected to receive ECMO treatment that their baby was participating in a clinical trial. The parents of the newborn receiving the conventional therapy would not be informed that their baby was involved in a trial, or that he or she had been randomized to receive the conventional therapy. Some

may argue that there is no harm done keeping one half of the subjects in the dark if it is the half that receives the usual treatment since they would have anyway if they had not participated in the trial. The ethical problem lies with the fact that collective ethics prevails when informed consent is randomized. The physicians put the welfare of society above the welfare of their patient when they focus on the trial instead of their patient. As doctors, they are sworn to put the welfare of their patient first. If they inform the patient as to the aspects of the trial and share the decision making process with them, they alleviate the tension between individual and collective ethics. In a sense, once the patient is fully involved in his care, and has made a fully informed choice as to whether to participate in a trial or not, then the pressure is off the physician to fulfill his role as a researcher and he can now fully focus on his role as a care-giver.

Another unforgettable case in the history of randomized clinical trials is that of the 1994 trial concerning the drug Antiviral Zidovudine Treatment (AZT), used for reducing the risk of transference of HIV from mother to unborn child [5]. As was the case with the ECMO treatment, there was strong, but not statistically significant evidence that AZT was very effective in preventing the spread of HIV. A trial was conducted using 50-50 randomization in order to test whether AZT would be more effective than a placebo. 476 patients were involved in the trial, 238 in each of the treatment and placebo groups. After the birth of each baby, an HIV test was administered in order to determine if the newborn had contracted the virus. The final results of the trial were that 60 (25.2%) newborns had contracted HIV in the placebo group, while only 20 (8.4%) newborns in the treatment group had contracted the virus. This amounted to a statistically significant difference between the two groups. Thus, both researchers and physicians had very strong evidence that AZT prevents the transference of HIV from mother to unborn child,

but the cost of that result was high. The 60 infants whose mothers received a placebo were born with HIV and would most likely die of AIDS at a young age. More infants could have been saved if 1:1 allocation had not been used in the trial. This is another circumstance where the greater good of society was put above the welfare of unborn children.

2.3 Alternatives to Randomized Clinical Trials

In clinical trials involving the research of treatments for life-threatening illnesses, such as the above examples of ECMO and AZT, randomization is often an unethical choice. There are other tools available to test the effectiveness of a new drug or treatment. Historical data, sequential clinical trials and adaptive clinical trials are all alternatives to randomized clinical trials. Although these methods do come with some disadvantages, they are much more ethically sound. This is most important when dealing with severe diseases, as all parties involved wish to give the patient the best possible chance for survival. If desperately ill patients are involved in a clinical trial, it is to their benefit that they have a better than a 50-50 chance of receiving a superior treatment.

Historical data can be a powerful tool in a trial if used properly. The abundance of historical data on any treatment is readily accessible for use in a clinical trial. The data can be compiled for use as a historical control. The advantage of using historical data as a pseudo-control group instead of live subjects is that patients do not have to be tested on what is already known as fact. For example, if a treatment has been in use for say, twenty years, and the results of this treatment have been well documented, accounting for differences in age, race or gender, then there is no reason to further test patients of the results of this treatment. The historical data can be used as a control in a clinical trial

where live subjects are tested on with a new, hopefully more effective treatment. The results of the trial are compared with the historical data to reach a permissible conclusion. Both time and money are saved when historical controls are used. The main advantage, however, is that patients can be subject to a hopefully superior treatment without the ethical worry of other patients being randomized to an inferior, conventional treatment. The choice of collective ethics over individual ethics does not have to be made.

Historical controls do have disadvantages. Historical data can be biased by time, location, race, gender, age, or other unknown factors. If the historical data is quite old, using the data as a control group may not be feasible. For instance, patients from ten years ago may not be comparable with present patients due to differing lifestyles. Historical data may be available, but the data for the control group may be incomparable with the treatment group. Data may be available for patients in the United States, while current patients live in Canada, or the data involves white male subjects, but the current trial is comprised of Asian women. Historical controls may be useful when they are available, but control groups must be matched as close as possible with treatment groups if systematic bias is to be minimized.

Sequential clinical trials are special case of adaptive designs. They differ from randomized clinical trials in that they incorporate a stopping rule. After a pre-determined number of patients have entered the trial testing is undertaken to determine if a statistically significant difference has been established between the two treatments under study. If a substantial difference has been observed from the accumulated information, then the trial ends. Otherwise, experimentation continues until the next set of experimental units has been run. The advantage of a sequential clinical trial over the usual randomized trial is that the trial can be stopped if the treatment shows clear signs of

being superior (or inferior) to the conventional treatment. This prevents unnecessary further experimentation. The problem with sequential trials is that 50-50 randomization is still used in the trial. In cases of acute, critical illness, full randomization of subjects is nonetheless considered unethical. Although sequential trials can be considered more efficient, they may be inappropriate in experiments concerning serious illnesses.

Adaptive clinical trials, although distinct from randomized clinical trials, still incorporate random assignment. The difference is the way in which the randomization is integrated into the trial. Unlike randomized clinical trials, adaptive clinical trials (or ACTs) make use of the information collected during the trial and use that information for the progression of the trial. This process has both advantages and disadvantages. The main problem with adaptive clinical trials is concerned with their adaptive nature: because the data collection process is adaptive, the treatment assignment probabilities change throughout the trial. This can make inference of the ACT the end of the trial quite difficult. So, the question is when is the extra work worthwhile? In other words, when are adaptive clinical trials an improvement on randomized clinical trials? To answer this, one must understand the main goal of ACTs. Adaptive clinical trials aim to minimize the risk of a patient receiving an inferior treatment. They are a trade-off between a randomized clinical trial and a non-randomized trial. Randomized clinical trials (or RCTs) simply intend to collect data on the success or failure of two or more treatments and determine which is superior. ACTs go a step further by identifying which treatment is performing best during the trial and allotting a higher proportion of subjects to the currently identified better treatment, (and thus a lower proportion of subjects to the inferior treatment(s)). The expectation is that as many patients as possible will receive the best care available. ACTs enable the clinician to perform a clinical trial and come to

a sound conclusion, but to do it in a more ethical way. Thus, adaptive clinical trials are a compromise between doing what is best for the patient and what is best for society.

Although randomized clinical trials will always be integral to medical experimentation, there are circumstances when they may not be morally feasible. Particularly, under cases of life-threatening illness, randomized clinical trials may not be ethically viable. When patients are very ill, even terminally ill, they will care little for expanding society's knowledge by taking part in a medical experiment. Their primary concern, and that of their doctor, will be to have the best possible care. The focus shifts from research to healthcare, from researcher and subject to doctor and patient. Randomization may be justified if the patient is fully informed about all the risks associated with the trial and the different treatments they may receive. Informed consent provides a moral justification for randomized clinical trials. Unfortunately, under desperate or terminal medical situations, informed consent may be unattainable. Patients are under extreme duress and may be unable to fully comprehend all the aspects of a randomized clinical trial. In such cases, adaptive designs are a morally safe alternative to fully randomized designs. ACTs provide a means for clinicians to study a new treatment under life-threatening illness while not sacrificing the care of their patient.

When clinicians enrol their patients in a well designed ACT, they ensure that their patients obtain the finest care available at that time of enrolment and gain information about a new, experimental treatment. Both goals are accomplished without any compromise of ethics due to the principle of interchangeability [16]. Pullman and Wang state that a design satisfies the principle of interchangeability if any two patients are ethically interchangeable, whether they are treated in the trial or afterwards. Adaptive designs satisfy this principle because any patient receiving treatment while in the trial is

ethically interchangeable with any patient receiving treatment after the trial. The principle dictates that the care that a patient receives, whether they are treated during the trial or afterwards, will be the best possible care available, given current information [16]. The patient's treatment depends only on when they fell ill. If treatment is administered during the trial, then the patient will receive the treatment identified as the best possible one given current information. A patient treated after the trial will receive what was proven to be the superior treatment.

An adaptive clinical trial begins much the same as a randomized clinical trial. There are usually two treatments, an experimental and a control, although there can be more. Ideally, the clinician will hopefully be in a state of equipoise at the start of the trial. Equipoise is ideal, but it is often not the case in Phase III of a clinical trial. After Phase I, a small trial to establish dosage, and Phase II, a study to determine toxicity, there is usually prior knowledge of any beneficiary performance of the drug or treatment. However, it is usually the practice at the start of Phase III to equally allocate the treatments to the subjects. Once the first subject, or group of subjects has received treatment, then data is collected on the success or failure of this treatment. This data is then used to adjust the probabilities of assignment to future subjects. So, if the treatment was successful, the probability that the next subject will receive that treatment increases, while the assignment probability of the other treatment decreases accordingly. As the trial progresses, the process continues and allocation probabilities are adjusted accordingly. It is important to note that while the assignment proportions change during the trial, every subject is still randomly allocated to his treatment. At the end of the trial, the hope is that not only will a superior treatment be revealed, but also that a higher percentage of patients will have had received that treatment.

There are many different types of adaptive designs. Perhaps the most famous is Wei and Durham's randomized play the winner design (1978). An adaptation of this design was used in the ECMO trial conducted at the University of Michigan [1]. The design begins with 1:1 allocation of two treatments, an experimental and a control. Two marbles are placed in an urn, each of a different colour to symbolize the two treatments. For this reason, the randomized play the winner design is one of the Urn models. One marble is drawn from the urn. Its colour corresponds to the treatment that the first patient is to receive. For example, if we have red and white marbles, red for experimental treatment and white for control, and a red marble is drawn then the patient will receive the experimental treatment. This is how randomization is incorporated into the trial.

After the experimental treatment has been administered, it will be deemed a success or failure. If the treatment was successful, then another red marble will be added to the urn. This will increase the probability of the next patient receiving that treatment. If it was a failure, then a white marble will be added, which increases the assignment probability of the control, or conventional treatment. Another marble is drawn and whatever colour it is, that is the treatment that the next patient will receive. Depending on success or failure of that treatment on this second patient, another marble of the appropriate colour will be added. In this way, the trial continues. All the while, the assignment probabilities change with every additional patient. That is what makes this design adaptive.

What makes this design particularly attractive is that as the trial progresses, the patient will in time have higher probability of receiving the better treatment. This is due to the design of the experiment. If one treatment is superior to the other, then there will be a higher concentration of that treatment's marbles in the urn, and thus a higher

probability of receiving that treatment. If there is a stopping rule set before the start of the trial, and, during the trial, this rule is satisfied or there is a statistically significant difference in the performance of the two treatments, then the trial can be stopped ahead of schedule and a valid conclusion can still be reached. In desperate medical situations, the need to find the best treatment possible is strong. It may be that the experimental treatment was not found to be different from (or better than) the conventional treatment. Whether the valid conclusion is positive or negative, the sooner it is made, the sooner more patients will receive the best care possible. One disadvantage of adaptive clinical trials is that it may not be immediately obvious if one treatment is out-performing the other, and the trial may carry on longer than necessary. ACTs improve upon RCTs by effectively treating as many patients as possible with the superior treatment within the trial.

There is some danger, however, that the urn model can skew results quickly. If the first few patients receive the treatment A, and these treatments are deemed successful, then there will be a higher concentration of type A marbles in the urn in the beginning of the trial. This can give little chance for the other treatment to "prove itself." Also, the opposite can happen, resulting in a high quantity of type B marbles in the urn after the first few patients. In both cases, the results can become skewed to favour one treatment. An even more extensive problem may be that there is an abundance of patients treated with one treatment, resulting in few patients treated with the other. This was the case with the ECMO trials at the University of Michigan in 1985 [1]. At the end of the trial there were ten patients who received ECMO (eleven, if you include the patient treated after the trial), but only one patient who had received the control treatment. In addition to

making it difficult to compare treatments this imbalance of treatments can at times make it impossible to draw a statistically significant conclusion.

Even with the above drawbacks of the randomized play the winner design, it remains an eloquent method for clinicians to compare treatments in life-threatening situations. The majority of patients receive the superior treatment and due to the principle of interchangeability, the ethical dilemma between individual and collective ethics is diminished. It is obvious that although RPW designs are not widely used in the medical community as of yet, incorporating adaptive designs such as the randomized play the winner design would be of great benefit to the patient.

2.3.1 Play the Winner Designs

Although 50:50 randomized designs are the most commonly used designs in clinical trials, they may not always be the best choice for an experimental design when the patients to be randomized are severely ill. Certainly, 50:50 randomized designs are the gold standard [3] in medical trials because they reduce systematic bias, are easy to implement and are relatively straightforward to analyze. In some cases, however, they are not the best choice in medical trials, ethically speaking. Doctors cannot justify using 50-50 randomization when their patients are suffering from life-threatening illnesses. Often, adaptive designs are a more attractive option to patients because they allocate a greater proportion of patients to the superior treatment.

Even though adaptive designs out-perform 50:50 randomized designs in terms of increasing the number of patients who receive the better treatment almost always, they are rarely used in practice. Simon [23] believes that this is the case because most methods have important deficiencies that render them unsuitable for application. Even

though adaptive designs may be more complex in implementation and analysis, the extra effort is definitely worthwhile when more patients receive the better treatment under study, especially in cases of life threatening illness. Important adaptive designs in the literature include Zelen's play the winner design [34] and Wei and Durham's randomized play the winner design [30].

Adaptive designs do not always have to include randomization. This is the case with Zelen's play the winner rule [34], or PW. This design is appropriate when there are two treatments under study and patients enter the trial sequentially. Furthermore, it is necessary to assume that the outcome of a trial is only dependent on the treatment administered and is dichotomous. When the first patient enters the trial, he is randomly assigned one of the two treatments with probability of one half. Once the response of that treatment is known, a marble is placed in an urn. If the treatment is a success, then a marble of type A corresponding to treatment A is placed in the urn, if the treatment is a failure, then a marble of type B corresponding to treatment B is placed in the urn. When the next patient enters the trial, a marble is drawn from the urn without replacement. Since there is only one marble to choose from, the patient receives that treatment. As long as all future trials are successful, all successive patients will receive the same treatment. As soon as a failure is observed, a marble of the other treatment is placed in the urn and the next patient to enter the trial is allocated to the other treatment. Then, all future patients will receive the other treatment until a failure is observed. If a patient enters the trial in between responses, and there are no marbles in the urn, then the patient is randomized.

Zelen is the first to admit that the design is not optimal, but he states it is nearly optimal, requires very few assumptions about prior distribution, and most importantly,

tends to assign more patients to the better treatment [34]. The main flaw with this design is that it is deterministic. Due to the design of the study, the researcher will know with almost complete certainty what the next treatment will be given the previous treatment. If a success was observed on the previous trial, the next patient will receive the same treatment, and if a failure was observed, then the next patient will receive the other treatment available. The deterministic property of the PW design can cause selection bias in the researcher.

A modification of the PW, aptly named the modified play the winner rule, or MPW, maximizes the chance of selection bias. MPW can be used in place of PW when the response to treatment is immediate, and the second patient enters the trial right after the first patient has completed treatment. For example, if the first patient's treatment was successful using treatment A(B), then the second patient receives treatment A(B). If it was a failure, then the second patient receives the opposite treatment. The next patient is never treated until the outcome of the previous patient is known. Due to the immediate response of this design, the researcher knows with probability one what treatment the next patient will be assigned once the first patient's response is ascertained. Moreover, on account of the lack of randomization in both the PW and MPW designs, there is also the possibility of systematic bias.

What makes the PW design particularly unattractive is that in practice the design is not quick to come to a conclusion and when there is an extreme delay between responses, patients are randomly allocated to treatments. Thus, at times the PW is not adaptive and does no better than 1:1 allocation. One way to improve these designs, as well as to reduce bias, is to include randomization in the design, as with Wei and Durham's randomized play the winner design.

Wei and Durham designed the randomized play the winner design in 1978 in the hope of improving Zelen's play the winner design. They wanted to build on Zelen's notion of maximizing the proportion of patients who receive better treatment, and on his ethical concern for reducing the length of the trial. The best way to solve the problems of a deterministic design with the possibility of selection bias and systematic bias was to include randomization into the design. This is the main difference between the PW, MPW and RPW designs. Adding randomization into the design is a simple change, but improves many characteristics of the design. Immediately, systematic bias is reduced. Furthermore, the allocation of patients is no longer deterministic, patients are randomly allocated to treatments, and thus selection bias is also reduced.

The assumptions necessary for the design are similar to those of the PW: the response of the patient must be dichotomous, either a success or a failure, and there are two treatments under study where the probability of success of treatment í is p_i, í=A, B, and where it is assumed $p_A \geq p_B$, without loss of generality. The trial begins with μ marbles of each type placed into an urn, each with one of two markings to represent the two treatments. When the first patient enters the trial, a marble is drawn from the urn and replaced. Recall that in the PW design, the marbles were drawn without replacement. If the marble drawn was of type A (B), then the patient will receive treatment A (B). When the response of the patient is available, additional marbles are placed into the urn to reflect that response and to alter the assignment probabilities of the two treatments. If the treatment was a success, then α marbles of the same type are placed into the urn. If the treatment was a failure, then β marbles of the opposite type are placed into the urn. Thus, the randomized play the winner rule is referred to as RPW(μ, α, β). Often, however, α is equal to β, so the design is also denoted as RPW(μ, α) When the next patient is available,

another marble is drawn from the urn and he receives the corresponding treatment. Note that this design allows for delayed responses to treatments. The main improvement of the RPW over the PW is that it decreases the number of patients assigned to the inferior treatment, and reduces selection bias.

It is common practice to only add one marble after obtaining the response of each patient. Often, the experiment commences with more than one marble of each type in the urn. If the initial urn size is increased, the conventional choice is to begin with five marbles of each type, or RPW(5, 1, 1). An increase in the initial marble count, however, shifts the focus of the trial away from adaptation of treatment allocation probabilities and towards randomization of patients. Due to the increased weight on randomization, the trial progression may speed up, and a significant conclusion may be reached sooner, but fewer patients may actually receive the superior treatment. In fact, one can expect the allocation proportions to be less extreme when the initial urn composition is increased, as the urn will not favour the superior treatment as highly [17].

But, what if more than one marble is drawn with replacement to determine the treatment to be administered, and then the same number of marbles is then added to the urn after the result of a treatment is known? What improvement, if any, will come from drawing, then adding, more than one marble? For example, what if a RPW(3, 3, 3) or RPW(5, 5, 5) design was conducted instead of a RPW(1, 1, 1) or RPW(5, 1, 1) design? Note that out of necessity, the trial must begin with at least the same number of marbles that will be drawn initially. Thus, the urn contains $2n+1$ marbles of each type at the beginning of the trial (where n is a natural number.) After the first patient enters the trial, $2n+1$ marbles are drawn, and then replaced. For convenience, an odd number of marbles should be drawn. If an even number of marbles is drawn, then there is the possibility of

having the same number of marbles of each type. This may result in having to randomize the patient to one of the two treatments, and could also lead to poor results in terms of maximizing the number of patients to the superior treatment. Since the number of marbles drawn is odd, one type of marble will always outnumber the other type. It follows that majority rules and the patient receives the treatment corresponding to the winning marble type. Once the response of the first patient is known, 2n+1 marbles are added to the urn. As before, if the treatment was a success, 2n+1 marbles of the same type are added to the urn, and if the treatment was a failure, 2n+1 marbles of the opposite are added to the urn. When the second patient enters the trial, 2n+1 marbles are again drawn and replaced. When the response of that patient is known, 2n+1 marbles of the appropriate type are added, and so on. This design works for any nonnegative integer value of n. Note that when n equals zero, the design reverts back to the classic RPW(1, 1, 1).

A RPW(2n+1, 2n+1, 2n+1) design would hopefully increase the immediate payoff to current patients by maximizing the number of patients who receive the superior treatment without sacrifice to the power the test. This improvement results from drawing, and then adding more marbles. The more marbles of any one type in the urn, the better chance the treatment corresponding to that marble type will be administered. Thus, multiplying the number drawn from one to 2n+1 skews the test to the superior treatment faster. Then, the superior treatment becomes apparent faster, and more patients receive that treatment earlier. These modifications do improve on the classic RPW(1, 1, 1) design in terms of immediate payoff to the patient.

Day [6] points out that adaptive clinical trials are better than sequential trials, which in turn are better than randomized clinical trials in terms of maximizing the

proportion of patients to the superior treatment. Just the opposite is true if the goal of the researcher is to maximize power. In situations where patients' health is critical, the ethical choice is to choose a higher proportion of successes over a higher significance level. The RPW design is a well-known design for its higher proportion of successes without sacrifice to power. The next chapter will highlight in detail which designs excel in which situations, and if modifications to the RPW design can improve results.

2.3.2 Neyman Allocation

The optimal allocation for minimizing the expected proportion of treatment failures can be found by fixing the variance of the test statistic and then calculating the optimal allocation ratio based on the ratio of the sample sizes of the two treatments, n_A/n_B [19]. When comparing two proportions, each assumed to be from a normal distribution, the variance of the difference of treatments A and B, $y_A - y_B$, is the squared denominator of the Z-test [19], given by:

$$\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}, \text{ where}$$

$$\sigma_i^2 = \frac{p_i q_i}{n_i}$$

and σ_i^2 is the variance of treatment i, i = A, B. Note that p_i is the probability of success of treatment i and q_i is the probability of failure of treatment i, equal to 1-p_i. We can set this equation equal to a constant, and with some calculations, obtain the ratio,

$$R = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{v(\theta)}{u(\theta)}},$$

where $v(\theta)$ and $u(\theta)$ are functions of the treatment effect, $\theta = \mu_A - \mu_B$. When $v(\theta)$ and $u(\theta)$ are equal, then we are left with the ratio of standard deviations of the two treatments, σ_A/σ_B, which is Neyman allocation. Neyman allocation minimizes the total number of patients in the treatment when the variance of the difference in sample proportions is fixed [18].

To use Neyman allocation in a trial, the variances of the probabilities of success of the two treatments must first be estimated. They can be estimated from historical data, previous trials or previous phases of the current trial. Then, patients are randomly allocated to treatments. First, the initial value of Q, the allocation rule, is calculated as below,

$$Q = \frac{\sigma_A}{\sigma_A + \sigma_B},$$

then patients are assigned treatments, where Q is the probability of allocating treatment A. This allocation rule maximizes the power of the test when the total sample size is fixed. When the first patient enters the trial, a table of random numbers, or a similar tool, is consulted. If the random number drawn is less than or equal to Q, then the patient receives treatment A, the experimental treatment. If the random number is greater than Q, then he receives treatment B, the conventional treatment. When the response of the patient is known, the probabilities of success are adjusted accordingly, which causes their standard deviations to change as well. As a result, Q is updated after every response. Neyman allocation is found to improve upon of the proportion of patients who receive the superior treatment when compared with 50-50 randomization, but only when $p_B < q_A$. If $p_B > q_A$, then Neyman allocation actually puts a higher proportion of patients on the inferior treatment [18], which is not ethical.

Adaptive designs, including Neyman allocation and play the winner designs, are underutilized alternatives to randomized designs in the medical field. Adaptive clinical trials are the only ethical choice for clinicians when treating severely ill patients as the patient's chances of receiving the better treatment are improved. The goal of this paper is not to argue the merit of the RPW design, but to improve upon it. Though adjustments to the experimental design and through simulation studies, it will be shown that it is possible to increase the proportion of patients who receive the superior treatment without significant loss of power.

Chapter 3

Simulation Design and Analysis

A simulation is an experiment carried out on a computer. Simulations are an attractive alternative to actual experiments. They can be used to model and analyse statistical systems. Observations generated from simulated experiments can be analysed in the same manner as those produced by an actual trial without the construction or experimentation of a real system. Simulations are particularly ideal to model medical experiments. Compared to the years it can take to accumulate all the subjects needed, as well as the expense involved with monitoring all the patients, simulations can not only save time and money but can also obtain results similar to those obtained by an authentic clinical trial. Furthermore, simulations can observe trends and strategies. They can test theories as well as perform experiments that may not otherwise have been possible.

The most significant advantage of using a simulated experiment over a real one is that there are no ethical issues to be considered. Since none of the subjects are actual patients, just random numbers, there is no conflict between individual and collective ethics. Possibly one draw back is that if the simulation results are misleading it can lead to the wrong conclusion. This danger, nonetheless, is much less worrisome than if the wrong conclusion is reached in an actual clinical trial. In this thesis simulations will be used to model adaptive clinical trials including the classical randomized play the winner design and variations on it.

3.1 Random Number Generation

Before a clinical trial can be simulated, a source of randomness is required for the simulation. Often this source comes from a random number generator. A random number generator can obtain a dataset that simulates the actual experiment. Simulations do not duplicate or replace actual clinical trials. For ethical reasons, as in the case of medical trials, it may be more appropriate to simulate experiments and obtain approximate results that one would expect on average rather than obtaining real data. The dataset generated usually has a uniform distribution, where random numbers are contained on the open subset (0, 1). Then, the sequence is transformed into the required distribution to obtain the desired dataset.

Not all random number generators are considered "good." A "bad" random number generator produces variables with non-random properties [10]. A good random number generator has the following aspects: it produces random variables that are independent and uniformly distributed, it is fast, efficient and easily implemented, and it expresses little deviation from the desired statistical properties.

Most generators operate using a recursive relationship. That is, the next number in the sequence is a function of one or more previous numbers in that sequence. The recursive relationship is of the form:

$$x_n = f(x_{n-1}, x_{n-2}, \ldots, x_{n-k}),$$

where $k<n$. To start the sequence, a seed is required. Just as there are good and bad generators, there are also good and bad seeds. In theory, any choice for a seed other than zero should not affect the results of the simulation. Sometimes, however, some seeds

work better for certain random number generators than others. It is best to select an odd numbered seed. Even values for seeds can lead to poor results. It is also beneficial to not use a random seed, such as the time of day, because it can be difficult to replicate results. If the random number generator requires a certain value or type of seed for independent, uniformly distributed results, it may not be such a wise choice, as it is too easy to choose a poor seed, and thus a non-random sequence. In order to reproduce the same dataset on each run, the same seed should be used in successive replications. Once the seed has been chosen, the sequence of random variables can be predicted with certainty. Thus, the numbers produced are called but pseudo-random numbers. Pseudo-random numbers are not truly random because the choice of the seed determines with absolute certainty the sequence of the random numbers [10]. Pseudo-random numbers are more desirable than fully random numbers because if the experiment needs to be repeated, the simulation will produce exactly the same outcome if the same seed is selected. In all simulations mentioned in this thesis the seed used was 12773.

There are numerous types of random number generators, some more widely used than others. Common ones include Linear-Congruential generators, Tausworthe generators, Fibonacci generators and Combined generators, which are a combination of two or more other types. In this paper, Linear-Congruential generators, (or LCG's), will be used for all simulations. LCG's rely on a recursive relationship to produce the sequence of random variables. The basic form of the LCG is:

$$x_n = ax_{n-1} \bmod m,$$

where a is the multiplier and m is the modulus, or remainder. A common variation on this form that is popular in generators today is:

$$x_n = (ax_{n-1} + b) \bmod m .$$

Generators of this form are also called Mixed Linear-Congruential generators because the equation includes both multiplication and addition. As before, there are good and bad choices for *m*, *a* and *b*. The multiplier, *a* as well as *b*, should be non-zero. The modulus *m* should be large. Both *m* and *b* should have no common factors between them, thus it may be best to choose prime numbers for both. A LCG that has been shown to perform very well is:

$$x_n = 7^5 x_{n-1} \bmod (2^{31} - 1).$$

This recursive relation generates a sequence of independent random variables distributed uniformly with lower limit zero and upper limit one, denoted *U(0, 1)*. It is the random number generator used in all simulations in this thesis.

After a sequence that behaves statistically like data from random numbers has been generated, it is necessary to check that the numbers are indeed random, independent and uniformly distributed. One simple, but important check is to insure that the sequence is uniformly distributed with values between zero and one. To do this, calculate the mean and variance of the random numbers. The mean should equal one half, which corresponds to the mean of a uniform distribution,

$$\bar{x} = \frac{a+b}{2} .$$

The variance of the uniform distribution,

$$\sigma^2 = \frac{(b-a)^2}{12},$$

should equal one twelfth. Other important tests include the Chi-square Goodness of Fit test, the Kolmogorov-Smirnov Test, and the Serial Correlation Test.

The Chi-square Goodness of Fit test, or the chi-square test, tests whether a sequence of random numbers satisfies the required distribution [10]. Furthermore, the test can verify if the sequence is independently and identically distributed *U(0, 1)*. It follows that the chi-square test can also verify if a sequence of numbers is random, and if the random number generator is working properly. To begin the test, the closed interval [0, 1] is divided into k equal parts, or cells. Then, for a sequence of size n, the expected frequency in each cell is n/k. The random numbers generated are sorted into their appropriate cells, and a count of the number of variables in each cell is taken. Then, this actual count is compared with the expected frequency for each cell. This comparison is made through the calculation of the chi-square statistic,

$$\chi^2 = \sum_{i=1}^{k} \left(\frac{o_i - e_i}{e_i} \right)^2 .$$

The above statistic calculates the sum of the adjusted, squared differences between the observed (o_i) and expected (e_i) counts for each cell. It has a chi-square distribution with *k-1* degrees of freedom. If the random number generator produces a dataset that perfectly fits the stipulated distribution, then the statistic would be zero. Due to the randomness of the generator, the difference will be nonzero. To test the null hypothesis,

H_o: *The random numbers are distributed U(0, 1)*, versus the alternative,

Ha: The random numbers are not distributed U(0, 1),

the chi-square statistic is compared with the critical value for the chi-square distribution with *k-1* degrees of freedom at significance level α. If the chi-square statistic is less than

the critical value, then the null hypothesis will not be rejected and the generated dataset can be considered uniformly random.

Before the random numbers generated using the aforementioned LCG were used in any simulation for this paper, they were tested to ensure they were independent and identically distributed *U(0, 1)*. The chi-square test was used to test the above null hypothesis. The value of the test statistic was 49.335, which was less than 66.34, the chi-square critical value at significance level 0.05 with 50-1 degrees of freedom. Thus, the null hypothesis was not rejected and the pseudo-random numbers generated and used in this thesis indeed behave statistically like data from uniformly distributed and random.

The chi-square test is best used for large samples and discrete distributions. It can be used as an approximate test for smaller samples or continuous distributions, but more specific tests exist. One such test is the Kolmogorov-Smirnov test, or K-S test, so named for the statisticians who developed it [10]. Like the chi-square test, the K-S test determines if a dataset follows a specific distribution. The test is based on the observation that the observed Cumulative Distribution Function (CDF) should be close numerically to the expected CDF. Two statistics are calculated to carry out the test, K^+ and K^-, which measure the maximum or minimum deviation of the empirical CDF above or below the expected CDF, respectively. The form of the expected CDF for the Uniform distribution is as follows:

$$F(x;a,b)=\frac{x-a}{b-a},\ x\in[a,b]$$

When $a = 0$ and $b = 1$, $F(x) = x$, and if x is greater than $j-1$ of the n random numbers, then the form of the observed CDF is $F(x) = j/n$. To carry out the test, the data must first be

sorted in increasing order. Then, to compare the observed and expected CDF 's, K^+ and K^- are calculated as below:

$$K^+ = \sqrt{n}\max_j\left(\frac{j}{n} - x_j\right),$$
$$K^- = \sqrt{n}\max_j\left(x_j - \frac{j-1}{n}\right).$$

If both K^+ and K^- are less than the Kolmogorov-Smirnov critical values for significance level α, then the null hypothesis cannot be rejected against the alternative, as below:

H_o: The random numbers are distributed U(0, 1),

Ha: The random numbers are not distributed U(0, 1).

One advantage the K-S test has over the chi-square test is that it is not necessary to group the data into cells. Grouping data into cells can be problematic. Often cells need to be combined if there are too few data points in any one cell. In addition, cell size can effect the conclusion of the test. The K-S test handles each observation individually, eliminating the need to group data, and making better use of it.

Besides testing for randomness in a generated sequence, it is also necessary to test for correlation between data. To apply the serial-correlation test, we calculate the sample autocovariance, or the sample covariance between numbers that are k values apart. The difference in location between the numbers is called the lag. The sample autocovariance at lag k is denoted R_k and is of the form:

$$R_k = \frac{1}{n-k}\sum_{i=1}^{n-k}\left(U_i - \frac{1}{2}\right)\left(U_{i+k} - \frac{1}{2}\right).$$

As n gets large, R_k becomes normally distributed,

$$N(0, 1/[144(n-k)]).$$

Thus, a *100(1–α)%* confidence interval for the autocovariance at lag k can be computed.

$$\text{If the interval, } R_k \pm z_{1-\alpha/2} \Big/ \left(12\sqrt{n-k}\right)$$

includes zero for all lags one through k, then the null hypothesis cannot be rejected against the alternative, as below:

H_o : *there are no correlations in the sequence*,

H_a: *there is a significant correlation.*

Note that if the null hypothesis is not rejected, evidence suggests the sequence is not independent. Testing independence is necessary in a generated sequence to ensure that the generator is working properly.

The generated pseudo-random numbers used in all simulations in this paper were tested using the serial correlation test to certify that there are no correlations for lags one through k in the sequence. The above null hypothesis was tested for lags one through ten, and all confidence intervals contained zero with the exception of lag five. Although the lag five confidence interval did not contain zero, the upper limit was extremely close to zero with a value of negative 0.005836. Therefore, overall, the null hypothesis was not rejected at significance level 0.05 and the pseudo-random numbers used in this thesis behaved as though they were not correlated.

There are many tests available beyond the ones mentioned that test the quality of the pseudo-random numbers, including Multi-Dimensional tests like the Serial Test and the Spectral Test [10]. What is important is not that the numbers generated pass every test available, but that the pseudo-random numbers are tested and deemed reliable.

Once tests have been conducted to ensure that the sequence generated is in fact random, then the sequence can be transformed to the desired distribution. This transformation can be obtained by various methods including inverse transformation,

rejection, convolution, composition and characterization. One of the simplest methods is inverse transformation which uses the property that a cumulative density function (CDF), *F(x)*, of a random variable, *u*, of any distribution is uniformly distributed between zero and one. This means that given a sequence of random variables *{u}* distributed *U(0, 1)*, the inverse of the CDF, $F^{-1}(u)$ can be used to transform the sequence to a sequence *{x}* of the desired distribution as long as the CDF is known. For example, suppose we wish to transform a sequence of pseudo-random numbers from a uniform distribution to an exponential distribution. Since the CDF of an exponential distribution is:

$$F(x) = 1 - e^{-\lambda x} = u, \; x \geq 0,$$

then the inverse CDF is:

$$x = -\frac{1}{\lambda}\ln(1-u), \; u \in (0,1]$$

Thus, if the pseudo-random numbers are substituted into the above equation at *u*, a sequence, *{x}*, of exponential random numbers can be obtained without intense calculations.

Simulation is a powerful tool that can help to answer questions unanswerable if attempted by normal means, whether it is due to expense or ethical issues. Simulation studies, although they cannot duplicate actual clinical trials, can be used to obtain ideas before a trial or speculate on theoretical results. After the assumptions necessary for good simulations have been validated, simulation studies can be carried out and results analysed. In an upcoming section, simulated clinical trials of various experimental designs will be compared.

3.2 An Overview of Simulation Procedures

All the simulation studies carried out begin with the generation of pseudo-random numbers. The performance of the random number generator is examined with various statistical methods before it is applied to all simulations. Once a pseudo-random number has been generated, it is then compared against another number or variable called a decision variable. For example, in most of the SAS programs used in this thesis, the decision variable was the probability of success of the superior treatment. If the pseudo-random number was less than the current value of the probability of success of the superior treatment, then a decision would be made to allocate another experimental unit to one treatment. Otherwise, the experimental unit would be assigned to the other treatment.

Although the probabilities of success are declared at the start of the simulation to indicate the superior and inferior treatments, they are not fixed. As the simulation progresses, the probabilities are updated based on the number of successes and failures of the corresponding treatments. Likewise, as the probabilities of success are updated, so too are other variables used in the calculations of the statistics analyzed. Variables such as the count of the total number of successes, overall and on each treatment, as well as the count of the total number of patients on each treatment were all updated after the generation of a pseudo-random number and the decision process.

Once one hundred pseudo-random numbers had been generated, the final value of the simulation variables were used in the calculation of statistics such as the power of the test, the expected proportion on the superior treatment, the odds ratio and the proportion of successes on each treatment and in total. These values were saved by the computer program. Then, all the variables would return to a zero count and the simulated

experiment would be repeated for another one hundred experimental units. The simulation was repeated at least two thousand times, and after each time the values of the statistics were saved in the memory of the program. After completion of the repetitions, an average was taken of the saved values of the statistics to get a more accurate estimate as well as a value for error. Those final values were used in the simulation analysis to compare various designs and study the monotonic properties of the randomized play the winner design.

3.3 Comparison of Various Designs

When carrying out an experiment, the design is just as important as the results, if not more important. The design of an experiment dictates what type of analysis should be done. Moreover, without the proper design, an experimenter may not be able to carry out the required analysis to obtain the desired results. In clinical trials, the experimental design must accomplish two goals: first, to carry out the required analysis, and, second, to carry out the trial in an ethically responsible way. This implies that the trial must take into consideration that the experimental units are human beings, and any randomization of treatments must be done in a careful manner as to not prolong the experiment, nor to cause any undue harm to patients.

Obviously, not all experimental designs are well suited for clinical trials. Often, clinical trials can be improved upon by incorporating prior knowledge or accumulating information when assigning treatments to patients, in other words, by making use of adaptive designs. But, there are numerous adaptive designs to choose from, and not all are appropriate for clinical trials. Different situations require different optimal adaptive designs. Adaptive trials can be designed to maximize the percentage of patients assigned

to the superior treatment, minimize the expected sample size or minimize the expected cost. In this paper, an experimental design is considered optimal when it maximizes the expected proportion of patients assigned to the superior treatment, without significant loss of power. The performance of various experimental designs will be compared. It will be shown that adaptive designs improve upon a 50:50 randomized design in this context. In particular, the randomized play the winner design, or RPW, is a powerful design that is useful when the response variable is acutely life threatening. Also, it will be shown that modifying the parameters in the RPW design improves the allocation of patients to the superior treatment while still remaining a powerful design.

Adaptive designs are considered to be a compromise between two goals, the goal of gathering information for future patients and the goal of immediate payoff to current patients. Different designs give different weight to the opposing goals. When the patient is extremely ill, then main focus should be on improving their health. To focus on gathering information would be unethical. To improve a patient's well being while they are in a trial can be accomplished by improving their chances of receiving a superior treatment. The program SAS (version 8 for Windows) was used for simulation studies used to compare and contrast experimental designs. All simulations are based on 5000 replications, each with a sample size of one hundred unless otherwise indicated. The goal of the simulation studies in this paper is to improve the patient's chances of being allocated to the superior treatment. It will be shown that this goal is accomplished using the RPW design where the number of balls drawn and then added is greater than one.

3.3.1 Expected Proportion on the Superior Treatment

The expected proportion of patients allocated to the superior treatment has been mentioned frequently in this paper. It is a very significant statistic, because it explains numerically how an adaptive design like the randomized play the winner rule, improves upon a completely randomized design. Essentially, the higher the expected proportion, the greater the number of patients receiving the superior treatment and in cases of life-threatening illness, the higher the number of possible lives saved. So, if the expected proportion is higher for an adaptive design as opposed to a randomized one, it can be called an ethically superior design because the patient's health is put before the goal of information gathering. The expected proportion for treatment A is calculated by

$$\frac{n_A}{n_A + n_B},$$

where n_i is the number of patients who received treatment i.

Table 1 gives the values of the expected proportion of patients allocated to treatment A for the various designs previously mentioned. Note that due to the nature of its design; 50-50 randomization always allocates one half of the subjects to one treatment and the other half to the other treatment. In all cases, the RPW designs have a much higher expected proportion as compared to the Neyman allocation and 50-50 randomization cases. For example, when $p_A = 0.8$ and $p_B = 0.3$, the expected proportion is 0.499 (0.049) for 50-50 randomization, 0.464 (0.056) for Neyman allocation, 0.752 (0.073) for RPW(1, 1, 1), 0.840 (0.076) for RPW(3, 3, 3) and 0.888 (0.074) for RPW(5, 5, 5). Another trend visible in Table 1 is that the increase in the expected proportion is greater between RPW(1, ,1, 1) and RPW(3, 3, 3) than between RPW(3, 3, 3) and RPW(5,

5, 5). It may be that as parameters increase, the increase in expected proportion becomes less drastic. This will explored further in the monotonicity section.

Furthermore, as the difference between p_A and p_B increases, we see an increase in the expected proportion because there is a higher chance that the patient will receive treatment A. Figure 1, which is a graph of the expected proportion on treatment A for various designs with p_A varying and p_B held constant at 0.1, shows that the RPW designs outperform 50-50 randomization and Neyman allocation in terms of allocating more patients to the superior treatment. In fact, the trend seems to be that the higher the number of marbles added, the higher the expected proportion of patients assigned to the superior treatment. As the expected proportion on the superior treatment increases, so does the standard deviation, but overall the number of patients receiving the superior treatment rises. This result is quite noteworthy. A simple modification to the randomized play the winner design that is easy to implement has produced a more ethical choice for clinicians who face the task of conducting clinical trials on dreadfully ill patients.

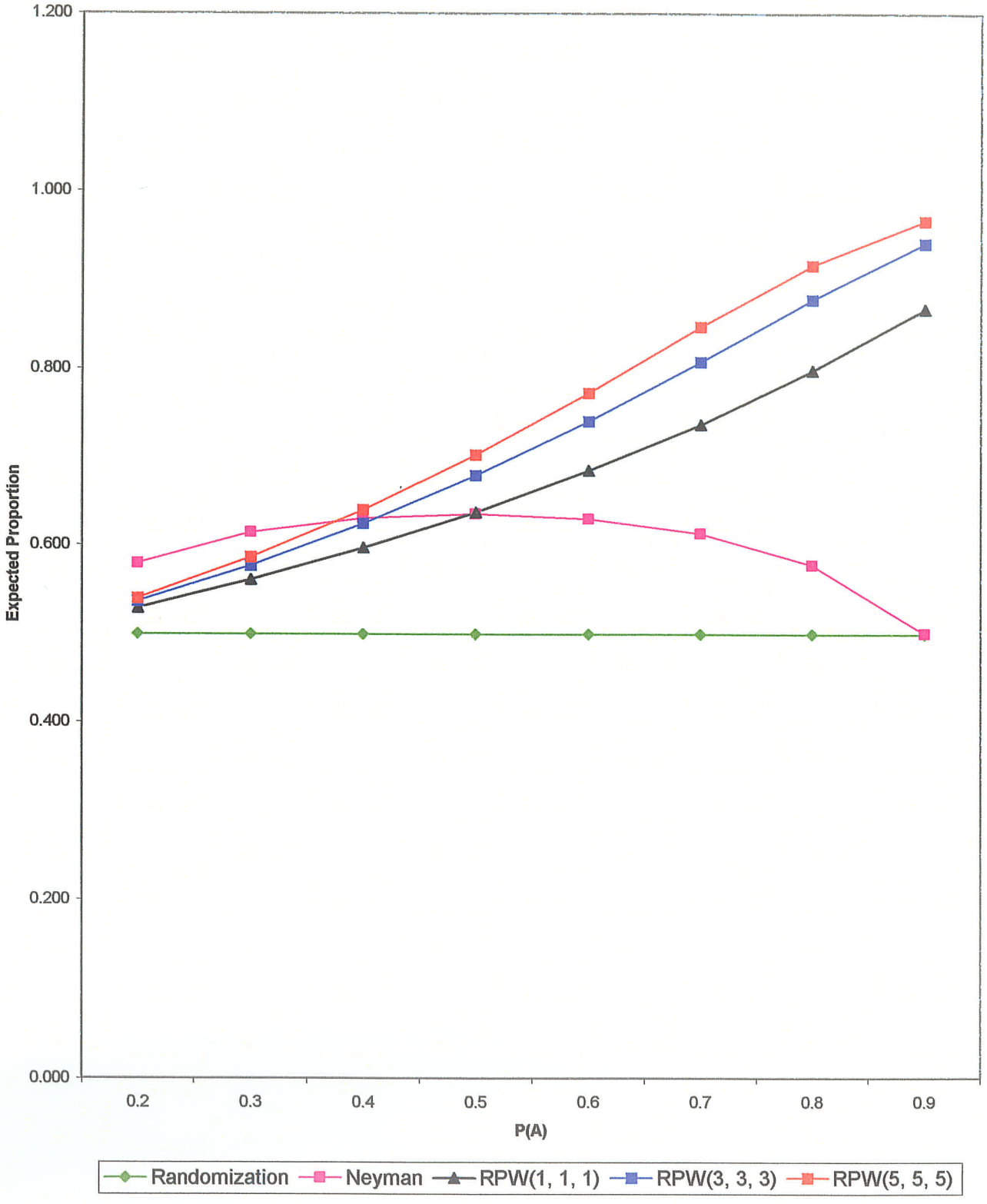
Type of Design	P(A) / P(B)	0.2	0.3	0.4	0.5
Randomization		0.499 (0.049)	0.499 (0.049)	0.499 (0.049)	0.499 (0.049)
Neyman		0.579 (0.067)	0.614 (0.350)	0.630 (0.062)	0.635 (0.062)
RPW(1, 1, 1)	**0.1**	0.529 (0.040)	0.561 (0.044)	0.597 (0.048)	0.637 (0.051)
RPW(3, 3, 3)		0.536 (0.041)	0.577 (0.046)	0.624 (0.051)	0.679 (0.057)
RPW(5, 5, 5)		0.540 (0.042)	0.586 (0.048)	0.639 (0.053)	0.702 (0.059)
Randomization			0.499 (0.049)	0.499 (0.049)	0.499 (0.049)
Neyman			0.537 (0.056)	0.555 (0.055)	0.560 (0.054)
RPW(1, 1, 1)	**0.2**		0.533 (0.049)	0.569 (0.053)	0.611 (0.570)
RPW(3, 3, 3)			0.541 (0.053)	0.589 (0.059)	0.646 (0.065)
RPW(5, 5, 5)			0.547 (0.056)	0.602 (0.062)	0.668 (0.069)
Randomization				0.499 (0.049)	0.499 (0.049)
Neyman				0.519 (0.052)	0.524 (0.051)
RPW(1, 1, 1)	**0.3**			0.538 (0.058)	0.580 (0.063)
RPW(3, 3, 3)				0.549 (0.068)	0.607 (0.075)
RPW(5, 5, 5)				0.557 (0.072)	0.627 (0.082)
Randomization					0.499 (0.049)
Neyman					0.507 (0.049)
RPW(1, 1, 1)	**0.4**				0.544 (0.070)
RPW(3, 3, 3)					0.559 (0.089)
RPW(5, 5, 5)					0.572 (0.099)

Table 1: Expected Proportion on the Superior Treatment (Standard deviations are given in parentheses)

Note to this and all subsequent tables: P(A) is the probability of success of treatment A and P(B) is the probability of success of treatment B unless otherwise indicated.

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		0.499	(0.049)	0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman		0.630	(0.062)	0.613	(0.064)	0.578	(0.067)	0.500	(0.077)
RPW(1, 1, 1)	**0.1**	0.684	(0.054)	0.737	(0.055)	0.798	(0.056)	0.867	(0.050)
RPW(3, 3, 3)		0.740	(0.059)	0.807	(0.059)	0.878	(0.050)	0.941	(0.035)
RPW(5, 5, 5)		0.772	(0.061)	0.847	(0.058)	0.916	(0.045)	0.967	(0.026)
Randomization		0.499	(0.049)	0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman		0.555	(0.055)	0.537	(0.056)	0.500	(0.059)	0.422	(0.067)
RPW(1, 1, 1)	**0.2**	0.659	(0.060)	0.713	(0.063)	0.777	(0.063)	0.851	(0.059)
RPW(3, 3, 3)		0.712	(0.069)	0.784	(0.070)	0.861	(0.061)	0.932	(0.044)
RPW(5, 5, 5)		0.744	(0.073)	0.827	(0.070)	0.904	(0.057)	0.961	(0.034)
Randomization		0.499	(0.049)	0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman		0.519	(0.052)	0.501	(0.053)	0.464	(0.056)	0.388	(0.063)
RPW(1, 1, 1)	**0.3**	0.630	(0.068)	0.685	(0.072)	0.752	(0.073)	0.829	(0.071)
RPW(3, 3, 3)		0.677	(0.082)	0.755	(0.084)	0.840	(0.076)	0.920	(0.057)
RPW(5, 5, 5)		0.708	(0.088)	0.800	(0.088)	0.888	(0.074)	0.953	(0.048)
Randomization		0.499	(0.049)	0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman		0.501	(0.050)	0.483	(0.051)	0.446	(0.054)	0.371	(0.062)
RPW(1, 1, 1)	**0.4**	0.595	(0.076)	0.652	(0.082)	0.722	(0.086)	0.802	(0.086)
RPW(3, 3, 3)		0.632	(0.099)	0.717	(0.102)	0.810	(0.098)	0.901	(0.079)
RPW(5, 5, 5)		0.659	(0.110)	0.761	(0.113)	0.861	(0.100)	0.938	(0.071)
Randomization		0.499	(0.049)	0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman		0.495	(0.050)	0.477	(0.051)	0.441	(0.054)	0.366	(0.061)
RPW(1, 1, 1)	**0.5**	0.553	(0.085)	0.612	(0.093)	0.683	(0.100)	0.768	(0.103)
RPW(3, 3, 3)		0.577	(0.117)	0.667	(0.127)	0.769	(0.128)	0.874	(0.110)
RPW(5, 5, 5)		0.593	(0.138)	0.705	(0.149)	0.820	(0.141)	0.914	(0.109)
Randomization				0.499	(0.049)	0.499	(0.049)	0.499	(0.049)
Neyman				0.483	(0.051)	0.446	(0.054)	0.371	(0.061)
RPW(1, 1, 1)	**0.6**			0.562	(0.107)	0.635	(0.117)	0.723	(0.126)
RPW(3, 3, 3)				0.598	(0.157)	0.711	(0.166)	0.829	(0.155)
RPW(5, 5, 5)				0.618	(0.198)	0.749	(0.204)	0.862	(0.180)
Randomization						0.499	(0.049)	0.499	(0.049)
Neyman						0.463	(0.056)	0.388	(0.063)
RPW(1, 1, 1)	**0.7**					0.577	(0.136)	0.668	(0.151)
RPW(3, 3, 3)						0.624	(0.217)	0.758	(0.220)
RPW(5, 5, 5)						0.642	(0.279)	0.776	(0.274)
Randomization								0.500	(0.049)
Neyman								0.422	(0.066)
RPW(1, 1, 1)	**0.8**							0.595	(0.181)
RPW(3, 3, 3)								0.647	(0.299)
RPW(5, 5, 5)								0.654	(0.369)

Figure 1: Expected Proportion to the Superior Treatment



3.3.2 Power of the Test

The concept of power is talked about frequently in statistics, and it is definitely an important tool for measuring the validity of tests of hypothesis. In this paper, the power of a test can be defined as the probability of correctly rejecting the null hypothesis. The test of hypothesis in question is:

$$H_O: p_A = p_B$$

$$H_A: p_A \neq p_B,$$

where p_A is the probability of success of the experimental treatment and p_B is the probability of success of the conventional treatment.

Table 2 gives values of the power of the above test of hypothesis for 50:50 allocation, Neyman allocation, RPW(1, 1, 1), RPW(3, 3, 3) and RPW(5, 5, 5). The probabilities found are relevant for a two-sided alternative hypothesis and are significant at the five per cent level. The standard normal critical value used for alpha equal to 0.025 was 1.96. Note that the power calculated is based on random and independent samples so all values are approximations. The probabilities were estimated by comparing two binomial proportions. The form of the probability calculated is given below:

$$Power = \Phi\left[\frac{|p_A - p_B|}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}} - z_{1-\alpha/2}\frac{\sqrt{\overline{pq}\left(1/n_A + 1/n_B\right)}}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}}\right],$$

where Φ is the standard normal density function, p_i is the probability of success of treatment i, q_i is equal to 1 - p_i, n_i is the sample size of treatment i,

$$\overline{p} = \frac{n_A p_A + n_B p_B}{n_A + n_B} \text{ and, } \overline{q} = 1 - \overline{p}\,.$$

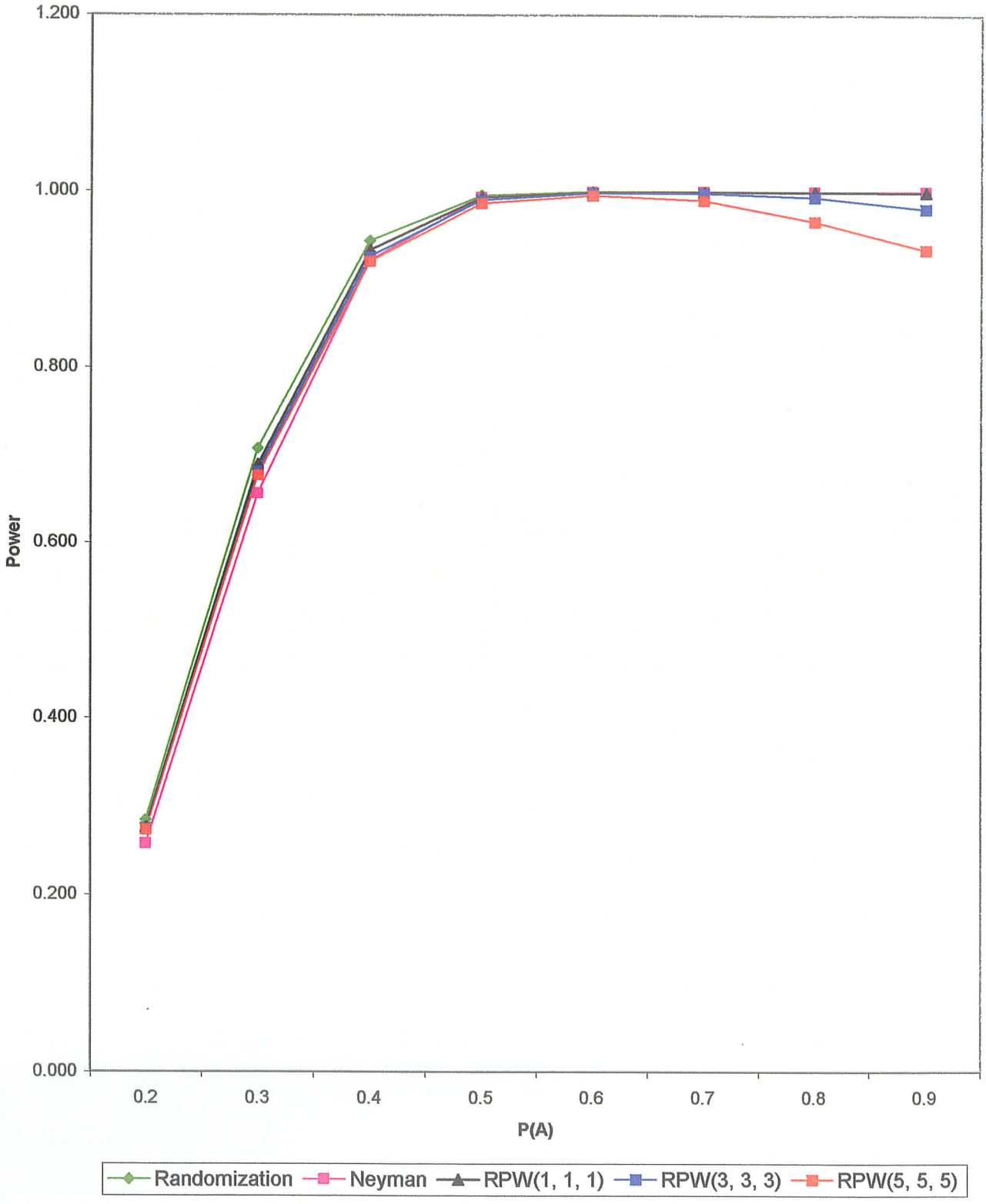
Type of Design	P(A) / P(B)	0.2		0.3		0.4		0.5	
Randomization		0.284	(0.012)	0.707	(0.011)	0.944	(0.003)	0.996	(0.000)
Neyman		0.257	(0.026)	0.655	(0.046)	0.922	(0.024)	0.994	(0.004)
RPW(1, 1, 1)	**0.1**	0.277	(0.012)	0.690	(0.019)	0.934	(0.011)	0.994	(0.003)
RPW(3, 3, 3)		0.275	(0.013)	0.682	(0.025)	0.926	(0.977)	0.991	(0.008)
RPW(5, 5, 5)		0.273	(0.013)	0.677	(0.027)	0.921	(0.022)	0.987	(0.014)
Randomization				0.207	(0.006)	0.586	(0.009)	0.891	(0.004)
Neyman				0.202	(0.009)	0.572	(0.020)	0.884	(0.012)
RPW(1, 1, 1)	**0.2**			0.203	(0.007)	0.567	(0.021)	0.870	(0.022)
RPW(3, 3, 3)				0.202	(0.009)	0.558	(0.029)	0.852	(0.038)
RPW(5, 5, 5)				0.200	(0.010)	0.551	(0.034)	0.836	(0.051)
Randomization						0.178	(0.003)	0.529	(0.006)
Neyman						0.177	(0.004)	0.526	(0.010)
RPW(1, 1, 1)	**0.3**					0.175	(0.006)	0.511	(0.023)
RPW(3, 3, 3)						0.173	(0.008)	0.497	(0.035)
RPW(5, 5, 5)						0.172	(0.009)	0.486	(0.044)
Randomization								0.167	(0.002)
Neyman								0.167	(0.002)
RPW(1, 1, 1)	**0.4**							0.165	(0.005)
RPW(3, 3, 3)								0.162	(0.008)
RPW(5, 5, 5)								0.160	(0.010)

Table 2: Power of the Test (Standard deviations are given in parentheses)

As to be expected, the highest values for power in Table 2 belong to the 50-50 randomization simulations. It has been well established that completely randomized experiments reduce bias and give a high value for power provided the sample size is large enough. Although it may be impossible to improve upon the high power of randomized trials, the main goal of adaptive trials is to improve upon the number of patients who receive the superior treatment without significant loss in power. Indeed, the values for power of the test for the randomized play the winner rule designs do not deviate too far from those of 50-50 randomization or Neyman allocation. In some cases, the power of the RPW designs comes very close to the Neyman allocation probabilities. For example,

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		1.000	(0.000)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
Neyman		1.000	(0.000)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
RPW(1, 1, 1)	**0.1**	1.000	(0.000)	1.000	(0.000)	1.000	(0.001)	0.999	(0.011)
RPW(3, 3, 3)		0.999	(0.005)	0.999	(0.009)	0.994	(0.040)	0.981	(0.059)
RPW(5, 5, 5)		0.996	(0.015)	0.990	(0.048)	0.966	(0.100)	0.934	(0.102)
Randomization		0.989	(0.001)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
Neyman		0.989	(0.002)	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
RPW(1, 1, 1)	**0.2**	0.980	(0.011)	0.998	(0.004)	0.999	(0.004)	0.998	(0.018)
RPW(3, 3, 3)		0.964	(0.031)	0.985	(0.033)	0.979	(0.062)	0.951	(0.096)
RPW(5, 5, 5)		0.945	(0.056)	0.956	(0.083)	0.920	(0.136)	0.869	(0.148)
Randomization		0.863	(0.005)	0.986	(0.001)	1.000	(0.000)	1.000	(0.000)
Neyman		0.861	(0.006)	0.986	(0.001)	1.000	(0.000)	1.000	(0.000)
RPW(1, 1, 1)	**0.3**	0.831	(0.033)	0.966	(0.023)	0.992	(0.015)	0.993	(0.029)
RPW(3, 3, 3)		0.798	(0.063)	0.925	(0.071)	0.943	(0.092)	0.911	(0.125)
RPW(5, 5, 5)		0.768	(0.087)	0.870	(0.123)	0.852	(0.168)	0.806	(0.174)
Randomization		0.513	(0.005)	0.863	(0.005)	0.989	(0.001)	1.000	(0.000)
Neyman		0.512	(0.006)	0.862	(0.006)	0.989	(0.002)	1.000	(0.000)
RPW(1, 1, 1)	**0.4**	0.492	(0.026)	0.820	(0.046)	0.956	(0.039)	0.980	(0.045)
RPW(3, 3, 3)		0.471	(0.047)	0.762	(0.097)	0.870	(0.127)	0.862	(0.152)
RPW(5, 5, 5)		0.453	(0.062)	0.701	(0.140)	0.761	(0.195)	0.745	(0.196)
Randomization		0.167	(0.002)	0.529	(0.006)	0.891	(0.004)	0.996	(0.000)
Neyman		0.167	(0.002)	0.527	(0.009)	0.884	(0.012)	0.994	(0.004)
RPW(1, 1, 1)	**0.5**	0.164	(0.006)	0.505	(0.034)	0.834	(0.065)	0.948	(0.068)
RPW(3, 3, 3)		0.160	(0.011)	0.468	(0.068)	0.731	(0.146)	0.797	(0.178)
RPW(5, 5, 5)		0.156	(0.015)	0.433	(0.094)	0.633	(0.198)	0.685	(0.207)
Randomization				0.178	(0.003)	0.586	(0.009)	0.944	(0.003)
Neyman				0.177	(0.004)	0.573	(0.019)	0.922	(0.023)
RPW(1, 1, 1)	**0.6**			0.174	(0.008)	0.552	(0.049)	0.867	(0.093)
RPW(3, 3, 3)				0.165	(0.018)	0.482	(0.105)	0.705	(0.196)
RPW(5, 5, 5)				0.156	(0.026)	0.419	(0.137)	0.609	(0.231)
Randomization						0.207	(0.006)	0.707	(0.011)
Neyman						0.202	(0.009)	0.657	(0.044)
RPW(1, 1, 1)	**0.7**					0.200	(0.016)	0.649	(0.082)
RPW(3, 3, 3)						0.179	(0.034)	0.519	(0.159)
RPW(5, 5, 5)						0.155	(0.045)	0.433	(0.182)
Randomization								0.283	(0.012)
Neyman								0.258	(0.026)
RPW(1, 1, 1)	**0.8**							0.267	(0.041)
RPW(3, 3, 3)								0.207	(0.072)
RPW(5, 5, 5)								0.161	(0.080)

Figure 2: Power of the Test



when p_A is equal to 0.8 and when p_B is equal to 0.7, then the power of the test using Neyman allocation is 0.202 (0.009) and the power using RPW(1, 1, 1) is 0.200 (0.016). We can also see in the tables that as the difference between p_A and p_B increases, the power of the test increases. This is due to the fact that we are testing for a difference between treatments A and B. The larger the observed difference, the easier it is to make a statistically significant conclusion. The trend seems to be that the more marbles that are drawn and then added, the lower the power of the test, and the more variable the results.

Figure 2 further emphasizes this trend. The graph shows the value of the power of the two-sided test for various designs with p_A varying between 0.2 and 0.9 while p_B is held constant at 0.1. Figure 2 clearly shows that while the randomized play the winner designs do have a lower level of power as compared to 50-50 randomization and Neyman allocation, the decrease is minor. Only when p_A, or both p_A and p_B become quite large does the decrease in the power seem to be more substantial, especially as the parameters of the RPW design increase. This will be discussed further in the monotonicity section. This is the price that must be paid if the expected proportion of patients allocated to the superior treatment is to increase. For the most part, however, the decrease in the power of the RPW designs is small and to be expected.

3.3.3 Odds Ratio

The odds ratio is a quotient of the number of successes versus failures of treatment A over the number of successes versus failures of treatment B. Obviously, if treatment A has a higher probability of success than treatment B, then the odds ratio should be greater than one. Without loss of generality, treatment A has been assumed to be the superior treatment. The data in Table 3 gives the odds ratio for various

probabilities of success of treatment A and B for 50-50 randomization, Neyman allocation, RPW(1, 1, 1), RPW(3, 3, 3) and RPW(5, 5, 5) experimental designs. In the table, all of the odds ratios are consistently greater than one, indicating that treatment A has a higher number of successes versus failures than treatment B. Furthermore, as p_A and p_B increase, so does the odds ratio for all designs.

When the odds ratios are compared for the various designs, it is evident that all designs improve upon 1:1 allocation. This implies that for the same treatment probabilities of success, randomization has a lower proportion of successes versus failures on the superior treatment than all other designs studied. In most cases, the randomized play the winner designs have a higher value for the odds ratio than Neyman allocation. Particularly, when the values of p_A and p_B are greater than 0.5, all RPW designs outperform Neyman allocation. Recall that when p_B is greater than q_A, Neyman allocation gives poor results and allocates less than fifty per cent of subjects to the superior treatment. For smaller values of p_A and p_B, Neyman allocation has the highest odds ratio of all the designs. If the randomized play the winner designs are compared, the RPW(3, 3, 3) improves upon the RPW(1, 1, 1) design in most cases, except when there is a large numerical difference between p_A and p_B. In addition, for some large values of p_A and p_B, RPW(5, 5, 5) has a higher odds ratio than RPW(3, 3, 3), indicating a higher number of successes versus failures on treatment A. It should be noted, however, that the odds ratio was found to be a highly variable statistic; in some cases the standard error was greater than the average value, so all conclusions drawn based on these values are only approximations.

Type of Design	P(A) / P(B)	0.2		0.3		0.4		0.5	
Randomization		2.94	(2.46)	5.04	(3.96)	7.88	(6.05)	11.95	(9.29)
Neyman		3.04	(2.36)	5.23	(3.86)	8.15	(5.87)	12.36	(8.87)
RPW(1, 1, 1)	**0.1**	3.01	(2.47)	5.16	(3.95)	8.06	(6.03)	12.27	(9.06)
RPW(3, 3, 3)		3.03	(2.47)	5.25	(4.07)	8.25	(6.15)	12.17	(8.46)
RPW(5, 5, 5)		3.00	(2.36)	5.23	(4.03)	8.08	(5.85)	11.84	(7.94)
Randomization				1.95	(1.10)	3.06	(1.73)	4.64	(2.68)
Neyman				2.06	(1.32)	3.24	(2.08)	4.91	(3.11)
RPW(1, 1, 1)	**0.2**			2.02	(1.26)	3.19	(2.00)	5.01	(3.62)
RPW(3, 3, 3)				2.03	(1.36)	3.27	(2.30)	5.12	(3.92)
RPW(5, 5, 5)				2.06	(1.37)	3.29	(2.36)	5.21	(3.96)
Randomization						1.71	(0.78)	2.60	(1.21)
Neyman						1.76	(0.86)	2.68	(1.33)
RPW(1, 1, 1)	**0.3**					1.76	(0.87)	2.70	(1.51)
RPW(3, 3, 3)						1.76	(0.92)	2.71	(1.54)
RPW(5, 5, 5)						1.78	(0.94)	2.78	(1.70)
Randomization								1.65	(0.73)
Neyman								1.66	(0.73)
RPW(1, 1, 1)	**0.4**							1.67	(0.76)
RPW(3, 3, 3)								1.69	(0.85)
RPW(5, 5, 5)								1.72	(0.97)

Table 3: Odds Ratio (Standard deviations are given in parentheses)

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		18.25	(14.33)	28.95	(24.42)	51.75	(49.19)	131.50	(139.48)
Neyman		18.73	(13.68)	29.81	(22.58)	53.94	(45.01)	137.97	(147.48)
RPW(1, 1, 1)	**0.1**	18.40	(13.06)	27.53	(18.37)	43.00	(26.46)	78.21	(47.93)
RPW(3, 3, 3)		17.03	(10.77)	23.47	(13.36)	31.11	(16.34)	43.91	(25.17)
RPW(5, 5, 5)		16.19	(9.66)	20.29	(10.77)	23.94	(12.48)	31.30	(20.17)
Randomization		7.08	(4.12)	11.20	(6.75)	20.09	(14.58)	51.57	(46.77)
Neyman		7.45	(4.58)	11.78	(7.35)	21.43	(14.82)	53.11	(44.72)
RPW(1, 1, 1)	**0.2**	7.85	(6.05)	12.79	(10.25)	22.79	(18.11)	49.53	(38.27)
RPW(3, 3, 3)		8.06	(6.47)	12.58	(9.51)	19.76	(13.48)	33.11	(21.64)
RPW(5, 5, 5)		8.06	(6.14)	12.11	(8.40)	16.84	(10.75)	25.95	(18.35)
Randomization		3.97	(1.91)	6.27	(3.19)	11.24	(7.31)	28.86	(23.87)
Neyman		4.07	(2.03)	6.50	(3.59)	11.93	(7.61)	29.77	(22.76)
RPW(1, 1, 1)	**0.3**	4.18	(2.52)	6.87	(4.97)	12.75	(10.20)	32.13	(28.06)
RPW(3, 3, 3)		4.35	(3.16)	7.31	(5.99)	12.80	(10.17)	24.99	(18.93)
RPW(5, 5, 5)		4.45	(3.43)	7.42	(5.93)	11.92	(8.80)	20.69	(16.05)
Randomization		2.51	(1.15)	3.98	(1.93)	7.13	(4.49)	18.27	(14.70)
Neyman		2.54	(1.15)	4.08	(2.08)	7.49	(4.63)	18.55	(13.37)
RPW(1, 1, 1)	**0.4**	2.56	(1.29)	4.15	(2.40)	7.74	(5.83)	20.51	(19.77)
RPW(3, 3, 3)		2.67	(1.66)	4.50	(3.73)	8.26	(6.96)	18.30	(15.39)
RPW(5, 5, 5)		2.71	(1.82)	4.65	(3.75)	8.23	(6.70)	16.19	(12.83)
Randomization		1.65	(0.72)	2.61	(1.12)	4.68	(2.73)	11.99	(9.45)
Neyman		1.67	(0.73)	2.66	(1.34)	4.93	(3.07)	12.19	(8.52)
RPW(1, 1, 1)	**0.5**	1.67	(0.76)	2.68	(1.66)	4.93	(3.41)	13.06	(12.13)
RPW(3, 3, 3)		1.72	(0.89)	2.86	(2.12)	5.44	(4.87)	13.28	(12.34)
RPW(5, 5, 5)		1.75	(1.03)	3.01	(2.51)	5.63	(4.90)	12.39	(10.67)
Randomization				1.72	(0.80)	3.09	(1.83)	7.93	(6.31)
Neyman				1.76	(0.94)	3.24	(2.05)	8.04	(5.69)
RPW(1, 1, 1)	**0.6**			1.75	(0.87)	3.18	(2.05)	8.48	(7.83)
RPW(3, 3, 3)				1.84	(1.17)	3.52	(3.12)	0.83	(0.16)
RPW(5, 5, 5)				1.91	(1.49)	3.80	(3.55)	9.02	(8.42)
Randomization						1.98	(1.19)	5.07	(4.04)
Neyman						2.07	(1.31)	5.18	(3.77)
RPW(1, 1, 1)	**0.7**					2.03	(1.29)	5.43	(5.32)
RPW(3, 3, 3)						2.18	(1.82)	6.16	(6.96)
RPW(5, 5, 5)						2.43	(2.32)	6.40	(6.51)
Randomization								2.94	(2.46)
Neyman								3.02	(2.31)
RPW(1, 1, 1)	**0.8**							3.11	(2.99)
RPW(3, 3, 3)								3.67	(4.27)
RPW(5, 5, 5)								4.08	(4.78)

3.3.4 Proportion of Successes

The results obtained from the RPW(2n+1, 2n+1, 2n+1) design must be verified to ensure that they are similar to results obtained from other designs. This helps to ascertain that there are no apparent flaws in the design and to ensure that the proportion of successes on the conventional treatment and on the experimental treatment were approximately equal to the allocation probabilities of those treatments. Table 4 gives the total proportion of successes on treatment A, which is equal to the total number of successes on treatment A divided by the total number of patients on treatment A. Table 5 gives similar data for treatment B. Both Table 4 and Table 5 show that the expected proportion of successes on treatment A and B are approximately equal to the allocation probabilities of treatments A and B, respectively. The table values are similar for all designs simulated: 50-50 randomization, Neyman allocation, RPW(1, 1, 1), RPW(3, 3, 3) and RPW(5, 5, 5). Thus, this is a good indication that the RPW design is allocating subjects to treatments properly.

Type of Design	P(A) / P(B)	0.2		0.3		0.4		0.5	
Randomization		0.200	(0.057)	0.299	(0.066)	0.399	(0.070)	0.499	(0.071)
Neyman		0.200	(0.053)	0.300	(0.059)	0.399	(0.062)	0.500	(0.064)
RPW(1, 1, 1)	**0.1**	0.199	(0.055)	0.298	(0.060)	0.397	(0.063)	0.497	(0.063)
RPW(3, 3, 3)		0.199	(0.054)	0.297	(0.060)	0.397	(0.062)	0.497	(0.062)
RPW(5, 5, 5)		0.198	(0.054)	0.297	(0.060)	0.396	(0.061)	0.496	(0.060)
Randomization				0.299	(0.066)	0.399	(0.070)	0.499	(0.071)
Neyman				0.300	(0.063)	0.400	(0.066)	0.500	(0.067)
RPW(1, 1, 1)	**0.2**			0.298	(0.062)	0.397	(0.065)	0.497	(0.064)
RPW(3, 3, 3)				0.297	(0.062)	0.397	(0.064)	0.497	(0.063)
RPW(5, 5, 5)				0.297	(0.062)	0.395	(0.063)	0.496	(0.062)
Randomization						0.399	(0.070)	0.499	(0.071)
Neyman						0.399	(0.068)	0.500	(0.069)
RPW(1, 1, 1)	**0.3**					0.397	(0.067)	0.497	(0.066)
RPW(3, 3, 3)						0.396	(0.067)	0.497	(0.065)
RPW(5, 5, 5)						0.395	(0.066)	0.495	(0.064)
Randomization								0.499	(0.071)
Neyman								0.500	(0.070)
RPW(1, 1, 1)	**0.4**							0.496	(0.068)
RPW(3, 3, 3)								0.496	(0.069)
RPW(5, 5, 5)								0.494	(0.067)

Table 4: Treatment A Proportion of Successes (Standard deviations are given in parentheses)

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		0.601	(0.070)	0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman		0.601	(0.063)	0.701	(0.059)	0.802	(0.054)	0.903	(0.043)
RPW(1, 1, 1)	**0.1**	0.598	(0.061)	0.698	(0.054)	0.799	(0.046)	0.899	(0.033)
RPW(3, 3, 3)		0.598	(0.058)	0.699	(0.052)	0.799	(0.042)	0.900	(0.031)
RPW(5, 5, 5)		0.596	(0.057)	0.698	(0.050)	0.798	(0.042)	0.900	(0.030)
Randomization		0.601	(0.070)	0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman		0.601	(0.067)	0.702	(0.063)	0.803	(0.058)	0.903	(0.047)
RPW(1, 1, 1)	**0.2**	0.598	(0.062)	0.698	(0.055)	0.799	(0.047)	0.899	(0.033)
RPW(3, 3, 3)		0.597	(0.059)	0.698	(0.053)	0.798	(0.043)	0.899	(0.031)
RPW(5, 5, 5)		0.595	(0.058)	0.697	(0.051)	0.798	(0.043)	0.900	(0.030)
Randomization		0.601	(0.070)	0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman		0.601	(0.069)	0.702	(0.065)	0.803	(0.060)	0.904	(0.049)
RPW(1, 1, 1)	**0.3**	0.597	(0.063)	0.698	(0.056)	0.799	(0.048)	0.899	(0.034)
RPW(3, 3, 3)		0.597	(0.061)	0.698	(0.054)	0.798	(0.044)	0.899	(0.031)
RPW(5, 5, 5)		0.595	(0.059)	0.696	(0.052)	0.798	(0.043)	0.899	(0.030)
Randomization		0.601	(0.070)	0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman		0.601	(0.070)	0.702	(0.066)	0.803	(0.061)	0.904	(0.050)
RPW(1, 1, 1)	**0.4**	0.597	(0.065)	0.697	(0.058)	0.798	(0.049)	0.899	(0.034)
RPW(3, 3, 3)		0.596	(0.064)	0.697	(0.056)	0.797	(0.045)	0.899	(0.032)
RPW(5, 5, 5)		0.594	(0.062)	0.696	(0.054)	0.797	(0.045)	0.899	(0.031)
Randomization		0.601	(0.070)	0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman		0.602	(0.071)	0.702	(0.067)	0.804	(0.061)	0.904	(0.050)
RPW(1, 1, 1)	**0.5**	0.596	(0.068)	0.696	(0.060)	0.797	(0.051)	0.899	(0.035)
RPW(3, 3, 3)		0.595	(0.068)	0.696	(0.059)	0.797	(0.047)	0.899	(0.033)
RPW(5, 5, 5)		0.592	(0.067)	0.694	(0.058)	0.796	(0.047)	0.899	(0.032)
Randomization				0.700	(0.065)	0.799	(0.057)	0.900	(0.042)
Neyman				0.702	(0.066)	0.803	(0.061)	0.904	(0.050)
RPW(1, 1, 1)	**0.6**			0.696	(0.064)	0.797	(0.053)	0.898	(0.037)
RPW(3, 3, 3)				0.694	(0.065)	0.795	(0.051)	0.898	(0.034)
RPW(5, 5, 5)				0.690	(0.066)	0.794	(0.052)	0.898	(0.035)
Randomization						0.799	(0.057)	0.900	(0.042)
Neyman						0.803	(0.060)	0.904	(0.049)
RPW(1, 1, 1)	**0.7**					0.796	(0.057)	0.898	(0.039)
RPW(3, 3, 3)						0.792	(0.060)	0.897	(0.040)
RPW(5, 5, 5)						0.791	(0.069)	0.896	(0.097)
Randomization								0.900	(0.042)
Neyman								0.903	(0.047)
RPW(1, 1, 1)	**0.8**							0.897	(0.044)
RPW(3, 3, 3)								0.893	(0.055)
RPW(5, 5, 5)								0.889	(0.091)

Type of Design	P(A) / P(B)	0.2		0.3		0.4		0.5	
Randomization		0.101	(0.043)	0.101	(0.043)	0.101	(0.043)	0.101	(0.043)
Neyman		0.097	(0.046)	0.096	(0.048)	0.096	(0.050)	0.096	(0.050)
RPW(1, 1, 1)	**0.1**	0.099	(0.043)	0.099	(0.045)	0.099	(0.047)	0.098	(0.049)
RPW(3, 3, 3)		0.098	(0.044)	0.098	(0.046)	0.097	(0.049)	0.096	(0.053)
RPW(5, 5, 5)		0.098	(0.044)	0.098	(0.046)	0.097	(0.050)	0.097	(0.055)
Randomization				0.210	(0.057)	0.210	(0.057)	0.210	(0.057)
Neyman				0.198	(0.060)	0.198	(0.061)	0.197	(0.062)
RPW(1, 1, 1)	**0.2**			0.198	(0.059)	0.198	(0.061)	0.197	(0.065)
RPW(3, 3, 3)				0.198	(0.059)	0.197	(0.063)	0.197	(0.068)
RPW(5, 5, 5)				0.197	(0.060)	0.196	(0.064)	0.194	(0.070)
Randomization						0.301	(0.065)	0.301	(0.065)
Neyman						0.297	(0.068)	0.297	(0.068)
RPW(1, 1, 1)	**0.3**					0.296	(0.068)	0.295	(0.072)
RPW(3, 3, 3)						0.297	(0.068)	0.296	(0.073)
RPW(5, 5, 5)						0.294	(0.069)	0.292	(0.075)
Randomization								0.400	(0.069)
Neyman								0.398	(0.071)
RPW(1, 1, 1)	**0.4**							0.395	(0.073)
RPW(3, 3, 3)								0.394	(0.075)
RPW(5, 5, 5)								0.390	(0.078)

Table 5: Treatment B Proportion of Successes (Standard deviations are given in parentheses)

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		0.101	(0.043)	0.101	(0.043)	0.101	(0.043)	0.101	(0.043)
Neyman		0.096	(0.050)	0.096	(0.048)	0.097	(0.046)	0.098	(0.043)
RPW(1, 1, 1)	**0.1**	0.098	(0.053)	0.097	(0.058)	0.095	(0.067)	0.092	(0.083)
RPW(3, 3, 3)		0.096	(0.060)	0.094	(0.068)	0.091	(0.087)	0.083	(0.125)
RPW(5, 5, 5)		0.096	(0.063)	0.093	(0.079)	0.088	(0.109)	0.072	(0.147)
Randomization		0.210	(0.057)	0.210	(0.057)	0.210	(0.057)	0.210	(0.057)
Neyman		0.197	(0.061)	0.198	(0.060)	0.198	(0.058)	0.199	(0.053)
RPW(1, 1, 1)	**0.2**	0.196	(0.069)	0.195	(0.076)	0.191	(0.087)	0.185	(0.110)
RPW(3, 3, 3)		0.196	(0.076)	0.193	(0.089)	0.187	(0.114)	0.172	(0.167)
RPW(5, 5, 5)		0.193	(0.081)	0.188	(0.103)	0.176	(0.144)	0.150	(0.203)
Randomization		0.301	(0.065)	0.301	(0.065)	0.301	(0.065)	0.301	(0.065)
Neyman		0.297	(0.067)	0.297	(0.066)	0.297	(0.064)	0.298	(0.060)
RPW(1, 1, 1)	**0.3**	0.294	(0.076)	0.292	(0.083)	0.288	(0.094)	0.278	(0.119)
RPW(3, 3, 3)		0.293	(0.082)	0.290	(0.097)	0.283	(0.126)	0.261	(0.191)
RPW(5, 5, 5)		0.290	(0.088)	0.283	(0.113)	0.267	(0.161)	0.229	(0.236)
Randomization		0.400	(0.069)	0.400	(0.069)	0.400	(0.069)	0.400	(0.069)
Neyman		0.398	(0.070)	0.398	(0.069)	0.398	(0.067)	0.398	(0.063)
RPW(1, 1, 1)	**0.4**	0.394	(0.078)	0.392	(0.085)	0.387	(0.097)	0.378	(0.128)
RPW(3, 3, 3)		0.391	(0.084)	0.387	(0.099)	0.379	(0.129)	0.352	(0.203)
RPW(5, 5, 5)		0.387	(0.090)	0.380	(0.117)	0.362	(0.172)	0.316	(0.258)
Randomization		0.500	(0.071)	0.500	(0.071)	0.500	(0.071)	0.500	(0.071)
Neyman		0.499	(0.071)	0.499	(0.070)	0.499	(0.068)	0.499	(0.064)
RPW(1, 1, 1)	**0.5**	0.494	(0.076)	0.493	(0.084)	0.489	(0.096)	0.479	(0.122)
RPW(3, 3, 3)		0.491	(0.080)	0.487	(0.096)	0.477	(0.129)	0.445	(0.207)
RPW(5, 5, 5)		0.486	(0.086)	0.479	(0.113)	0.461	(0.170)	0.411	(0.267)
Randomization				0.601	(0.070)	0.601	(0.070)	0.601	(0.070)
Neyman				0.601	(0.069)	0.601	(0.066)	0.600	(0.063)
RPW(1, 1, 1)	**0.6**			0.594	(0.078)	0.592	(0.089)	0.582	(0.115)
RPW(3, 3, 3)				0.587	(0.087)	0.578	(0.119)	0.545	(0.202)
RPW(5, 5, 5)				0.581	(0.102)	0.560	(0.163)	0.511	(0.266)
Randomization						0.700	(0.065)	0.700	(0.065)
Neyman						0.701	(0.063)	0.700	(0.059)
RPW(1, 1, 1)	**0.7**					0.692	(0.077)	0.683	(0.101)
RPW(3, 3, 3)						0.681	(0.105)	0.654	(0.185)
RPW(5, 5, 5)						0.665	(0.148)	0.619	(0.252)
Randomization								0.801	(0.006)
Neyman								0.801	(0.053)
RPW(1, 1, 1)	**0.8**							0.789	(0.081)
RPW(3, 3, 3)								0.767	(0.156)
RPW(5, 5, 5)								0.736	(0.227)

Table 6 gives the total proportion of successes on both treatments A and B divided by the total number of patients treated. Again, the same properties present in Tables 4 and 5 are present here. The values in this table hold more significance however because the greater the total proportion of successes, the greater the number of patients cured. Moreover, a higher total proportion of successes may imply that the design is better at allocating patients to the superior treatment. The proportions for 50-50 randomization increase linearly as the probabilities of success of treatments A and B increase. The proportions for Neyman allocation are greater than those of 50-50 randomization until the probability of success of treatment B is greater than the probability of failure of treatment A. For example, when $p_B = 0.7$ and $p_A = 0.9$, then the total proportion of successes is 0.777 (0.041), which is lower than 0.80 (0.040), the value for 50-50 randomization.

The proportions for the randomized play the winner designs are almost consistently higher than 50-50 randomization and Neyman allocation except when both p_A and p_B are small. Furthermore, as the number of balls drawn then added increases from one to three to five, so does the proportion of successes. This trend can be visualised by Figure 3. Figure 3 is a graph of total proportion of successes for various values of p_A while p_B is held constant at 0.1. As the graph indicates, the randomized play the winner rules perform better than both 1:1 allocation and Neyman allocation in terms of increasing the number of successes in the trial. In addition, RPW(5, 5, 5) outperforms RPW(3, 3, 3), which outperforms RPW(1, 1, 1). Note, however, that the increase in the total proportion of successes is greater between RPW(1, 1, 1) and RPW(3, 3, 3) than RPW(3, 3, 3) and RPW(5, 5, 5). This may indicate that as the parameters of the design increase, the improvement in the total number of successes is not as significant. Even so,

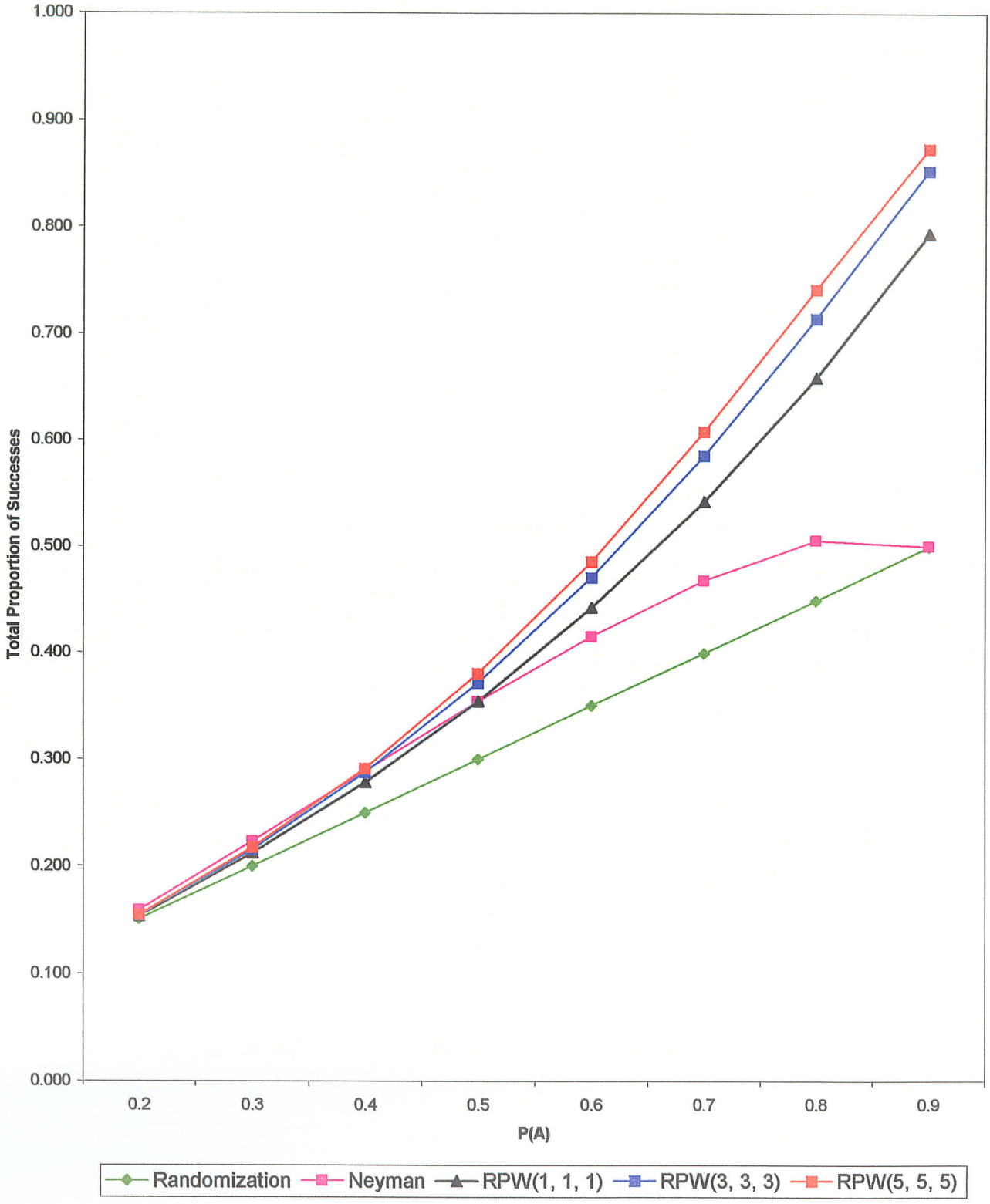
the more balls drawn, then added to the urn, the higher the total proportion of successes in the trial. This implies that more patients are successfully treated when randomized play the winner designs are used in clinical trials.

Type of Design	P(A) / P(B)	0.2	0.3	0.4	0.5
Randomization		0.150 (0.036)	0.200 (0.040)	0.250 (0.043)	0.300 (0.046)
Neyman		0.159 (0.036)	0.223 (0.041)	0.289 (0.044)	0.354 (0.045)
RPW(1, 1, 1)	**0.1**	0.154 (0.036)	0.213 (0.042)	0.279 (0.049)	0.355 (0.054)
RPW(3, 3, 3)		0.154 (0.037)	0.215 (0.044)	0.287 (0.051)	0.372 (0.059)
RPW(5, 5, 5)		0.154 (0.037)	0.217 (0.045)	0.291 (0.053)	0.380 (0.061)
Randomization			0.250 (0.044)	0.300 (0.046)	0.350 (0.004)
Neyman			0.254 (0.009)	0.311 (0.045)	0.368 (0.046)
RPW(1, 1, 1)	**0.2**		0.254 (0.044)	0.314 (0.049)	0.383 (0.053)
RPW(3, 3, 3)			0.255 (0.044)	0.319 (0.050)	0.395 (0.057)
RPW(5, 5, 5)			0.255 (0.044)	0.320 (0.050)	0.400 (0.058)
Randomization				0.350 (0.048)	0.400 (0.049)
Neyman				0.351 (0.048)	0.404 (0.049)
RPW(1, 1, 1)	**0.3**			0.353 (0.049)	0.416 (0.052)
RPW(3, 3, 3)				0.356 (0.049)	0.423 (0.054)
RPW(5, 5, 5)				0.355 (0.049)	0.425 (0.055)
Randomization					0.450 (0.050)
Neyman					0.450 (0.050)
RPW(1, 1, 1)	**0.4**				0.454 (0.051)
RPW(3, 3, 3)					0.457 (0.051)
RPW(5, 5, 5)					0.457 (0.052)

Table 6: Total Proportion of Successes (Standard deviations are given in parentheses)

Type of Design	P(A) / P(B)	0.6		0.7		0.8		0.9	
Randomization		0.350	(0.047)	0.400	(0.048)	0.449	(0.049)	0.500	(0.049)
Neyman		0.415	(0.045)	0.468	(0.045)	0.506	(0.044)	0.500	(0.050)
RPW(1, 1, 1)	**0.1**	0.443	(0.060)	0.542	(0.063)	0.659	(0.065)	0.794	(0.058)
RPW(3, 3, 3)		0.471	(0.065)	0.585	(0.068)	0.714	(0.063)	0.853	(0.048)
RPW(5, 5, 5)		0.485	(0.068)	0.608	(0.070)	0.741	(0.062)	0.873	(0.042)
Randomization		0.401	(0.047)	0.450	(0.049)	0.500	(0.049)	0.550	(0.049)
Neyman		0.422	(0.045)	0.469	(0.043)	0.501	(0.041)	0.496	(0.043)
RPW(1, 1, 1)	**0.2**	0.464	(0.058)	0.557	(0.061)	0.667	(0.063)	0.795	(0.058)
RPW(3, 3, 3)		0.486	(0.062)	0.593	(0.066)	0.717	(0.062)	0.853	(0.049)
RPW(5, 5, 5)		0.497	(0.065)	0.613	(0.068)	0.742	(0.062)	0.873	(0.043)
Randomization		0.450	(0.050)	0.500	(0.050)	0.549	(0.049)	0.600	(0.048)
Neyman		0.455	(0.048)	0.500	(0.045)	0.532	(0.043)	0.532	(0.045)
RPW(1, 1, 1)	**0.3**	0.489	(0.056)	0.574	(0.059)	0.676	(0.061)	0.797	(0.057)
RPW(3, 3, 3)		0.504	(0.058)	0.603	(0.063)	0.720	(0.061)	0.852	(0.050)
RPW(5, 5, 5)		0.512	(0.062)	0.620	(0.066)	0.743	(0.062)	0.871	(0.044)
Randomization		0.500	(0.050)	0.550	(0.050)	0.599	(0.048)	0.650	(0.047)
Neyman		0.500	(0.049)	0.545	(0.047)	0.578	(0.002)	0.585	(0.046)
RPW(1, 1, 1)	**0.4**	0.519	(0.054)	0.595	(0.056)	0.688	(0.058)	0.800	(0.056)
RPW(3, 3, 3)		0.528	(0.055)	0.616	(0.060)	0.725	(0.060)	0.851	(0.050)
RPW(5, 5, 5)		0.531	(0.057)	0.628	(0.063)	0.744	(0.062)	0.869	(0.047)
Randomization		0.550	(0.050)	0.600	(0.049)	0.649	(0.047)	0.700	(0.045)
Neyman		0.550	(0.050)	0.595	(0.048)	0.632	(0.047)	0.646	(0.046)
RPW(1, 1, 1)	**0.5**	0.555	(0.051)	0.622	(0.052)	0.705	(0.055)	0.807	(0.054)
RPW(3, 3, 3)		0.559	(0.051)	0.635	(0.055)	0.731	(0.058)	0.849	(0.054)
RPW(5, 5, 5)		0.558	(0.053)	0.641	(0.058)	0.746	(0.062)	0.865	(0.052)
Randomization				0.650	(0.048)	0.700	(0.046)	0.750	(0.043)
Neyman				0.649	(0.047)	0.690	(0.046)	0.711	(0.044)
RPW(1, 1, 1)	**0.6**			0.657	(0.049)	0.728	(0.050)	0.817	(0.051)
RPW(3, 3, 3)				0.661	(0.050)	0.742	(0.054)	0.849	(0.056)
RPW(5, 5, 5)				0.662	(0.051)	0.750	(0.059)	0.859	(0.059)
Randomization						0.750	(0.043)	0.800	(0.040)
Neyman						0.747	(0.044)	0.777	(0.041)
RPW(1, 1, 1)	**0.7**					0.758	(0.045)	0.833	(0.046)
RPW(3, 3, 3)						0.763	(0.046)	0.852	(0.053)
RPW(5, 5, 5)						0.764	(0.050)	0.855	(0.060)
Randomization								0.850	(0.036)
Neyman								0.842	(0.037)
RPW(1, 1, 1)	**0.8**							0.859	(0.038)
RPW(3, 3, 3)								0.865	(0.043)
RPW(5, 5, 5)								0.865	(0.047)

Figure 3: Total Proportion of Successes



3.4 Monotonicity of RPW(μ, α, β; γ)

In order to study the monotonic properties on the randomized play the winner design, a new notation will be introduced to fully represent all the different values the parameters may take. The general case of the randomized play the winner rule design shall be denoted as RPW(μ, α, β; γ). The parameters in the design include μ, the number of balls at the start of the trial of each type, α, the number of balls added after the response of the patient was a success, β, the number of balls added when the response was a failure and γ, the number of balls drawn. Note that it is possible that at the beginning of the trial for the number of balls of each type to be unequal. If that were the case then μ would actually have two levels, and the randomized play the winner design would be denoted as RPW(μ_1, μ_2; α, β; γ). Beginning the trial with an uneven sample size may favour one treatment over the other and also may reduce power. In this paper only the case where the trial begins with an equal number of balls will be explored.

Varying any or all of the parameters in the randomized play the winner design enables us to see its monotonic properties. If the parameters are allowed to systematically increase, and a trend emerges with respect to the parameters, then the design can be considered monotonic. The monotonicity of the RPW(μ, α, β; γ) design is explored with respect to both the expected proportion on the superior treatment and the power of the test. The parameters are increased simultaneously as well as individually while holding all others constant to determine how strongly each parameter affects the results of the statistic under study. All simulations use 2000 replications, each with a sample size of one hundred patients unless otherwise indicated. As before, the program SAS was used for simulation studies. The goals of the monotonicity study are to not only determine which parameters hold the most influence on the RPW design, but also to

hopefully discover an optimum level for the parameters in the design to maximize the expected proportion of patients allocated to the superior treatment without significant sacrifices in power.

3.4.1 Monotonicity in all Parameters Concurrently

Studying the monotonic properties of the RPW(2n+1, 2n+1, 2n+1; 2n+1) design enables us to understand the trends of the design as n increases. Tables 7 and 8 show the monotonic properties of various randomized play the winner designs for the expected proportion on the superior treatment and the power of the test for various probabilities of treatment A and B. Also, the tables include values for 50-50 randomization so we are able to compare the randomized play the winner designs with the classical 50:50 randomized design.

Table 7 shows an increasing trend in the expected proportion as n increases while Table 8 shows a decreasing trend in the power of the test. These trends are clearly visible in Figure 4, which plots the expected proportion against the power of the test for various values of n as well as randomization when $p_A = 0.7$ and $p_B = 0.4$. The trend visible in the graph is that as the number of balls in the urn increases, there is a higher chance that a patient will receive the superior treatment, but there is also less likely of a chance of reaching a valid conclusion at the end of the trial. The intuition behind these trends is that as n increases, the trial focuses more on allocating patients to the superior treatment and less on information gathering. So, increasingly more patients receive a better treatment, but the trial becomes less balanced, causing a substantial decrease in power.

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1; 2n+1) design, where n is							
		0	1	2	4	8	16	32	64
P(A) = 0.2	0.500	0.528	0.536	0.539	0.545	0.551	0.555	0.559	0.562
P(B) = 0.1	(0.048)	(0.040)	(0.041)	(0.041)	(0.043)	(0.043)	(0.044)	(0.048)	(0.047)
P(A) = 0.5	0.500	0.637	0.678	0.702	0.732	0.765	0.801	0.827	0.851
P(B) = 0.1	(0.048)	(0.051)	(0.058)	(0.058)	(0.064)	(0.068)	(0.070)	(0.075)	(0.075)
P(A) = 0.9	0.500	0.868	0.941	0.967	0.982	0.989	0.992	0.992	0.992
P(B) = 0.1	(0.048)	(0.051)	(0.036)	(0.027)	(0.018)	(0.013)	(0.010)	(0.009)	(0.010)
P(A) = 0.5	0.500	0.544	0.559	0.573	0.585	0.608	0.632	0.655	0.674
P(B) = 0.4	(0.048)	(0.070)	(0.090)	(0.098)	(0.114)	(0.137)	(0.158)	(0.179)	(0.193)
P(A) = 0.7	0.500	0.652	0.718	0.763	0.814	0.871	0.907	0.929	0.934
P(B) = 0.4	(0.048)	(0.082)	(0.104)	(0.114)	(0.120)	(0.121)	(0.122)	0.113	(0.112)
P(A) = 0.9	0.500	0.803	0.901	0.938	0.963	0.973	0.975	0.976	0.975
P(B) = 0.4	(0.048)	(0.086)	(0.079)	(0.070)	(0.059)	(0.054)	(0.062)	(0.057)	(0.061)
P(A) = 0.8	0.500	0.633	0.712	0.750	0.799	0.813	0.814	0.781	0.782
P(B) = 0.6	(0.048)	(0.116)	(0.166)	(0.205)	(0.234)	(0.283)	(0.318)	(0.367)	(0.375)
P(A) = 0.9	0.500	0.722	0.830	0.862	0.892	0.887	0.879	0.838	0.842
P(B) = 0.6	(0.048)	(0.126)	(0.155)	(0.180)	(0.193)	(0.233)	(0.268)	(0.328)	(0.333)
P(A) = 0.9	0.500	0.592	0.650	0.658	0.640	0.614	0.611	0.592	0.603
P(B) = 0.8	(0.048)	(0.181)	(0.298)	(0.366)	(0.428)	(0.467)	(0.479)	(0.487)	(0.485)

Table 7: Expected Proportion on the Superior Treatment for RPW(2n+1, 2n+1, 2n+1; 2n+1) (Standard deviations are given in parentheses)

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1; 2n+1) design, where n is							
		0	1	2	4	8	16	32	64
P(A) = 0.2	0.284	0.277	0.275	0.274	0.272	0.270	0.268	0.267	0.265
P(B) = 0.1	(0.012)	(0.012)	(0.013)	(0.131)	(0.014)	(0.015)	(0.016)	(0.017)	(0.017)
P(A) = 0.5	0.996	0.994	0.990	0.987	0.978	0.957	0.917	0.861	0.796
P(B) = 0.1	(0.000)	(0.003)	(0.008)	(0.015)	(0.033)	(0.070)	(0.128)	(0.186)	(0.239)
P(A) = 0.9	1.000	0.999	0.980	0.933	0.868	0.817	0.789	0.786	0.783
P(B) = 0.1	(0.000)	(0.011)	(0.060)	(0.102)	(0.120)	(0.113)	(0.099)	(0.098)	(0.096)
P(A) = 0.5	0.167	0.165	0.162	0.160	0.157	0.152	0.146	0.139	0.133
P(B) = 0.4	(0.002)	(0.005)	(0.008)	(0.010)	(0.013)	(0.102)	(0.025)	(0.031)	(0.035)
P(A) = 0.7	0.863	0.820	0.759	0.698	0.605	0.479	0.384	0.335	0.322
P(B) = 0.4	(0.005)	(0.046)	(0.100)	(0.142)	(0.192)	(0.227)	(0.240)	(0.233)	(0.236)
P(A) = 0.9	1.000	0.980	0.862	0.745	0.649	0.600	0.585	0.583	0.590
P(B) = 0.4	(0.000)	(0.045)	(0.152)	(0.197)	(0.207)	0.206	0.210	0.206	0.215
P(A) = 0.8	0.586	0.553	0.482	0.418	0.340	0.278	0.237	0.204	0.189
P(B) = 0.6	(0.008)	(0.049)	(0.105)	(0.138)	(0.159)	(0.160)	(0.157)	(0.148)	(0.140)
P(A) = 0.9	0.944	0.867	0.706	0.611	0.534	0.496	0.443	0.407	0.386
P(B) = 0.6	(0.003)	(0.093)	(0.194)	(0.230)	(0.242)	(0.245)	(0.244)	(0.246)	(0.233)
P(A) = 0.9	0.284	0.267	0.208	0.163	0.119	0.097	0.089	0.085	0.089
P(B) = 0.8	(0.012)	(0.040)	(0.072)	(0.080)	(0.078)	(0.071)	(0.061)	(0.056)	(0.052)

Table 8: Power of the Test for RPW(2n+1, 2n+1, 2n+1; 2n+1) (Standard deviations are given in parentheses)

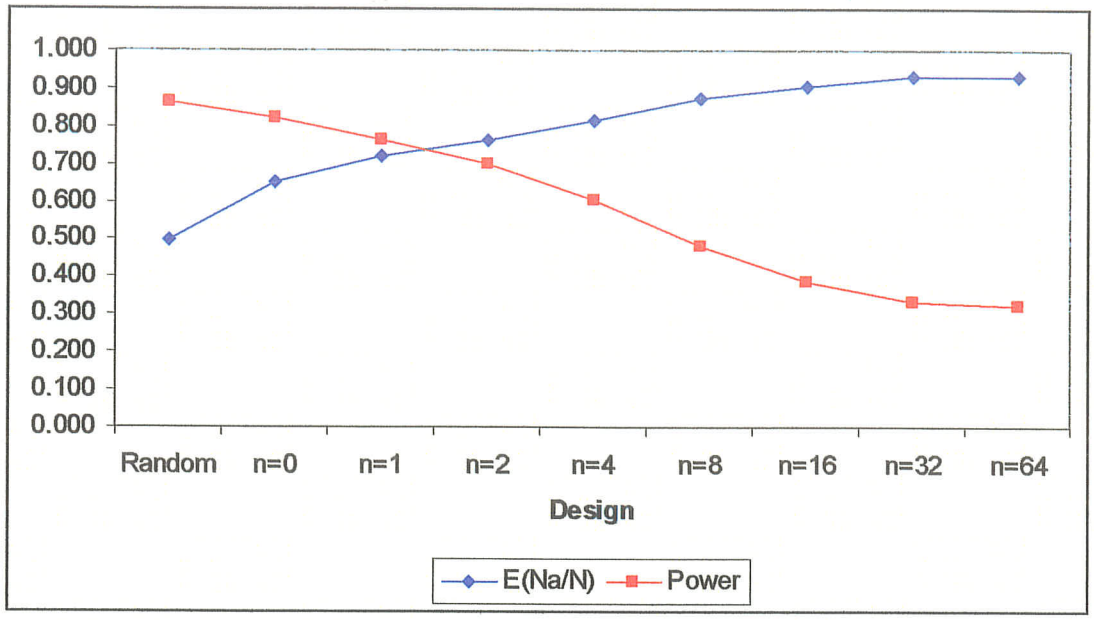


Figure 4: Monotonicity of RPW(2n+1, 2n+1, 2n+1; 2n+1)

Note that there are some exceptions to these trends, particularly in Table 7 when p_B is large. This is likely due to the large number of type B balls and the higher chance of drawing one of them. If a type B ball is drawn early in the trial, and the response is a success, then a large amount of B balls are then added to the urn. This will further increase the chances of receiving treatment B. The probability of receiving treatment B increases with each success on that treatment. So, more patients may actually receive the inferior treatment and the trial can easily become skewed as n increases. For the most part, however, as n increases the expected proportion on the superior treatment is quite an improvement over the classical randomized design, which only allocates fifty per cent of patients to either treatment. The values for power of the test are highest when 50-50 randomization is used, and then steadily decrease as n increases, except when n = 64, p_A = 0.9 and p_B = 0.8. While it is beneficial to have the expected proportion on the better treatment to be as high as possible, any trial would be a waste if no valid statistical

conclusion could be drawn. Thus, it is necessary to choose a value of n so that the RPW(2n+1, 2n+1, 2n+1; 2n+1) design gives valid statistical results with less ethical compromise to the patients in the trial.

3.4.2 Monotonicity in μ

Increasing the number of balls of each type at the beginning of the trial, denoted μ, reduces the rate at which the treatment assignment probabilities adjust. If the treatment assignment probabilities adjust slowly, then it takes longer for the trial to reveal the superior treatment, and thus a smaller proportion of patients receive that treatment. This has a slowing effect on the trial and provides a level playing field for the treatments. In a sense, more weight is given to randomization than adaptation. Note that all other variables are kept constant at three (similar properties are expected at other fixed values.) When the number of balls added to the urn after a response are small compared to the number of balls already in the urn, those added do little to change the treatment probabilities. Thus, the larger the initial size of the urn, the slower the trial progresses, and the less extreme the results.

Table 9 shows the monotonic properties of various randomized play the winner designs for the expected proportion on the superior treatment and Table 10 shows the monotonic properties for the power of the test. In Table 9 we see that as n increases, a decreasing trend can be seen in the expected proportion to the superior treatment. In Table 10 an increasing trend can be seen in the power of the test as n increases. For large values of n, the initial urn size is quite large. Whether more balls of type A or type B are added after a response, the number added is so few in comparison to the total number of balls in the urn that it has little effect to change to the treatment probabilities. The effect

will not be seen until towards the end of the trial, in which case the trial could have been already skewed to the inferior treatment. These results indicate that the initial urn size places a substantial role in the outcome of the trial. A large initial urn size in comparison to the number of balls drawn and added will reduce the effect of adding or drawing balls, which slows the adaptation of the treatment assignment probabilities. This reduces the number of patients who receive the superior treatment overall but actually increases the power of the test. Particularly, if both p_A and p_B are large, it is best to keep the initial urn size small.

Treatment probability	RPW(2n+1, 3, 3; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2	0.536	0.536	0.535	0.533	0.530	0.526	0.520
P(B) = 0.1	(0.041)	(0.041)	(0.040)	(0.040)	(0.039)	(0.039)	(0.039)
P(A) = 0.5	0.678	0.675	0.669	0.658	0.641	0.616	0.587
P(B) = 0.1	(0.058)	(0.057)	(0.056)	(0.054)	(0.051)	(0.048)	(0.046)
P(A) = 0.9	0.941	0.931	0.912	0.881	0.832	0.767	0.692
P(B) = 0.1	(0.036)	(0.036)	(0.038)	(0.040)	(0.043)	(0.045)	(0.047)
P(A) = 0.5	0.559	0.558	0.555	0.550	0.542	0.534	0.524
P(B) = 0.4	(0.090)	(0.088)	(0.084)	(0.078)	(0.070)	(0.062)	(0.055)
P(A) = 0.7	0.718	0.712	0.696	0.675	0.646	0.612	0.578
P(B) = 0.4	(0.104)	(0.100)	(0.093)	(0.086)	(0.077)	(0.068)	(0.060)
P(A) = 0.9	0.901	0.887	0.861	0.822	0.765	0.700	0.636
P(B) = 0.4	(0.079)	(0.078)	(0.077)	(0.074)	(0.071)	(0.067)	(0.061)
P(A) = 0.8	0.712	0.701	0.683	0.653	0.621	0.587	0.557
P(B) = 0.6	(0.166)	(0.154)	(0.139)	(0.121)	(0.101)	(0.083)	(0.068)
P(A) = 0.9	0.830	0.816	0.789	0.744	0.691	0.637	0.589
P(B) = 0.6	(0.155)	(0.144)	(0.131)	(0.116)	(0.100)	(0.083)	(0.069)
P(A) = 0.9	0.650	0.646	0.632	0.605	0.577	0.552	0.532
P(B) = 0.8	(0.298)	(0.267)	(0.226)	(0.181)	(0.138)	(0.102)	(0.078)

Table 9: Expected Proportion on the Superior Treatment for RPW(2n+1, 3, 3; 3)

(Standard deviations are given in parentheses)

Treatment probability	RPW(2n+1, 3, 3; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2	0.041	0.275	0.275	0.276	0.276	0.278	0.279
P(B) = 0.1	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
P(A) = 0.5	0.990	0.991	0.992	0.993	0.994	0.995	0.996
P(B) = 0.1	(0.008)	(0.007)	(0.006)	(0.005)	(0.003)	(0.002)	(0.001)
P(A) = 0.9	0.980	0.990	0.997	1.000	1.000	1.000	1.000
P(B) = 0.1	(0.060)	(0.040)	(0.021)	(0.009)	(0.000)	(0.000)	(0.000)
P(A) = 0.5	0.162	0.162	0.163	0.164	0.165	0.166	0.166
P(B) = 0.4	(0.008)	(0.008)	(0.007)	(0.006)	(0.005)	(0.004)	(0.003)
P(A) = 0.7	0.759	0.768	0.786	0.805	0.826	0.842	0.853
P(B) = 0.4	(0.100)	(0.091)	(0.074)	(0.057)	(0.039)	(0.024)	(0.014)
P(A) = 0.9	0.862	0.899	0.943	0.979	0.995	0.999	1.000
P(B) = 0.4	(0.152)	(0.123)	(0.083)	(0.039)	(0.009)	(0.002)	(0.000)
P(A) = 0.8	0.482	0.497	0.518	0.543	0.563	0.577	0.584
P(B) = 0.6	(0.105)	(0.096)	(0.078)	(0.056)	(0.036)	(0.021)	(0.012)
P(A) = 0.9	0.706	0.744	0.798	0.858	0.901	0.925	0.937
P(B) = 0.6	(0.194)	(0.172)	(0.136)	(0.087)	(0.047)	(0.023)	(0.011)
P(A) = 0.9	0.208	0.226	0.248	0.268	0.280	0.286	0.287
P(B) = 0.8	(0.072)	(0.065)	(0.054)	(0.040)	(0.028)	(0.020)	(0.017)

Table 10: Power of the Test for RPW(2n+1, 3, 3; 3) (Standard deviations are given in parentheses)

3.4.3 Monotonicity in α

Increasing the number of balls added after a successful response has the effect of speeding up the trial and producing more extreme results. In all the simulations, all other variables are kept constant at three, but similar results are expected at other fixed values. Steadily raising α quickly skews the trial towards one treatment when the treatment probabilities are small. The reason behind this is that after only one successful treatment, the number of balls added is large compared to the urn size. Therefore, when the next balls are drawn to determine the next treatment, the treatment probabilities have changed dramatically, especially for large values of n. For example, if treatment A had a

successful treatment, then for the next treatment selection, the number of type A balls would greatly outweigh the number of type B balls, thus giving treatment A a much higher probability of selection. Further successes on treatment A would only improve its chance of selection, and further skew results.

Tables 11 and 12 reflect this analysis. It should be pointed out that these tables give the more extreme results on average when compared with other monotonicity tables measuring the same statistic. The values for the expected proportion are higher and the values for power are lower. This is most likely owing to how rapidly the trial skews as a result of increasing the number of balls added after a successful response.

Treatment probability	RPW(3, 2n+1, 3; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2 P(B) = 0.1	0.536 (0.041)	0.549 (0.051)	0.574 (0.073)	0.618 (0.114)	0.678 (0.179)	0.713 (0.266)	0.709 (0.350)
P(A) = 0.5 P(B) = 0.1	0.674 (0.058)	0.732 (0.069)	0.809 (0.078)	0.885 (0.082)	0.924 (0.102)	0.924 (0.157)	0.901 (0.227)
P(A) = 0.9 P(B) = 0.1	0.941 (0.036)	0.964 (0.028)	0.977 (0.026)	0.984 (0.028)	0.983 (0.040)	0.978 (0.070)	0.966 (0.122)
P(A) = 0.5 P(B) = 0.4	0.559 (0.090)	0.581 (0.132)	0.605 (0.208)	0.614 (0.317)	0.596 (0.413)	0.569 (0.465)	0.564 (0.482)
P(A) = 0.7 P(B) = 0.4	0.718 (0.104)	0.768 (0.134)	0.809 (0.191)	0.797 (0.286)	0.744 (0.381)	0.693 (0.438)	0.674 (0.458)
P(A) = 0.9 P(B) = 0.4	0.901 (0.079)	0.923 (0.095)	0.923 (0.140)	0.889 (0.228)	0.829 (0.329)	0.771 (0.399)	0.745 (0.427)
P(A) = 0.8 P(B) = 0.6	0.712 (0.166)	0.731 (0.233)	0.713 (0.332)	0.662 (0.420)	0.620 (0.466)	0.601 (0.481)	0.594 (0.487)
P(A) = 0.9 P(B) = 0.6	0.830 (0.155)	0.830 (0.216)	0.787 (0.312)	0.720 (0.404)	0.666 (0.455)	0.643 (0.472)	0.635 (0.478)
P(A) = 0.9 P(B) = 0.8	0.650 (0.298)	0.627 (0.375)	0.590 (0.442)	0.565 (0.478)	0.554 (0.492)	0.549 (0.495)	0.546 (0.496)

Table 11: Expected Proportion on the Superior Treatment for RPW(3, 2n+1, 3; 3)

(Standard deviations are given in parentheses)

Treatment probability	RPW(3, 2n+1, 3; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2	0.275	0.270	0.258	0.234	0.190	0.143	0.107
P(B) = 0.1	(0.013)	(0.017)	(0.028)	(0.050)	(0.079)	(0.098)	(0.097)
P(A) = 0.5	0.990	0.977	0.904	0.666	0.401	0.301	0.273
P(B) = 0.1	(0.008)	(0.035)	(0.123)	(0.270)	(0.321)	(0.338)	(0.339)
P(A) = 0.9	0.980	0.941	0.892	0.855	0.837	0.831	0.829
P(B) = 0.1	(0.060)	(0.098)	(0.118)	(0.121)	(0.119)	(0.118)	(0.118)
P(A) = 0.5	0.162	0.155	0.140	0.110	0.077	0.056	0.047
P(B) = 0.4	(0.008)	(0.015)	(0.027)	(0.038)	(0.036)	(0.026)	(0.018)
P(A) = 0.7	0.759	0.672	0.533	0.406	0.299	0.199	0.146
P(B) = 0.4	(0.100)	(0.166)	(0.235)	(0.264)	(0.235)	(0.179)	(0.120)
P(A) = 0.9	0.862	0.766	0.705	0.685	0.629	0.488	0.414
P(B) = 0.4	(0.152)	(0.198)	(0.225)	(0.250)	(0.267)	(0.262)	(0.208)
P(A) = 0.8	0.482	0.407	0.303	0.193	0.125	0.104	0.093
P(B) = 0.6	(0.105)	(0.141)	(0.157)	(0.140)	(0.111)	(0.092)	(0.059)
P(A) = 0.9	0.706	0.628	0.551	0.390	0.248	0.220	0.215
P(B) = 0.6	(0.194)	(0.234)	(0.257)	(0.261)	(0.227)	(0.186)	(0.137)
P(A) = 0.9	0.208	0.160	0.112	0.082	0.069	0.072	0.074
P(B) = 0.8	(0.072)	(0.082)	(0.082)	(0.074)	(0.062)	(0.057)	(0.053)

Table 12: Power of the Test for RPW(3, 2n+1, 3; 3) (Standard deviations are given in parentheses)

It is important to note that both the increasing trends seen in Table 11 as well as the decreasing trends seen in Table 12 are only valid for small values of p_A, p_B and n. It may seem counter-intuitive, but when p_A, p_B and n all have extremely high values there is actually a decreasing trend seen in the expected proportion and an increasing trend in the power. This exception is in all likelihood due to the high probability of choosing treatment B. If the trial consisted of a series of successes on treatment B, (which is very likely because p_B is high) then the trial could become skewed toward the inferior treatment, even if treatment A has a higher probability of selection. If, in this case, there was a large number of successful treatments on treatment B early in the trial, then the

number of B balls would outweigh the number of A balls and treatment B would most likely be selected more often than treatment A. This obviously reduces the number of patients allocated to the superior treatment. In conclusion, the number of balls added after a successful response should not be set too high, otherwise the trial will quickly be skewed and produce extreme values for the expected proportion on the superior treatment and for the power of the test.

3.4.4 Monotonicity in β

Next we explore the monotonic properties of the number of balls added after a failed response, denoted as β. Increasing the number of balls added after a failed response has the opposite effect of increasing the number of balls added after a successful response. Instead of an increasing trend in the expected proportion on the better treatment, we see a decreasing trend as n increases. In addition, a non-decreasing trend can be seen in power as n increases. One reasoning for these differing trends is that adding more balls of the opposite treatment after a failed response will actually decrease the chances of selecting that failed treatment again.

In essence, even if the superior treatment has a larger selection probability, the occurrence of a failure of that treatment early in the trial will decrease its selection probability substantially, especially for large values of n. Given a large value of n, it is quite easy to skew results to the inferior treatment as the number of balls added after a failed response will outnumber the initial urn size. Thus, it is not only counter-productive, but also unethical to have large values for the number of balls added after a failed response because patients have a lower chance of receiving the superior treatment.

Tables 13 and 14 show the monotonic properties of β for the expected proportion on the superior treatment and the power of the test, respectively.

Treatment probability	RPW(3, 3, 2n+1; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2	0.536	0.529	0.524	0.521	0.519	0.519	0.518
P(B) = 0.1	(0.041)	(0.053)	(0.032)	(0.030)	(0.029)	(0.029)	(0.028)
P(A) = 0.5	0.678	0.643	0.618	0.604	0.596	0.592	0.589
P(B) = 0.1	(0.058)	(0.049)	(0.043)	(0.039)	(0.037)	(0.036)	(0.035)
P(A) = 0.9	0.941	0.917	0.881	0.847	0.822	0.807	0.800
P(B) = 0.1	(0.036)	(0.043)	(0.049)	(0.050)	(0.048)	(0.046)	(0.045)
P(A) = 0.5	0.559	0.547	0.537	0.533	0.530	0.528	0.528
P(B) = 0.4	(0.090)	(0.069)	(0.055)	(0.048)	(0.044)	(0.042)	(0.042)
P(A) = 0.7	0.718	0.679	0.646	0.626	0.614	0.609	0.606
P(B) = 0.4	(0.104)	(0.083)	(0.068)	(0.058)	(0.053)	(0.050)	(0.049)
P(A) = 0.9	0.901	0.878	0.841	0.804	0.778	0.763	0.755
P(B) = 0.4	(0.079)	(0.076)	(0.073)	(0.068)	(0.062)	(0.059)	(0.057)
P(A) = 0.8	0.712	0.686	0.653	0.629	0.616	0.609	0.605
P(B) = 0.6	(0.166)	(0.128)	(0.100)	(0.081)	(0.070)	(0.066)	(0.062)
P(A) = 0.9	0.830	0.818	0.785	0.751	0.726	0.712	0.704
P(B) = 0.6	(0.155)	(0.126)	(0.107)	(0.091)	(0.080)	(0.074)	(0.071)
P(A) = 0.9	0.650	0.658	0.653	0.637	0.622	0.614	0.607
P(B) = 0.8	(0.298)	(0.246)	(0.188)	(0.144)	(0.118)	(0.103)	(0.094)

Table 13: Expected Proportion on the Superior Treatment for RPW(3, 3, 2n+1; 3)

(Standard deviations are given in parentheses)

Treatment probability	RPW(3, 3, 2n+1; 3) design, where n is						
	1	2	4	8	16	32	64
P(A) = 0.2	0.041	0.277	0.279	0.280	0.280	0.280	0.281
P(B) = 0.1	(0.013)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
P(A) = 0.5	0.990	0.994	0.995	0.995	0.996	0.996	0.996
P(B) = 0.1	(0.008)	(0.003)	(0.002)	(0.001)	(0.008)	(0.001)	(0.001)
P(A) = 0.9	0.980	0.992	0.998	1.000	1.000	1.000	1.000
P(B) = 0.1	(0.060)	(0.038)	(0.018)	(0.007)	(0.006)	(0.002)	(0.000)
P(A) = 0.5	0.162	0.164	0.166	0.166	0.167	0.167	0.167
P(B) = 0.4	(0.008)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
P(A) = 0.7	0.759	0.803	0.828	0.839	0.845	0.847	0.848
P(B) = 0.4	(0.100)	(0.059)	(0.035)	(0.024)	(0.019)	(0.016)	(0.016)
P(A) = 0.9	0.862	0.915	0.961	0.984	0.992	0.995	0.996
P(B) = 0.4	(0.152)	(0.116)	(0.070)	(0.039)	(0.023)	(0.015)	(0.010)
P(A) = 0.8	0.482	0.523	0.552	0.567	0.574	0.577	0.579
P(B) = 0.6	(0.105)	(0.075)	(0.046)	(0.030)	(0.022)	(0.019)	(0.017)
P(A) = 0.9	0.706	0.755	0.820	0.864	0.888	0.898	0.903
P(B) = 0.6	(0.194)	(0.164)	(0.118)	(0.081)	(0.057)	(0.044)	(0.039)
P(A) = 0.9	0.208	0.236	0.263	0.279	0.287	0.290	0.292
P(B) = 0.8	(0.072)	(0.060)	(0.043)	(0.030)	(0.021)	(0.016)	(0.014)

Table 14: Power of the Test for RPW(3, 3, 2n+1; 3) (Standard deviations are given in parentheses)

3.4.5 Monotonicity in γ

When studying the monotonicity of γ, the number of balls drawn, one must consider that the number drawn cannot exceed the total initial urn size. Ideally, the number drawn should be equal to one half the number of balls in the urn at the start of the trial. This provides a nice compromise between the expected proportion on the superior treatment and the power of the test. Drawing too many balls will result in a low power and an inability to make a sound conclusion, while drawing too few will give rise to a lower probability that a patient may receive the superior treatment. It is important to determine a balance between increased proportion of patients on the better treatment and

a valid statistical conclusion. Depending on the situation, one might find it more appealing to focus on one property more than the other, and thus drawing more or fewer balls. Table 15 shows an increasing trend for the expected proportion concerning the number balls drawn while all other variables are kept constant. Likewise, Table 16 shows a decreasing trend for the power of the test regarding the number of balls drawn while all other variables are kept constant at five. These trends have been present throughout the entire monotonicity section to some degree. The main concept to grasp is that there is no perfect combination of variables for the perfect experiment, what is key is that once the monotonic properties are known, the variables can be manipulated to suit the experiments needs and goals.

Treatment probability	RPW(5, 5, 5; γ) where γ is				
	1	3	5	7	9
P(A) = 0.2	0.529	0.536	0.539	0.661	0.738
P(B) = 0.1	(0.041)	(0.041)	(0.041)	(0.042)	(0.040)
P(A) = 0.5	0.636	0.678	0.702	0.848	0.922
P(B) = 0.1	(0.051)	(0.058)	(0.058)	(0.052)	(0.041)
P(A) = 0.9	0.865	0.941	0.967	0.994	0.998
P(B) = 0.1	(0.050)	(0.036)	(0.027)	(0.009)	(0.005)
P(A) = 0.5	0.541	0.559	0.573	0.788	0.897
P(B) = 0.4	(0.070)	(0.090)	(0.098)	(0.086)	(0.062)
P(A) = 0.7	0.649	0.718	0.763	0.939	0.983
P(B) = 0.4	(0.082)	(0.104)	(0.114)	(0.053)	(0.023)
P(A) = 0.9	0.798	0.901	0.938	0.990	0.997
P(B) = 0.4	(0.086)	(0.079)	(0.070)	(0.017)	(0.007)
P(A) = 0.8	0.632	0.712	0.750	0.964	0.992
P(B) = 0.6	(0.118)	(0.166)	(0.205)	(0.060)	(0.020)
P(A) = 0.9	0.718	0.830	0.862	0.985	0.997
P(B) = 0.6	(0.124)	(0.155)	(0.180)	(0.038)	(0.011)
P(A) = 0.9	0.592	0.650	0.658	0.954	0.994
P(B) = 0.8	(0.177)	(0.298)	(0.366)	(0.145)	(0.033)

Table 15: Expected Proportion on the Superior Treatment for RPW(5,5,5; γ) (Standard deviations are given in parentheses)

Treatment probability	RPW(5, 5, 5; γ) where γ is				
	1	3	5	7	9
P(A) = 0.2	0.277	0.275	0.274	0.222	0.178
P(B) = 0.1	(0.012)	(0.013)	(0.013)	(0.022)	(0.026)
P(A) = 0.5	0.994	0.990	0.987	0.858	0.554
P(B) = 0.1	(0.003)	(0.008)	(0.015)	(0.149)	(0.259)
P(A) = 0.9	1.000	0.980	0.933	0.800	0.761
P(B) = 0.1	(0.009)	(0.060)	(0.102)	(0.105)	(0.074)
P(A) = 0.5	0.165	0.162	0.160	0.119	0.082
P(B) = 0.4	(0.005)	(0.008)	(0.010)	(0.025)	(0.025)
P(A) = 0.7	0.822	0.759	0.698	0.330	0.191
P(B) = 0.4	(0.044)	(0.100)	(0.142)	(0.167)	(0.100)
P(A) = 0.9	0.982	0.862	0.745	0.534	0.479
P(B) = 0.4	(0.042)	(0.152)	(0.197)	(0.158)	(0.115)
P(A) = 0.8	0.552	0.482	0.418	0.202	0.152
P(B) = 0.6	(0.051)	(0.105)	(0.138)	(0.104)	(0.062)
P(A) = 0.9	0.871	0.706	0.611	0.386	0.329
P(B) = 0.6	(0.088)	(0.194)	(0.230)	(0.157)	(0.099)
P(A) = 0.9	0.269	0.208	0.163	0.149	0.130
P(B) = 0.8	(0.039)	(0.072)	(0.080)	(0.056)	(0.036)

Table 16: Power of the Test for RPW(5, 5, 5; γ) (Standard deviations are given in parentheses)

3.5 Simulations with Real Data

Besides studying the properties of an experimental design, it is essential to try to simulate actual experiments and compare results obtained with those of real clinical trials. Only then will it be possible to say if the RPW(2n+1, 2n+1, 2n+1) design can improve on results obtained in actual clinical trials. Recall that the RPW designs are most appropriate for cases of life-threatening illness. That is why the two clinical trials we have chosen to compare with the RPW design are the ECMO trials and AZT trials previously mentioned in chapter two. In the following sections the RPW design will be

compared with real data and simulations based on the clinical trials in terms of the expected proportion on the superior treatment and the power of the test.

3.5.1 Comparison with U.K. ECMO Trial

The ECMO trial used for comparison in this section is actually the U.K. ECMO Study of 1993 to 1995. The study consisted of 185 infant patients who were randomized into either the conventional treatment or the ECMO treatment for treatment of PPHN. 93 patients were allocated to the ECMO treatment, of which 63 survived and 92 patients received the conventional treatment, of which 38 survived. This breaks down to a mortality rate of 32.3% on the ECMO treatment and 58.7% on the conventional treatment [25]. Differences between the two randomized groups were presented as relative risks and as absolute percentage differences or differences between means or medians. Statistical tests used to calculate a statistically significant difference include chi-square tests, Fisher's exact tests, t tests and median tests [25]. The study concluded with overwhelming evidence that the ECMO treatment was superior to the conventional one, with a p-value of 0.0005 [25].

Although the ECMO treatment had been proven without refute to be the better treatment, the randomized clinical trial did not come without a cost. The actual proportion of patients who received the superior treatment was 50.3%, which is as expected for a randomized clinical trial. A higher proportion of patients would most likely have received the ECMO treatment if an adaptive design had been used instead of a classical randomized design. A simulation study was done using 2000 replications, each with a sample size of 185 to compare the expected proportion of patients on ECMO with the actual proportion in the clinical trial. Table 17 gives the expected proportion on the

superior treatment for various values of p_A and p_B. In the table it is evident that even small treatment assignment probabilities still exhibit a higher expected proportion than 50.3%. In addition, the proportions tend to increase as n increases. Table 18 gives values of the power of the test for various values of p_A and p_B and increasing values of n. As n increases, the power of the test decreases, making it more difficult to reach a valid conclusion. Nonetheless, it is still possible to make a valid statistical conclusion using a RPW(2n+1, 2n+1, 2n+1) design and is quite worthwhile if more patients can be treated successfully with the superior treatment. Note especially when the treatment assignment values are close to those actually observed in the study ($p_A = 0.6$ and $p_B = 0.3$) that the expected proportion definitely improves on the study results and the power is only slightly diminished. Therefore, although the simulations do not provide indisputable proof that a RPW design will save more lives than a 50:50 randomized design, it can be concluded that the RPW design would have been a better choice in the U.K. ECMO study.

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1) design, where n is					
		0	1	2	4	8	16
P(A) = 0.2	0.499	0.530	0.536	0.541	0.545	0.551	0.554
P(B) = 0.1	(0.036)	(0.029)	(0.030)	(0.031)	(0.031)	(0.032)	(0.033)
P(A) = 0.5	0.499	0.640	0.682	0.705	0.735	0.770	0.803
P(B) = 0.1	(0.036)	(0.037)	(0.042)	(0.044)	(0.042)	(0.051)	(0.054)
P(A) = 0.9	0.499	0.881	0.954	0.976	0.990	0.994	0.995
P(B) = 0.1	(0.036)	(0.037)	(0.024)	(0.017)	(0.010)	(0.007)	(0.006)
P(A) = 0.6	0.499	0.631	0.685	0.716	0.761	0.814	0.865
P(B) = 0.3	(0.036)	(0.051)	(0.061)	(0.068)	(0.075)	(0.080)	(0.078)
P(A) = 0.5	0.499	0.545	0.564	0.575	0.591	0.615	0.642
P(B) = 0.4	(0.036)	(0.052)	(0.065)	0.074	(0.090)	(0.106)	(0.122)
P(A) = 0.7	0.499	0.657	0.732	0.778	0.839	0.902	0.940
P(B) = 0.4	(0.036)	(0.063)	(0.077)	(0.087)	(0.090)	(0.086)	(0.073)
P(A) = 0.9	0.499	0.821	0.926	0.958	0.980	0.986	0.987
P(B) = 0.4	(0.036)	(0.007)	(0.052)	(0.047)	(0.033)	(0.209)	(0.030)
P(A) = 0.8	0.499	0.645	0.742	0.785	0.840	0.852	0.850
P(B) = 0.6	(0.036)	(0.097)	(0.132)	(0.170)	(0.198)	(0.251)	(0.285)
P(A) = 0.9	0.499	0.744	0.870	0.900	0.928	0.917	0.908
P(B) = 0.6	(0.036)	(0.104)	(0.117)	(0.140)	(0.147)	(0.196)	(0.228)
P(A) = 0.9	0.499	0.608	0.682	0.666	0.663	0.598	0.601
P(B) = 0.8	(0.036)	(0.158)	(0.282)	(0.369)	(0.431)	(0.476)	(0.484)

Table 17: Expected Proportion on the Superior Treatment, Comparison with U.K. ECMO Study (Standard deviations are given in parentheses)

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1) design, where n is					
		0	1	2	4	8	16
P(A) = 0.2	0.476	0.468	0.466	0.465	0.463	0.461	0.459
P(B) = 0.1	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)	(0.013)	(0.013)
P(A) = 0.5	1.000	1.000	1.000	1.000	1.000	0.999	0.996
P(B) = 0.1	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.005)	(0.017)
P(A) = 0.9	1.000	1.000	0.995	0.963	0.881	0.824	0.805
P(B) = 0.1	(0.000)	(0.000)	(0.030)	(0.079)	(0.116)	(0.109)	(0.101)
P(A) = 0.6	0.987	0.980	0.969	0.957	0.925	0.849	0.728
P(B) = 0.3	(0.001)	(0.066)	(0.018)	(0.031)	(0.065)	(0.138)	(0.212)
P(A) = 0.5	0.274	0.271	0.267	0.264	0.259	0.251	0.241
P(B) = 0.4	0.002	0.006	0.010	0.013	0.018	0.027	0.035
P(A) = 0.7	0.987	0.975	0.947	0.904	0.801	0.605	0.448
P(B) = 0.4	(0.001)	(0.012)	(0.043)	(0.087)	(0.164)	(0.247)	(0.265)
P(A) = 0.9	1.000	0.999	0.932	0.804	0.660	0.612	0.598
P(B) = 0.4	(0.000)	0.006	(0.104)	(0.182)	(0.209)	(0.206)	(0.209)
P(A) = 0.8	0.847	0.806	0.710	0.611	0.464	0.368	0.321
P(B) = 0.6	(0.003)	(0.049)	(0.126)	(0.189)	(0.239)	(0.250)	(0.246)
P(A) = 0.9	0.998	0.973	0.812	0.674	0.564	0.550	0.523
P(B) = 0.6	(0.001)	(0.038)	(0.179)	(0.242)	(0.270)	(0.279)	(0.279)
P(A) = 0.9	0.475	0.444	0.326	0.230	0.146	0.099	0.092
P(B) = 0.8	(0.009)	(0.048)	(0.115)	(0.127)	(0.113)	(0.090)	(0.073)

Table 18: Power of the Test, Comparison with U.K. ECMO Study (Standard deviations are given in parentheses)

3.5.2 Comparison with AZT Trial

A study of the drug AZT was carried out from 1991 to 1994 to study the effectiveness of the drug in prevention of HIV transmission from mother to infant. A stratified (with respect to institute) permuted block design [5] was used to allocate 476 women to treatment using either of AZT or a placebo. At the end of the study, the success rates for AZT and the placebo were 92.8% and 75.2%, respectively [33]. The study had a very strong and statistically valid conclusion with a p-value of 2×10^{-7} [33]. The test statistic for testing the equality of AZT and the placebo was based on the difference between their corresponding Kaplan-Meier estimates [33]. Of the 409 women who gave birth to live babies, 53 infants had contracted HIV, 13 out of the 205 who were treated with AZT and 40 out of the 204 who had been administered a placebo. The actual proportion of patients who received the superior treatment was 50.1%, which is concurrent with a randomized trial.

Although the study reached a very powerful and irrefutable conclusion, it may have been possible to obtain nearly the same results and save more infants from contracting HIV if an adaptive clinical trial had been used. A simulation study was done to compare the RPW(2n+1, 2n+1, 2n+1) design with the AZT study in terms of the expected proportion on the superior treatment and the power of the test. 2000 replications were used, each with a sample size of 476.

Tables 19 and 20 provide the results for different combinations of p_A and p_B. Note that in Table 19 there is an overall increasing trend as n increases and all of the expected proportions are greater than 50.1%. Table 20 displays a non-increasing trend as n increases. One comparison to note in particular is when the treatment allocation probabilities are similar to those actually observed in the AZT study. When $p_A = 0.95$

and $p_B = 0.8$, we see that the expected proportion on the superior treatment is quite higher than fifty per cent and the power of the test is still acceptable for drawing a valid conclusion as long as n is small. It is important to point out that for smaller values of n the power of the test is the same as that of 50-50 randomization. The conclusion is that a RPW(2n+1, 2n+1, 2n+1) design can allocate more patients to a superior treatment without substantial loss of power and could have saved more infants from infection of HIV if used in this study.

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1) design, where n is					
		0	1	2	4	8	16
P(A) = 0.2 P(B) = 0.1	0.500 (0.023)	0.530 (0.018)	0.536 (0.019)	0.541 (0.018)	0.545 (0.019)	0.550 (0.021)	0.555 (0.020)
P(A) = 0.5 P(B) = 0.1	0.500 (0.023)	0.641 (0.024)	0.683 (0.026)	0.708 (0.027)	0.738 (0.030)	0.774 (0.032)	0.808 (0.035)
P(A) = 0.9 P(B) = 0.1	0.500 (0.023)	0.890 (0.023)	0.963 (0.014)	0.985 (0.009)	0.995 (0.004)	0.998 (0.003)	0.998 (0.002)
P(A) = 0.5 P(B) = 0.4	0.500 (0.023)	0.545 (0.034)	0.563 (0.043)	0.578 (0.048)	0.593 (0.057)	0.619 (0.070)	0.648 (0.083)
P(A) = 0.7 P(B) = 0.4	0.500 (0.023)	0.661 (0.041)	0.741 (0.051)	0.793 (0.054)	0.860 (0.058)	0.927 (0.048)	0.968 (0.036)
P(A) = 0.9 P(B) = 0.4	0.500 (0.023)	0.835 (0.043)	0.945 (0.030)	0.976 (0.022)	0.990 (0.016)	0.994 (0.012)	0.995 (0.011)
P(A) = 0.8 P(B) = 0.6	0.500 (0.023)	0.651 (0.065)	0.762 (0.100)	0.834 (0.117)	0.884 (0.151)	0.912 (0.176)	0.886 (0.254)
P(A) = 0.9 P(B) = 0.6	0.500 (0.023)	0.760 (0.072)	0.901 (0.081)	0.946 (0.083)	0.957 (0.101)	0.956 (0.127)	0.936 (0.193)
P(A) = 0.9 P(B) = 0.8	0.500 (0.023)	0.618 (0.124)	0.713 (0.264)	0.714 (0.356)	0.657 (0.443)	0.618 (0.479)	0.591 (0.489)
P(A) = 0.95 P(B) = 0.8	0.500 (0.023)	0.692 (0.140)	0.805 (0.253)	0.780 (0.338)	0.709 (0.428)	0.664 (0.466)	0.635 (0.480)

Table 19: Expected Proportion on the Superior Treatment, Comparison with AZT Study (Standard deviations are given in parentheses)

Treatment probability	50-50 random	RPW(2n+1, 2n+1, 2n+1) design, where n is					
		0	1	2	4	8	16
P(A) = 0.2	0.865	0.862	0.862	0.861	0.860	0.859	0.858
P(B) = 0.1	(0.001)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
P(A) = 0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
P(B) = 0.1	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
P(A) = 0.9	1.000	1.000	1.000	0.990	0.904	0.836	0.813
P(B) = 0.1	(0.000)	(0.000)	(0.001)	(0.043)	(0.110)	(0.108)	(0.099)
P(A) = 0.5	0.592	0.586	0.582	0.577	0.570	0.557	0.538
P(B) = 0.4	(0.001)	(0.006)	(0.011)	(0.015)	(0.020)	(0.033)	(0.049)
P(A) = 0.7	1.000	1.000	1.000	0.999	0.982	0.850	0.548
P(B) = 0.4	(0.000)	(0.000)	(0.000)	(0.003)	(0.035)	(0.161)	(0.268)
P(A) = 0.9	1.000	1.000	0.994	0.898	0.692	0.622	0.611
P(B) = 0.4	(0.000)	(0.000)	(0.022)	(0.135)	(0.211)	(0.208)	(0.213)
P(A) = 0.8	0.998	0.994	0.962	0.871	0.641	0.441	0.416
P(B) = 0.6	(0.000)	(0.005)	(0.049)	(0.126)	(0.259)	(0.319)	(0.330)
P(A) = 0.9	1.000	1.000	0.949	0.762	0.585	0.569	0.564
P(B) = 0.6	(0.000)	(0.004)	(0.080)	(0.205)	(0.271)	(0.288)	(0.291)
P(A) = 0.9	0.865	0.826	0.609	0.400	0.201	0.097	0.093
P(B) = 0.8	(0.002)	(0.054)	(0.203)	(0.235)	(0.182)	(0.108)	(0.085)
P(A) = 0.95	0.999	0.980	0.733	0.585	0.371	0.186	0.207
P(B) = 0.8	(0.000)	(0.037)	(0.248)	(0.287)	(0.275)	(0.203)	(0.176)

Table 20: Power of the Test, Comparison with AZT Study (Standard deviations are given in parentheses)

3.6 Recommendations

Although the randomized play the winner rule design is not appropriate for all clinical trials, when it is appropriate it is definitely an ethical experimental design to choose. The comparison study in the previous section demonstrates that the RPW design outperforms other designs in terms of allocating more patients to the superior treatment. When patients are afflicted with life-threatening illness, this can amount to saving more lives. What is especially appealing about the design is that both the total number of successes and the expected proportion on the superior treatment are higher than if an

alternate design had been used, but the power of the test is not necessarily compromised, hence a statistically valid conclusion can still be obtained. Furthermore, due to the compromise between information gathering and payoff to current patients, the ethical concerns often found in randomized trials are eliminated when an adaptive clinical trial is used.

There has been much discussion as to the monotonic properties of the RPW design. The parameters can be raised or lowered, depending on the focus on the trial, whether it is maximizing the power of the test or the number of patients on the superior treatment. Certainly, taking a large sample size will always help to increase the power of the test. To increase the expected proportion on the superior treatment, increasing all parameters simultaneously, such as a RPW(3, 3, 3) or a RPW(5, 5, 5) design will provide better results than a RPW(1, 1, 1) design. One cautionary note is that if the treatment allocation probabilities are expected to be high, especially the probability on the inferior treatment, then all the parameters should be kept small to prevent skewing results to favour the inferior treatment.

The main message to take away from this in depth study of the RPW(μ, α, β; γ) design is that there are other options to use in lieu of randomization in a clinical trial, and when those options are exercised, the results are not compromised. Equal randomization is thought of as the gold standard [3] of research in the medical field. It may be so in terms of obtaining the most powerful test. Where the patient's chances of receiving the best treatment are concerned, randomized trials are deficient. It has been shown in the previous simulation studies that adaptive clinical trials consistently outperform randomized clinical trials in terms of allocating a higher percentage of patients to the superior treatment. In situations where the disease or affliction under study is severely

life-threatening, adaptive clinical trials are the only ethical option. The RPW(2n+1, 2n+1, 2n+1; 2n+1) design is perfectly suited for studies of severe illness because it is highly customizable. The parameters can be modified to focus more on goal of information gathering or on immediate payoff to the patient. The RPW design is highly valuable to both the medical community and to patients as an alternative to randomization and should not be overlooked.

Chapter 4

Conclusions

4.1 Summary

Wei and Durham's Randomized Play the Winner Rule is an adaptive design well suited to medical trials studying the treatment of severe illness. Compared to other designs, the RPW design determines during the study which treatment is currently identified as superior and then allocates a greater proportion of patients to that treatment. Moreover, the RPW design is still able to reach a statistically valid conclusion as the power of the test is comparable to that of randomized clinical trials. The main attraction of the design is that it makes a compromise between individual and collective ethics, resulting in a better choice for clinicians wishing to conduct medical trials on very ill patients.

Adaptive designs are indeed a better choice than randomized clinical trials in cases of testing treatments on life-threatening illness. In chapter two, the advantages and disadvantages of randomized clinical trials over adaptive clinical trials were discussed. Randomized designs are well known, highly regarded designs that reduce many types of bias and have solid methods for inference. However, adaptive designs are better suited to trials of ethical question since they focus more on the patient's concerns of randomization than on the validity of the trial. In particular, two cases that would have benefited from

the use of adaptive clinical trials were the UK ECMO trial of 1993-1995 and the AZT trial of 1991-1994. Randomizing patients in these cases was unethical. Using an adaptive design might have saved more lives and still resulted in statistically valid conclusions.

Simulation studies were used in chapter three to demonstrate the superiority of the RPW design over 50-50 randomization and Neyman allocation. The total number of successes, the expected proportion on the superior treatment, and the odds ratio were all found to have higher values overall compared to other designs. In addition, the power of the test was found to be quite comparable to other designs, indicating that a valid conclusion could be obtained. Later in the chapter the monotonic properties of RPW(μ, α, β; γ) were explored. It was found that the higher the values the parameters, the greater the expected proportion on the superior treatment and the lower the power of the test. There were some significant exceptions, particularly when the allocation probability of the inferior treatment was high.

In general, adaptive designs are ethically appropriate but underutilized in medical clinical trials. Incorporating designs such as the randomized play the winner design would benefit patients of severe illness without sacrificing the goal of the medical trial: to further medical knowledge and save future patients' lives.

4.2 Future Research

Although much research has been done to as to the design of the randomized play the winner rule, more research is needed as to the inference of the design. Solid, statistical tests that are quickly and easily computable are lacking for adaptive designs. If

these designs are ever to be adopted fully by the medical community, reliable statistical methods for inference must be developed.

Another area of possible future research is to expand the design of the randomized play the winner rule to include more than two treatments. Simulation studies could be done to compare this design to the two treatment randomized play the winner design as well as other experimental designs. Also, a monotonicity study of the multiple treatment adaptive design would be insightful as to not only how the design operates, but also which values to choose for parameters of the design to achieve desired results.

References

1. Bartlett, R. H., *et al.*, Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study, *Pediatrics*, 1985; 76: 479-487.
2. Bartlett, R. H., Andrews, A. F., Toomasian, J. M., *et al.*, Extracorporeal Membrane Oxygenation for Newborn Respiratory Failure: 45 Cases, *Surgery*, 1982; 92:425-433.
3. Berry, Donald A., Comment: Ethics and ECMO, *Statistical Science*, 1985; 4:306-310.
4. Begg, Colin B., On Inferences from Wei's Biased Coin Design for Clinical Trials, *Biometrika*, 1990; 77:467-484.
5. Connor, E. M., ,Sperling, R. S., Gelber, R., *et al.*, and the Pediatric AIDS Clinical Trials Group Protocol 076 Study Group, Reduction of Maternal-Infant Transmission of Human Immunodeficiency Virus Type I with Zidovudine Treatment, *New England Journal of Medicine*, 1994; 331:1173-1180.
6. Day, N. E., Two Stage Designs for Clinical Trials, *Biometrics*, 1969; 25:111-118.
7. Freedman, B., Equipoise and the Ethics of Clinical Research, *New England Journal of Medicine*, 1987; 317:141-145.
8. Fried, C., *Medical Experimentation: Personal Integrity and Social Policy*, New York, NY: Elsevier Science Inc., 1974.
9. Gross, S. J., Bifano, E. M., D'Eugenio, D. B., Hakanson, D. O. and Hingre, R. V. Prospective Randomized Controlled Trial of Conventional treatment or Transport

for ECMO in Infants with Severe Persistent Pulmonary Hypertension (PPHN): Two Year Follow Up (Abstract), *Pediatrics Res.*, 1994; 17A: 86.

10. Jain, Raj, *The Art of Computer Systems Performance Analysis,* New York, NY: John Wiley and Sons, Inc., 1991.
11. Kirkpatrick, B. V., *et al.*, Use of Extracorporeal Membrane Oxygenation for Respiratory Failure in Term Infants, *Pediatrics*, 1983; 72:872-876.
12. Moore, David S., *The Basic Practice of Statistics* (4th Edition), New York, NY: W. H. Freeman and Company, 1995.
13. O'Rourke, P. P., *et al.*, Extracorporeal Membrane Oxygenation and Conventional Hypertension of the Newborn: A Prospective Randomized Study, *Pediatrics*, 1989; 84:957-963.
14. Palmer, Christopher R. and Rosenberger, William F., Ethics and Practice: Alternative Designs for Phase III Randomized Clinical Trials, *Controlled Clinical Trials*, 1999; 20:172-186.
15. Petrucci, Ralph H., and Harwood, William, S., *General Chemistry: Applications Principles and Modern Applications* (6th Edition), Englewood Cliffs, NJ: Prentice-Hall Inc., 1993.
16. Pullman, Daryl and Wang, Xikui, Adaptive Designs, Informed Consent, and the Ethics of Research, *Controlled Clinical Trials*, 2001; 22:203-210.
17. Rosenberger, William F., Randomized Play the Winner Clinical Trials: Review and Recommendations, *Controlled Clinical Trials*, 1999; 20:328-342.
18. Rosenberger, William F. *et al.*, Optimal Adaptive Designs for Binary Response Trials, *Biometrics*, 2001; 57:909-913.

19. Rosenberger, William F. and Lachin, John M., *Randomization in Clinical Trials*, New York, NY: John Wiley and Sons, Inc., 2002.

20. Rosenberger, William F. and Lachin, John M., The Use of Response-Adaptive Designs in Clinical Trials, *Controlled Clinical Trials*, 1993; 14:471-484.

21. Rosner, Bernard, *Fundamentals of Biostatistics* (5th Edition), New York, NY: Duxbury Press, 2000.

22. Royall, Richard M., Ethics and Statistics in Randomized Clinical Trials, *Statistical Science*, 1991; 6:1 52-88.

23. Simon, Richard, Adaptive Treatment Assignment Methods and Clinical Trials, *Biometrics*, 1977; 33:743-749.

24. Troug, Robert D., Informed Consent and Research Design in Critical Care Medicine, *Critical Care*, 1999; 3:3 29-33.

25. UK Collaborative ECMO Trial Group, UK Collaborative Randomized Trial of Neonatal Extracorporeal Membrane Oxygenation, *Lancet*, 1996; 348:75-82.

26. Wang, Xikui and Pullman, Daryl, Play the Winner Rule and Adaptive Designs of Clinical Trials, *International Journal of Mathematics & Mathematical Science*, 2001; 27:4 229-236.

27. Ware, James H., Investigating Therapies of Potentially Great Benefit: ECMO, *Statistical Science*, 1989; 4:4 298-340.

28. Wei, L. J., Exact Two-Sample Permutation Tests Bases on the Randomized Play the Winner Rule, *Biometrika*, 1988; 75:603-606.

29. Wei, L. J. *et al.* Statistical Inference with Data-Dependent Treatment Allocation Rules, *Journal of the American Statistical Association*, 1990; 85:156-162.

30. Wei, L. J. and Durham, S., The Randomized Play the Winner Rule in Medical Trials, *Journal of the American Statistical Association*, 1978; 73:840-843.

31. Wetmore, N., McEwen, D., O'Connor, M., and Bartlett, R. H., Defining Indications for Artificial Organ Support in Respiratory Failure, *Transactions of the American Society of Artificial Internal Organs*, 1979; 25:459-461.

32. Veatch, R. M., *A Theory of Medical Ethics*, New York, NY: Basic Books, 1981.

33. Yao, Q. and Wei, L. J., Play the Winner for Phase II/III Clinical Trials, *Statistics in Medicine*, 1996; 15:2413-2423.

34. Zelen, M., Play the Winner Rule and the Controlled Clinical Trial, *Journal of the American Statistical Association*, 1969; 64:131-146.

Appendix: SAS Programs

1. Chi-square Test

```
options linesize=80 pagesize=50 nodate;

/*This program is for the Chi-square Test for Randomness.  The purpose
of this program is to determine that the random numbers generated are
in fact random.  If the Chi-square statistic is not significant, then
there will be insufficient evidence to reject the null hypothesis, Ho:
The numbers are random.*/

data generate;
  seed=12773;
  array x{1000};
  array y{50};
  chisquare = 0;  /*set chisquare value to zero*/
  chi = 0;  /*set chisquare in each subset to zero*/

do i = 1 to 100;  /*repeat simulation 100 times*/
        do k = 1 to 50;
                y{k} = 0;  /*set all counts in subsets to zero*/
        end;
        do j = 1 to 1000;  /*generate 1000 random numbers*/
                link randgen;
                x{j}=rand;  /*create an array of random numbers*/
        do k = 1 to 50;
                        if x{j} > (k-1)/50 & x{j} <= k/50
                        then y{k} = y{k} + 1;  /*count of random # in
each subset*/
                end;
        end;
        do k = 1 to 50;
                chi = chi + (y{k} - 1000/50)**2/(1000/50);
        end;
        chisquare = chisquare + chi;  /*cummulative sum over all
subsets*/
        chi = 0;  /*reset each subset to zero*/
end;

chisquare = chisquare / 100; /*take average of chisquare variables*/
output;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;  /*set up random number generator*/
return;

proc print;
 var chisquare;
 title "Chi-square Test for Randomness";
run;
```

2. Serial Correlation Test

```
options linesize=80 pagesize=50 nodate;

/*This program is for the Serial Correlation Test.  The purpose of this
program is to determine that the random numbers generated are
independent from one another.  If the 100(1-a)% confidence interval for
the autocovariance does contain zero, then there will be insufficient
evidence to reject the null hypothesis, Ho:  There is no correlation.*/

data generate;
  seed=12773;
  array x{100};
  Lag = 0;
/*n = 100;  number of random numbers
  k = 10;  lags 1 to 10*/

do j = 1 to 100;
      link randgen;
      x{j}=rand;
end;

do k = 1 to 10;
      Lag = k;
      R = 0;  /*autocovariance statistic*/
    do j = 1 to 100-k;
            R = R + (x{j}-0.5)*(x{j+k}-0.5)/(100-k);
    end;
      UCL = R + 1.96/(12*sqrt(100-k));
    LCL = R - 1.96/(12*sqrt(100-k));
    output;
end;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;  /*set up random number generator*/
return;

proc print;
 var Lag R LCL UCL;
 title "Serial Correlation Test";
run;
```

3. 50-50 Randomization

```
options linesize=80 pagesize=50 nodate;
/*Program for 50:50 Randomization*/

data sim;
  seed=12773;
  patsN = 100;  /*total # of patients treated*/
  array RedS{5000};  /*total # of successes using trmt 1*/
  array WhtS{5000};  /*total # of successes using trmt 2*/
  array TotalS{5000};  /*total # of successes*/
  array RedPat{5000};  /*total # of patients on trmt 1*/
  array WhtPat{5000};  /*total # of patients on trmt 2*/
  array TotalPat{5000};  /*total # patients*/
  array RedFail{5000};
  array WhtFail{5000};

  dimS = dim(totalS);  /*# of simulations*/
  Z = 1.96;  /*critical value corresponding to given alpha*/
  p1 = 0.8;  /* prob of success of red treatment*/
  p2 = 0.3;  /* prob of success of white treatment*/
  q1 = 1 - p1;
  q2 = 1 - p2; /*variables for power calculation*/


do i = 1 to dimS;  /*perform simulation 5000 times*/
  RedS{i} = 0;
  WhtS{i} = 0;
  TotalS{i} = 0;
  RedPat{i} = 0;
  WhtPat{i} = 0;
  TotalPat{i} = 0;  /*set arrays to zero for each simulation*/

  do j = 1 to patsN;
      link randgen;
    if rand > 0.5
    then link RedTrmt; /*patient receives Red Trmt*/
    else link WhtTrmt; /*patient receives White Trmt*/
  end;

  TotalS{i} = RedS{i} + WhtS{i};
  TotalPat{i} = RedPat{i} + WhtPat{i};
  Red = RedS{i}/RedPat{i};
  White = WhtS{i}/WhtPat{i};
  Total = TotalS{i}/TotalPat{i};
  n1 = RedPat{i};
  n2 = WhtPat{i};
  pbar = (n1*p1 +n2*p2)/(n1 + n2);
  qbar = 1 - pbar;
  Denom = ((p1*q1/n1)+(p2*q2/n2))**0.5;
  Term1 = (p1 - p2)/Denom;
  Term2 = ((pbar*qbar*((1/n1)+(1/n2)))**0.5)/Denom;
  Term3 = Term1 - Z*Term2;
  Power = probnorm(Term3);
  RedFail{i} = RedPat{i} - RedS{i};
  WhtFail{i} = WhtPat{i} - WhtS{i};
```

```
  OddsRatio = (RedS{i}/RedFail{i})/(WhtS{i}/WhtFail{i});
  ExpProp = n1/(n1+n2);
  TestStat = (Red-White)/sqrt(Total*(1-Total)*((1/n1)+(1/n2)));
  Pvalu = 2*(1-probnorm(TestStat));
  output;
end;

go to done;

RedTrmt:  /*trmt 1*/
  RedPat{i} = RedPat{i} + 1;
  link randgen;
  if rand < p1
  then RedS{i} = RedS{i} + 1;
return;

WhtTrmt:  /*trmt 2*/
  WhtPat{i} = WhtPat{i} + 1;
  link randgen;
  if rand < p2
  then WhtS{i} = WhtS{i} + 1;
return;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;
return;

done:
  keep Red White Total Power OddsRatio ExpProp Pvalu;
run;

proc means;
  var Red White Total Power OddsRatio ExpProp Pvalu;
  title "50:50 Randomization";
run;
```

4. RPW(1, 1, 1)

```
options linesize=80 pagesize=50 nodate;
/*Program for Classic Urn Model, RPW(1, 1, 1)*/

data sim;
  seed=12773;
  AddBall = 1;  /*number of balls to add*/
  patsN = 100;  /*total # of patients treated*/
  NBall = 1;  /*# balls of each trmt at start*/
  array p{2};  /*array prob of success of all trmts*/
  p{1} = 0.8;  /*probablity of success for trmt 1; red ball*/
  p{2} = 0.3;  /*probablity of success for trmt 1; white ball*/
  array RedS{5000};  /*total # of successes using trmt 1*/
  array WhtS{5000};  /*total # of successes using trmt 2*/
  array TotalS{5000};  /*total # of successes*/
  array RedPat{5000};  /*total # of patients on trmt 1*/
  array WhtPat{5000};  /*total # of patients on trmt 2*/
  array TotalPat{5000};  /*total # patients*/
  array RedFail{5000};
  array WhtFail{5000};

  dimS = dim(totalS);  /*# of simulations*/
  Z = 1.96;  /*critical value corresponding to given alpha*/
  p1 = p{1};  /* prob of success of red treatment*/
  p2 = p{2};  /* prob of success of white treatment*/
  q1 = 1 - p1;
  q2 = 1 - p2; /*variables for power calculation*/

do i = 1 to dimS;  /*perform simulation 5000 times*/
  RedS{i} = 0;
  WhtS{i} = 0;
  TotalS{i} = 0;
  RedPat{i} = 0;
  WhtPat{i} = 0;
  TotalPat{i} = 0;  /*set arrays to zero for each simulation*/
  RedBall = NBall;
  WhtBall = NBall;

  do j = 1 to patsN;
      TotalBall = RedBall + WhtBall;
    ProbRed = RedBall / TotalBall;
      link DrawBall;
      if BallOne = 1 then link RedTrmt;  /*if draw red, then trmt 1*/
    else link WhtTrmt;  /*if draw white, then trmt 2*/
  end;

  TotalS{i} = RedS{i} + WhtS{i};
  TotalPat{i} = RedPat{i} + WhtPat{i};
  Red = RedS{i}/RedPat{i};
  White = WhtS{i}/WhtPat{i};
  Total = TotalS{i}/TotalPat{i};
  n1 = RedPat{i};
  n2 = WhtPat{i};
  pbar = (n1*p1 +n2*p2)/(n1 + n2);
  qbar = 1 - pbar;
```

```
  Denom = ((p1*q1/n1)+(p2*q2/n2))**0.5;
  Term1 = (p1 - p2)/Denom;
  Term2 = ((pbar*qbar*((1/n1)+(1/n2)))**0.5)/Denom;
  Term3 = Term1 - Z*Term2;
  Power = probnorm(Term3);
  RedFail{i} = RedPat{i} - RedS{i};
  WhtFail{i} = WhtPat{i} - WhtS{i};
  OddsRatio = (RedS{i}/RedFail{i})/(WhtS{i}/WhtFail{i});
  ExpProp = n1/(n1+n2);
  TestStat = (Red-White)/sqrt(Total*(1-Total)*((1/n1)+(1/n2)));
  Pvalu = 2*(1-probnorm(TestStat));
  output;
end;

go to done;

DrawBall:
  link randgen;
  if rand < ProbRed
  then BallOne = 1;  /*draw red ball*/
  else BallOne = 0;  /*draw white ball*/
return;

RedTrmt:  /*trmt 1*/
  RedPat{i} = RedPat{i} + 1;
  link randgen;
  if rand < p{1}
  then
  do;
    RedBall = RedBall + AddBall;
    RedS{i} = RedS{i} + 1;
  end;
  else WhtBall = WhtBall + AddBall;
return;

WhtTrmt:  /*trmt 2*/
  WhtPat{i} = WhtPat{i} + 1;
  link randgen;
  if rand < p{2}
  then
  do;
    WhtBall = WhtBall + AddBall;
    WhtS{i} = WhtS{i} + 1;
  end;
  else RedBall = RedBall + AddBall;
return;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;
return;

done:
  keep Red White Total Power OddsRatio ExpProp Pvalu;
run;
```

```
proc means;
  var Red White Total Power OddsRatio ExpProp Pvalu;
  title "Classic Urn Model, RPW(1, 1, 1)";
run;
```

5. RPW(2n+1, 2n+1, 2n+1)

```
options linesize=80 pagesize=50 nodate;
/*Program for RPW(2n+1, 2n+1, 2n+1)*/

data sim;
  seed=12773;
  patsN = 100;  /*total # of patients treated*/
  AddBall = 3;  /*number of balls to add*/
  NBall = AddBall;  /*# balls of each trmt at start*/
  TestBall = (AddBall-1)/2;
  array p{2};  /*array prob of success of all trmts*/
  p{1} = 0.8;  /*probablity of success for trmt 1; red ball*/
  p{2} = 0.3;  /*probablity of success for trmt 2; white ball*/
  array RedS{5000};  /*total # of successes using trmt 1*/
  array WhtS{5000};  /*total # of successes using trmt 2*/
  array TotalS{5000};  /*total # of successes*/
  array RedPat{5000};  /*total # of patients on trmt 1*/
  array WhtPat{5000};  /*total # of patients on trmt 2*/
  array TotalPat{5000};  /*total # patients*/
  array RedFail{5000};
  array WhtFail{5000};

  dimS = dim(totalS);  /*# of simulations*/
  Z = 1.96;  /*critical value corresponding to given alpha*/
  p1 = p{1};
  p2 = p{2};
  q1 = 1 - p1;
  q2 = 1 - p2; /*variables for power calculation*/

do i = 1 to dimS;  /*perform simulation 5000 times*/
  RedS{i} = 0;
  WhtS{i} = 0;
  TotalS{i} = 0;
  RedPat{i} = 0;
  WhtPat{i} = 0;
  TotalPat{i} = 0;  /*set arrays to zero for each simulation*/
  RedBall = NBall;
  WhtBall = NBall;

  do j = 1 to patsN;
      TotalBall = RedBall + WhtBall;
    ProbWht = WhtBall / TotalBall;
    link DrawBall;
    if Ball > TestBall then link RedTrmt;  /*if draw 2 red, then trmt
1*/
    else link WhtTrmt;  /*if draw 2 white, then trmt 2*/
  end;

  TotalS{i} = RedS{i} + WhtS{i};
  TotalPat{i} = RedPat{i} + WhtPat{i};
  Red = RedS{i}/RedPat{i};
  White = WhtS{i}/WhtPat{i};
  Total = TotalS{i}/TotalPat{i};
  n1 = RedPat{i};
  n2 = WhtPat{i};
```

```
  pbar = (n1*p1 +n2*p2)/(n1 + n2);
  qbar = 1 - pbar;
  Denom = ((p1*q1/n1)+(p2*q2/n2))**0.5;
  Term1 = (p1 - p2)/Denom;
  Term2 = ((pbar*qbar*((1/n1)+(1/n2)))**0.5)/Denom;
  Term3 = Term1 - Z*Term2;
  Power = probnorm(Term3);
  RedFail{i} = RedPat{i} - RedS{i};
  WhtFail{i} = WhtPat{i} - WhtS{i};
  OddsRatio = (RedS{i}/RedFail{i})/(WhtS{i}/WhtFail{i});
  ExpProp = n1/(n1+n2);
  TestStat = (Red-White)/sqrt(Total*(1-Total)*((1/n1)+(1/n2)));
  Pvalu = 2*(1-probnorm(TestStat));
  output;
end;

go to done;

DrawBall:
  Ball = 0;
  do k = 1 to AddBall;  /*# balls drawn = # balls added*/
    link randgen;
    if rand > ProbWht  /*if random # > P(Wht), then red trmt*/
    then Ball = Ball + 1;
  end;
return;

RedTrmt:  /*trmt 1*/
  RedPat{i} = RedPat{i} + 1;
  link randgen;
  if rand < p{1}
  then
  do;
    RedBall = RedBall + AddBall;
    RedS{i} = RedS{i} + 1;
  end;
  else WhtBall = WhtBall + AddBall;
return;

WhtTrmt:  /*trmt 2*/
  WhtPat{i} = WhtPat{i} + 1;
  link randgen;
  if rand < p{2}
  then
  do;
    WhtBall = WhtBall + AddBall;
    WhtS{i} = WhtS{i} + 1;
  end;
  else RedBall = RedBall + AddBall;
return;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;
return;
```

```
done:
  keep Red White Total Power OddsRatio ExpProp Pvalu;
run;

proc means;
  var Red White Total Power OddsRatio ExpProp Pvalu;
  title "RPW(2n+1, 2n+1, 2n+1)";
run;
```

6. Neyman Allocation

```
options linesize=80 pagesize=50 nodate;
/*Program for Neyman Allocation*/

data sim;
  seed=12773;
  patsN = 100;  /*total # of patients treated*/
  array RedS{5000};  /*total # of successes using trmt 1*/
  array WhtS{5000};  /*total # of successes using trmt 2*/
  array TotalS{5000};  /*total # of successes*/
  array RedPat{5000};  /*total # of patients on trmt 1*/
  array WhtPat{5000};  /*total # of patients on trmt 2*/
  array TotalPat{5000};  /*total # patients*/
  array RedFail{5000};
  array WhtFail{5000};

  dimS = dim(totalS);  /*# of simulations*/
  Z = 1.96;  /*critical value corresponding to given alpha*/
  p1 = 0.7;  /* prob of success of red treatment*/
  p2 = 0.4;  /* prob of success of white treatment*/
  q1 = 1 - p1;
  q2 = 1 - p2; /*variables for power calculation*/

do i = 1 to dimS;  /*perform simulation 5000 times*/
  InitRed = 10;
  InitWht = 10;
  InitRedS = InitRed*p1;
  InitWhtS = InitWht*p2;
  RedPat{i} = InitRed;
  WhtPat{i} = InitWht;
  TotalPat{i} = 0;
  RedS{i} = InitRedS;
  WhtS{i} = InitWhtS;
  TotalS{i} = 0; /*set arrays to zero for each simulation*/


  pr = p1;
  qr = 1 - pr;
  pw = p2;
  qw = 1 - pw;  /*probabilities for Neyman allocation*/

  Q = sqrt(pr*qr)/(sqrt(pr*qr)+sqrt(pw*qw)); /*initial Q value*/

  do j = 1 to patsN;

     link randgen;
    if rand < Q
    then link RedTrmt; /*patient receives Red Trmt*/
    else link WhtTrmt; /*patient receives White Trmt*/
    Q = sqrt(pr*qr)/(sqrt(pr*qr)+sqrt(pw*qw));  /*adaptive Q value*/
 end;

  TotalS{i} = RedS{i} + WhtS{i} - InitRedS - InitWhtS;
  TotalPat{i} = RedPat{i} + WhtPat{i} - InitRed - InitWht;
  Red = (RedS{i}-InitRedS)/(RedPat{i}-InitRed);
```

```
  White = (WhtS{i}-InitWhtS)/(WhtPat{i}-InitWht);
  Total = TotalS{i}/TotalPat{i};

  n1 = RedPat{i} - InitRed;
  n2 = WhtPat{i} - InitWht;
  pbar = (n1*p1 +n2*p2)/(n1 + n2);
  qbar = 1 - pbar;
  Denom = ((p1*q1/n1)+(p2*q2/n2))**0.5;
  Term1 = (p1 - p2)/Denom;
  Term2 = ((pbar*qbar*((1/n1)+(1/n2)))**0.5)/Denom;
  Term3 = Term1 - Z*Term2;
  Power = probnorm(Term3);
  RedFail{i} = (RedPat{i}-InitRed) - (RedS{i}-InitRedS);
  WhtFail{i} = (WhtPat{i}-InitWht) - (WhtS{i}-InitWhtS);
  OddsRatio = ((RedS{i}-InitRedS)/RedFail{i})/((WhtS{i}-
InitWhtS)/WhtFail{i});
  ExpProp = n1/(n1+n2);
  output;
end;

go to done;

RedTrmt:  /*trmt 1*/
  RedPat{i} = RedPat{i} + 1;
  link randgen;
  if rand < p1
  then RedS{i} = RedS{i} + 1;
  pr = RedS{i}/RedPat{i}; /*update red variables*/
  qr = 1 - pr;
return;

WhtTrmt:  /*trmt 2*/
  WhtPat{i} = WhtPat{i} + 1;
  link randgen;
  if rand < p2
  then WhtS{i} = WhtS{i} + 1;
  pw = WhtS{i}/WhtPat{i}; /*update white variables*/
  qw = 1 - pw;
return;

randgen:
   a=7**5;
   b=2**31-1;
   seed=mod(a*seed,b);
   rand=seed/b;
return;

done:
  keep i Red White Total Power OddsRatio ExpProp;
run;

proc means;
  var Red White Total Power OddsRatio ExpProp;
  title "Neyman Allocation";
run;
```