Robust Assignment Probabilities for Clinical Trials

by

Fred Ackah

Master of Science Project

September 2004

Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta, Canada

ACKNOWLEDGEMENTS

I am very grateful to my supervisor, Dr. D. Wiens, for his helpful and crucial comments and suggestions throughout the preparation of this project.

Special thanks to my wife Yvonne Ackah, to Grace Takyi and to all friends for their help in writing this project.

1 INTRODUCTION

1.1 BACKGROUND INFORMATION

Suppose that n subjects enter a clinical study and are to be assigned to one of the p treatments. Corresponding to each subject is a vector $\mathbf{x}$ of prognostic factors or covariates, upon which the mean response relies through a vector $\mathbf{z}(\mathbf{x})$ of possible regressors. Upon observing $\mathbf{x}$, the experimenter is to assign the subject to one of the groups. The aim is to obtain an efficient and robust estimate of the difference in the mean responses to the treatments. A complicating factor calling for a robust solution is that the fitted model relating the response to treatment/ covariates effects may be only approximately valid.

Ethical considerations often dictate that there should be some randomness in the assignment of subjects to treatments. Randomization is required to avoid biases which can weaken or destroy the usefulness and applicability of the results of the trial. The most important source of bias is the selection bias which arises when the choice of the next patient to enter the trial is affected by knowledge of the treatment that the patient will receive. A famous non-clinical example is the Lanarkshire milk experiment (Student, 1931) in which teachers swapped treatment and controls so that undernourished children received the milk. The paramount importance of avoiding bias in clinical trials is emphasized by Chalmers (1990). Thus, we shall derive a probability $\rho(\mathbf{x})$ according to which a subject exhibiting covariates $\mathbf{x}$ is to be assigned to the treatments or groups.

As an example, assume there are p treatments $(0, 1, ..., p)$ in a sequential clinical trial. When a patient arrives, the clinician should assign him a treatment immediately. Suppose the prognostic factor of the patients is Age with the following categories:

A	Age under 20
B	Age between 20 and 40
C	Age above 40

.

Also suppose we already have ten patients in the trial with the prognostic factor as follows:

Patient	Age	Treatment
1	C	1
2	A	0
3	B	.
4	C	.
5	A	.
6	C	.
7	B	.
8	B	.
9	A	p
10	B	.

Now the 11^{th} patient is coming with prognostic factor Age = A, then which treatment should we assign to him/her?

1.2 A REVIEW OF THE ALLOCATION SCHEMES IN CLINICAL TRIALS

1.2.1 a) COMPLETELY RANDOM ALLOCATION

This is the simplest method of allocating subjects in groups. We assign subjects randomly into p groups (treatments and the control, for example) with the same probability, independently of the assignment for the other subjects, and also ignore prognostic factors. In this method the probability of allocation is $1/p$.

Compared to the perfectly balanced experiment, completely random allocation has the advantage that bias is avoided while in the perfectly balanced experiment the experimenter may know for certain what the next assignment may be.

This method suffers from some disadvantages. When there are small numbers of subjects in the experiment, the final distribution of groups can be very unbalanced.

1.2.2 b) PERMUTED BLOCK DESIGNS

One of the best allocation schemes that helps in eliminating imbalance in treatments is the permuted block procedure (Zelen 1974). This design divides the subjects into blocks of even length. For example, if the total number of subjects in the trial is 40 and the number of groups is two (i.e. treatment and control groups), then we divide these subjects into blocks of 20 each. Within each block randomly assign 20 subjects to treatment and control groups respectively.

In general, if we have n patients, we should divide them into blocks with length $2k$ ($0 < k < n/2$), and within each block randomly assign k subjects to treatment and k subjects to the control groups.

The permuted blocks can be quite effective in eliminating unbalanced designs. But the disadvantage is that at some point, the experimenter knows for certain which treatment the next arrival will be assigned, which brings about bias. Also it does not make provision for prognostic factors.

1.2.3 c) EFRON'S BIASED COIN DESIGN

Efron (1971) provided the biased coin design.

Suppose at a certain stage in the experiment a new subject arrives and we already have D more treatment assignments than control assignments. We will assign the new subject as follows:

If $D > 0$, assign to the treatment group with probability q and to the control group with probability p.

If $D = 0$, assign to the treatment group with probability $1/2$ and to the control group with probability $1/2$.

If $D < 0$, assign to the treatment group with probability p and to the control group with probability q.

Here $p > q$, $p + q = 1$. Efron suggested that $p = 2/3$, saying that this number is big enough to yield generally good designs.

This biased coin design method can achieve greater balance than completely random allocation. But when there are few subjects, it is likely to be out of balance.

It also does not make provision for prognostic factors.

1.2.4 d) MINIMIZATION METHOD OF TAVES

In this method, Taves (1974) considered the number of patients in each group who had the same characteristics as the patient about to be assigned. By doing this, we minimize the differences between the treatment groups. Let d_i be the absolute difference when the new subject is added to treatment i and d_j the corresponding absolute difference when the new treatment is added treatment j. Traves proposed that the $(n+1)^{th}$ subject be added to treatment i when $d_i \leq d_j$.

The use of randomization occurs only when the placement of the new patient makes no difference in the comparability of the two groups. The disadvantage with this method is that, it works well only when the number of patient is large. This is because maintaining balance across each stratum becomes difficult when the number of patients is small.

1.2.5 e) THE METHOD OF POCOCK AND SIMON

Pocock and Simon (1975) suggested a procedure for treatment assignment that concentrates on minimizing imbalance within the levels of each individual prognostic factor. The improvement of this method compared with the method of Traves is that we have more randomization in this method.

Let d_{irk} be the range of treatment difference for level r when treatment k were assigned to the new patient who has level r in factor i. This measures

the "lack of balance". The term G_k is the sum of d_{irk} with respect to i. It represents the total amount of imbalance.

Hence treatments with small values of G_k have a higher probability of being chosen. The aim here is to balance the number of patients in each stratum (i.e., each cell of the factorial array) that receives each treatment. Such balance rapidly becomes impossible for even a few covariates unless the number of patients is large.

1.2.6 f) THE METHOD OF BEGG AND IGLEWICZ

Begg and Iglewicz (1980) proposed a scheme for balancing the treatments. In their method, the primary endpoint is the comparison of two treatments with respect to an outcome variable that is related to treatment and binary prognostic factors. But sometimes, not all prognostic factors are binary data. In that case, we can combine some similar levels together, and only levels for each prognostic factor can be combined together.

Suppose M_{--}, M_{-+}, M_{+-}, M_{++} represents the current treatment-factor totals, the first subscript representing treatment. Now, we will calculate the difference between the treatment imbalances in each of the two levels, that is, $(M_{--} - M_{-+}) - (M_{+-} - M_{++})$. If the result is negative, we will assign the new patient treatment $(+)$. Otherwise, if the result is positive, we will assign treatment $(-)$ to the new subject.

1.2.7 g) ATKINSON'S D_A - OPTIMUM DESIGNS

Atkinson (1982) used optimal design theory to provide a procedure of the biased coin type for sequential clinical trials in the presence, or absence of prognostic factors.

Silvey (1980) stated that optimal design aims at choosing a design matrix $\mathbf{X}$ that minimizes various functions of the covariance matrix of the regression coefficients $\hat{\boldsymbol{\beta}}$, in a model in which the outcomes are to be regressed on the treatment assignments and possibly the covariates. In this case, the linear

model is $E(Y) = \mathbf{X}^T\boldsymbol{\beta}$ with independent observations of variance σ^2. Here $\hat{\boldsymbol{\beta}}$ is $(p+q) \times 1$, where p is the number of treatments and q is the number of covariates. We find $\mathbf{X}$ such that (the trace, determinant etc. of) $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ is minimized.

Let $\mathbf{M} = n^{-1}(\mathbf{X}^T\mathbf{X})$ and let $d(\mathbf{x}) = \mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}$ be the standardized variance of the parameter estimates.

In clinical trials, interest is often in contrasts between treatment effects. Suppose the contrasts are s linear combinations which are elements of the vector $\mathbf{A}\boldsymbol{\beta}$, where $\mathbf{A}$ is an $s \times p$ matrix of rank $s < p$. The covariance matrix of the least squares estimate $\mathbf{A}\boldsymbol{\beta}$ is proportional to $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T$. The analogue of D-optimality is to maximize the determinant of $\{\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T\}^{-1}$. This is called $D_{\mathbf{A}}$-optimality by Sibson (1974).

The corresponding standardized variance is:

$$d_A(\mathbf{x}) = \mathbf{x}^T\mathbf{M}^{-1}\mathbf{A}^T\{\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T\}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{x}. \tag{1}$$

The standardized variance represents the variance of the predicted response over the design region. Sequential $D_{\mathbf{A}}$-optimum designs are generated one trial at a time by adding an observation at that $\mathbf{x}$ where (1) is a maximum. The design region consists of p points, the i^{th} of which corresponds to allocating treatment i to the next patient. Let the corresponding value of $d_A(\mathbf{x})$ be $d_A(i)$. The sequential construction of the optimum design allocates the $(n+1)^{th}$ subject to the treatment (i) for which $d_A(i)$ is a maximum.

Atkinson proposed that if randomization is required, we can choose treatment i with probability

$$P_i = \frac{d_A(i)}{\sum_{i=1}^{p} d_A(i)}. \tag{2}$$

For example, with two treatments and with no prognostic factors (i.e. $i = 1, 2$), $\mathbf{A} = (1, -1), \mathbf{M} = diag\{\frac{n_i}{n}\}$, a diagonal matrix. Substituting this in (2), shows that for the criterion of $D_{\mathbf{A}}-$optimality,

$$d_A(1) = \frac{n_2}{n_1},\ d_A(2) = \frac{n_1}{n_2}$$

and the resulting probability of selecting treatment 1 is

$$p_1 = \frac{n_2^2}{n_1^2 + n_2^2}.$$

In our example, $\mathbf{x}_1$ is the vector of indicator variables for the treatments and $\mathbf{x}_2$ is the vector of indicator variable for the prognostic factor. These are then

$$\mathbf{x}_1 = \begin{cases} (1,0,0)^T, & \text{for Treatment 1,} \\ (0,1,0)^T, & \text{for Treatment 2,} \\ (0,0,1)^T, & \text{for Treatment 3,} \end{cases}$$

$$\mathbf{x}_2 = \begin{cases} (0,1)^T, & \text{if Age = A,} \\ (1,0)^T, & \text{if Age = B,} \\ (0,0)^T, & \text{if Age = C.} \end{cases}$$

Also, we defined our contrast as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \end{pmatrix},$$

since interest is in only the difference between treatments effects.

In our example

$$\mathbf{X} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

and we calculate that

$$\mathbf{M} = \frac{(\mathbf{X}^T\mathbf{X})}{n} = \begin{pmatrix} .3 & 0 & 0 & 0.1 & 0.1 \\ 0 & 0.4 & 0 & 0.2 & 0 \\ 0 & 0 & 0.3 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.4 & 0 \\ 0.1 & 0 & 0.2 & 0 & 0.3 \end{pmatrix}.$$

Partition $\mathbf{A}$ as $\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \vdots \mathbf{0} \end{pmatrix}$ for

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix},$$

and partition $\mathbf{M}^{-1}$ as

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{M}^{11} : 3 \times 3 & \mathbf{M}^{12} : 3 \times 2 \\ \mathbf{M}^{21} : 2 \times 3 & \mathbf{M}^{22} : 2 \times 2 \end{pmatrix}.$$

Define

$$\mathbf{B} = \mathbf{A}_1^T(\mathbf{A}_1\mathbf{M}^{11}\mathbf{A}_1^T)^{-1}\mathbf{A}_1 = \begin{pmatrix} 0.2083 & -0.1167 & -0.0917 \\ -0.1167 & 0.1667 & -0.0500 \\ -0.0917 & -0.0500 & 0.1417 \end{pmatrix}.$$

For treatment 1, $\mathbf{x}_1^T = (1, 0, 0)$, $\mathbf{x}_2^T = (0, 1)$ and then

$$d_A(1) = \mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{11}\mathbf{x}_1 + 2\mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 + \mathbf{x}_2^T\mathbf{M}^{21}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 = 3.5088.$$

For treatment 2, $\mathbf{x}_1^T = (0, 1, 0)$, $\mathbf{x}_2^T = (0, 1)$ and

$$d_A(2) = \mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{11}\mathbf{x}_1 + 2\mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 + \mathbf{x}_2^T\mathbf{M}^{21}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 = 7.0615.$$

For treatment 3, $\mathbf{x}_1^T = (0, 0, 1)$, $\mathbf{x}_2^T = (0, 1)$ and

$$d_A(3) = \mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{11}\mathbf{x}_1 + 2\mathbf{x}_1^T\mathbf{M}^{11}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 + \mathbf{x}_2^T\mathbf{M}^{21}\mathbf{B}\mathbf{M}^{12}\mathbf{x}_2 = 0.8772.$$

We will assign the new patient treatment 1 with probability

$$p_1 = \frac{d_A(1)}{d_A(1) + d_A(2) + d_A(3)} = 0.3065,$$

assign the new patient treatment 2 with probability

$$p_2 = \frac{d_A(2)}{d_A(1) + d_A(2) + d_A(3)} = 0.6169$$

and assign the new patient treatment 3 with probability

$$p_3 = \frac{d_A(3)}{d_A(1) + d_A(2) + d_A(3))} = 0.0766.$$

1.3 SUMMARY OF THE PROBLEMS IN §1.2

In these methods reviewed, the response function was assumed to be exactly correct. If the linear model is not correct, then the least squares estimator is biased, and the resulting classical optimal designs may yield large bias (compared with the variance). The presence of $f(\mathbf{x})$ in the model biases our estimate $\hat{\boldsymbol{\theta}}$, so optimal assignment probabilities should aim at minimizing the mean square error (mse) rather than merely the variance.

The classical optimal designs which minimize only the variance alone are not optimal anymore because of the contribution of the bias. Box and Draper (1959) made apparent the dangers of designing a regression experiment with the assumption that the response function is exactly correct. They found that very small departures from the model assumptions can eliminate any supposed gains arising form the use of a design which minimizes variance alone. See also Wiens (1992). Huber (1981) points out that "deviations from linearity that are too small to be detected are already large enough to tip the balance away from the 'optimal designs', which assume exact linearity and put observations on the extreme points of the observable range, toward the 'naive' ones which distribute the observations more or less evenly over the entire design space".

Also, when the model assumption is violated we cannot do model adequacy testing by using the classical optimal designs. We only make observations on the extreme points of the space S in question.

In practice, the relationship between the response variable and the explanatory variable is usually only approximate. This often results in violations of the model assumptions. Therefore there is the need to study optimal designs under possible small violations of the model assumptions.

1.4 THE METHOD OF N. E. HECKMAN (1987)

Heckman (1987) suggested that we find an optimal probability $P(x)$ such that $MSE(\widehat{\boldsymbol{\beta}(x)}, P(x), \mathbf{X})$ is minimax. In this paper, $\widehat{\boldsymbol{\beta}(x)}$ is the estimator that minimizes the maximum mean square error $MSE(\widehat{\boldsymbol{\beta}(x)}) = E\left[(\boldsymbol{\beta}(x) - \widehat{\boldsymbol{\beta}(x)})^2\right]$, where the maximum is taken over all the functions $Y = \beta_0 + \beta_1\delta + f(\delta, x', x)$ and $\boldsymbol{\beta}(x)$ is the true parameter. The class used in this paper is

$$\digamma = \{f_i : |n^{-1/2} f_i(x',x)| \leq \eta_i(x')|x - x'|\}.$$

The class $\digamma$ was required such that f_i, the disturbance functions is "small " and the term $\mathbf{v}_i^T(x)\widehat{\boldsymbol{\beta}(x)}$ is still the leading term in our approximate model. The radius $\eta_i(x^{'})$ is assumed known, $\mathbf{X}$ is the design matrix and $i = 1, 2$.

The value of the maximum mean squared error of $\widehat{\boldsymbol{\beta}(x)}$ for the allocation vector $P(x)$, and covariate x, is denoted by $MSE(\widehat{\boldsymbol{\beta}(x)}, P(x), \mathbf{X})$. The optimal assignment probability $P(x)$ is found by minimizing the maximum $MSE(\widehat{\boldsymbol{\beta}(x)}, P(x), \mathbf{X})$, with this maximum taken over $\digamma$.

The author also stated that determining $\text{MSE}(\widehat{\boldsymbol{\beta}(x)}, P(x), \mathbf{X})$ involved solving a system of integral equations, so she used asymptotic approximation to find the optimal probabilities of allocation.

This method does not apply to the cases with more than two treatments and more than one prognostic factor.

1.5 THE METHOD OF D. P. WIENS (2000)

In this paper, Wiens obtained minimax assignment probabilities using a neighbourhood structure which is defined by

$$E\left[\text{Outcome}|\text{Treatment } i, \text{ covariates } \mathbf{x}\right] = \theta_i + \boldsymbol{\gamma}^T\mathbf{v}(\mathbf{x}) + f_i(\mathbf{x}) \text{ for } f_i \in \digamma,$$

where $\digamma$ is a class of functions satisfying

$$\int_S \left(1, \mathbf{v}^T(\mathbf{x})\right) f_i(\mathbf{x})\mu(d\mathbf{x}) = 0^T, \quad \int_S f_i^2(\mathbf{x})\mu(d\mathbf{x}) \leq \eta_i^2, i = 1, 2, ..., p. \qquad (3)$$

Here, $f_i(\mathbf{x})$ is the disturbance functions introduced because of the approximate nature of the assumed model, p is the number of groups, $\mathbf{v}(\mathbf{x})$ is a vector of regressors, μ is a measure on S and the radii η_i are assumed known and small.

The least squares estimate was used in his paper. The loss function is maximized over the class above and the corresponding expression was minimized over the probability of allocation given a prognostic factor.

For example, in the case where we have only two groups the optimal assignment probability of allocation obtained was given by

$$\rho_2(\mathbf{x}) = \frac{\sigma_2}{\sigma_1 + \sigma_2},$$

where σ_1 and σ_2 are the standard deviations which are estimated by the median absolute deviations of the corresponding residuals. In the case where the σ_i are constant, the optimal assignment probability is given by $\rho_i(\mathbf{x}) = \frac{1}{p}, i = 1, ..., p.$

There is increased imbalance across prognostic factors when this method is used.

In this project, I intend to compute robust assignment probabilities and compare the robust allocations with those resulting from Atkinson's method.

2 METHOD

With reference to Wiens's technical report (Wiens 2000), the loss function was given by

$$L(\rho_1, \rho_2, ..., \rho_p) = |\mathbf{W}\mathbf{B}^{-1}\mathbf{Q}\mathbf{B}^{-1}\mathbf{W}^T| \left(1 + \max_{f_1,...,f_p} ||\mathbf{R}\mathbf{b}(f_1, ..., f_p)||^2\right)$$

with

$$\max_{f_1,...,f_p} ||\mathbf{R}\mathbf{b}(f_1, ..., f_p)||^2 = \left(\max_{||\boldsymbol{\alpha}||=1} \sum_{i=1}^{p} \eta_i \sqrt{\boldsymbol{\alpha}^T \mathbf{R}\mathbf{G}_i\mathbf{R}^T\boldsymbol{\alpha}}\right)^2 \tag{4}$$

and the definitions of $\mathbf{R}, \mathbf{b}(f_1, ..., f_p), \mathbf{G}_i$ can be found in Wiens (2000).

For $p > 2$ the final maximum in (4) is too intractable to admit an exact analysis. We shall replace it with an upper bound which is given (Wiens 2000) by

$$L(\rho_1, \rho_2, ..., \rho_p) \leq \frac{\eta_1^2 \sigma_1^{2(p-2)}}{p} L'(\rho_1, \rho_2, ..., \rho_p). \tag{5}$$

Here

$$L'(\rho_1, \rho_2, ..., \rho_p) = \frac{1}{|\mathbf{D}|} \left[v tr\mathbf{D} + \left\{ \sum_{i=1}^{p} \frac{k_i}{w_i} \sqrt{\frac{g_i}{s_i} \left(tr\mathbf{D} - \frac{s_i}{w_i^2} \right)} \right\}^2 \right], \tag{6}$$

$w_i = \sigma_i^2/\sigma_1^2$ is the variance ratio, the '*noise/noise*' ratio $k_i = \eta_i/\eta_1$ and the noise to contamination ratio $v = \sigma_1^2/\eta_1$. Also, $\mathbf{D} = diag(s_1, s_2/w_2^2, .., s_p/w_p^2)$.

For $p = 2$ the inequality in (5) is an equality. We aim to find optimal assignment probabilities such that the upper bound $L'(\rho_1, \rho_2, ..., \rho_p)$ is minimized over $(\rho_1, \rho_2, ..., \rho_p)$. We define

$$\begin{aligned} a &= \int_S \mu(d\mathbf{x}), \\ s_i &= \int_S \rho_i(\mathbf{x}) m(\mathbf{x}) \mu(d\mathbf{x}), \\ g_i &= \int_S (\rho_i(\mathbf{x}) m(\mathbf{x}) - \frac{s_i}{a})^2 \mu(d\mathbf{x}). \end{aligned}$$

Further details can be found in Wiens (2000). In this paper, it was shown that for a two group problem, the solution was

$$\begin{aligned} \rho_1(\mathbf{x}) &= \left[\frac{\tau_{1,1}}{m(\mathbf{x})} + \tau_{1,2} \right]_0^1, \\ \rho_2(\mathbf{x}) &= 1 - \rho_1(\mathbf{x}), \end{aligned}$$

with the constants $\tau_{1,1}, \tau_{1,2}$ chosen to minimize L' in (6). This suggests that we might consider assignment probabilities

$$\rho_i(\mathbf{x}) = \frac{\left[\frac{\tau_{i,1}}{m(\mathbf{x})} + \tau_{i,2} \right]^+}{\sum_{j=1}^{p} \left[\frac{\tau_{j,1}}{m(\mathbf{x})} + \tau_{j,2} \right]^+}, i = 1, 2, ...p. \tag{7}$$

(Here, $[x]^+ = \max(x, 0)$.) This is guaranteed to be nonnegative, and since the $\rho_i(\mathbf{x})$'s sum to 1 they must also be ≤ 1. The form (7) is overparameterized,

since for instance we could divide each τ by $|\tau_{1,2}|$. Thus we take $\tau_{1,2} = \pm 1$ and minimize over the remaining $2p-1$ constants. Note that the solution shown to be optimal when $p = 2$ is a special case of this class of functions.

As an example, when $p = 3$, we need to find $\tau_{1,1}, \tau_{2,1}, \tau_{2,2}, \tau_{3,1}, \tau_{3,2}$ such that (6) is minimized. Then the optimal assignment probabilities will be

$$\begin{aligned}
\rho_1(\mathbf{x}) &= \frac{\left[\frac{\tau_{1,1}}{m(\mathbf{x})} + 1\right]^+}{\sum_{j=1}^3 \left[\frac{\tau_{j,1}}{m(\mathbf{x})} + \tau_{j,2}\right]^+}, \\
\rho_2(\mathbf{x}) &= \frac{\left[\frac{\tau_{2,1}}{m(\mathbf{x})} + \tau_{2,2}\right]^+}{\sum_{j=1}^3 \left[\frac{\tau_{j,1}}{m(\mathbf{x})} + \tau_{j,2}\right]^+}, \\
\rho_3(\mathbf{x}) &= \frac{\left[\frac{\tau_{3,1}}{m(\mathbf{x})} + \tau_{3,2}\right]^+}{\sum_{j=1}^3 \left[\frac{\tau_{j,1}}{m(\mathbf{x})} + \tau_{j,2}\right]^+},
\end{aligned}$$

where $m(\mathbf{x})$ is the density (with respect to counting measure μ) of the standardized covariate (e.g. age of the subject).

We impose additional constraints

$$w_i = k_i, i = 1, ..., p,$$

These conditions means that variance and bias are of the same relative order within each group. So

$$g_i = \int_S \rho_i^2(\mathbf{x}) m^2(\mathbf{x}) \mu(d\mathbf{x}) - \frac{s_i^2}{a},$$

(6) reduces to

$$L'(\rho_1, \rho_2, ..., \rho_p) = \frac{1}{|\mathbf{D}|} \left[v tr\mathbf{D} + \left\{ \sum_{i=1}^p \sqrt{\frac{g_i}{s_i} \left(tr\mathbf{D} - \frac{s_i}{w_i^2} \right)} \right\}^2 \right], \qquad (8)$$

and the problem reduces to the minimization of (8) .

2.1 Example

With reference to our example in §1, we have $p = 3$ treatments and a discrete covariate $x = Age$ with three categories ($J = 3$). We let μ be counting measure on the discrete space S. Then in this case, $m_j = P(x = j)$ for $j \in \{A, B, C\}$. We have

$$\begin{aligned} S &= \{1, 2, 3\}, \\ a &= \sum_{i=1}^{3} 1 = 3, \\ s_i &= \sum_{j=1}^{3} \rho_i(j) m_j \\ g_i &= \sum_{j=1}^{3} \rho_i^2(j) m_j^2 - \frac{1}{3} s_i^2, \ i = 1, 2, 3. \end{aligned}$$

We now substitute these values into (8) and minimize numerically for $\tau_{i,1}$, $\tau_{i,2}$ using (7).

To minimize (8), we used Matlab function "FMINSEACH" which finds the unconstrained minimum of a scalar function of several variables starting at an initial value. At this initial value, (8) is found using the expression for (7). This function value is compared with the previously calculated value and based on the minimum of the two, new $\tau's$ are calculated. The procedure is repeated until the minimum $\tau's$ are obtained.

In this minimization, we assumed $w = (w_1, w_2, w_3) = (1, 3, 5)$, to show the effects of differing variability within the treatments and $v = 1$, to show the experimenter's equal importance of variance versus bias.

Table 1 shows the minimizing values of $\tau_{i,1}$and $\tau_{i,2}$ for three and four treatments when $(m_A, m_B, m_C) = (0.3, 0.5, 0.2)$.

We now calculate the corresponding robust assignment probabilities for

Table 1. Minimizing values of the constants for 3 and 4 treatments; $(m_A, m_B, m_C) = (0.3, 0.5, 0.2)$		
i	$\tau_{i,1}$	$\tau_{i,2}$
$p = 3, w = [1, 3, 5]$		
1	3.2896	1.
2	5.1150	−1.5549
3	5.7536	−1.7490
$p = 4, w = [1, 3, 5, 7]$		
1	1.6268	1
2	2.0678	2.4241
3	2.3578	2.4774
4	2.497	2.5156

three treatments as

$$\rho_1(j) = \left[\frac{3.2896}{m_j} + 1\right]^+ / \left(\left[\frac{3.2896}{m_j} + 1\right]^+ + \left[\frac{5.1150}{m_j} - 1.5549\right]^+ + \left[\frac{5.7536}{m_j} - 1.7490\right]^+\right),$$

$$\rho_2(j) = \left[\frac{5.1150}{m_j} - 1.5549\right]^+ / \left(\left[\frac{3.2896}{m_j} + 1\right]^+ + \left[\frac{5.1150}{m_j} - 1.5549\right]^+ + \left[\frac{5.7536}{m_j} - 1.7490\right]^+\right),$$

$$\rho_3(j) = \left[\frac{5.7536}{m_j} - 1.7490\right]^+ / \left(\left[\frac{3.2896}{m_j} + 1\right]^+ + \left[\frac{5.1150}{m_j} - 1.5549\right]^+ + \left[\frac{5.7536}{m_j} - 1.7490\right]^+\right).$$

Here, $m_j = P(Age = j)$. Since the 11^{th} patient was in age class A, if we let $m_A = 0.3$ then we will assign the new patient treatment 1 with $\rho_1(A) = 0.2665$, treatment 2 with $\rho_2(A) = 0.3452$ and treatment 3 with $\rho_3(A) = 0.3883$.

Table 2 shows minimizing values for $\tau's$ for various values of m_j and Table 3 also shows their corresponding assignment probabilities.

So ρ depends on neither treatment nor age of the patient. The probabilities (ρ) depend on the variance ratios $w = (w_{1,}w_2, w_3)$.

Table 2. Minimizing values of the constants for 3 treatments, 3 age levels, and a variety of probability distributions on 'Age'.

$\{m_j\}$	Treatment $i=1$		Treatment $i=2$		Treatment $i=3$	
	$\tau_{1,1}$	$\tau_{1,2}$	$\tau_{2,1}$	$\tau_{2,2}$	$\tau_{3,1}$	$\tau_{3,2}$
$(.01,.5,.49)$	2.3588	1	1.8552	2.1721	2.0806	2.4557
$(.1,.5,.4)$	3.9347	1	6.1180	−1.5549	6.8818	−1.7490
$(.3,.5,.2)$	3.2896	1	5.1150	−1.5549	5.7536	−1.7490
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	2.0990	1	1.9906	2.2644	2.2737	2.4432
$(.6,.3,.1)$	1.5790	1	2.4552	−1.5549	2.7617	−1.7490
$(.1,.3,.6)$	1.5790	1	2.4552	−1.5549	2.7617	−1.7490
$(.3,.1,.6)$	1.5790	1	2.4552	−1.5549	2.7617	−1.7490
$(.5,.25,.25)$	1.3761	1	−0.7982	7.2150	1.0942	2.1600
$(.25,.5,.25)$	1.3761	1	−0.7982	7.2150	1.0942	2.1600

Table 3. Robust assignment probabilities for 3 treatments, 3 age levels, and a variety of probability distributions on 'Age'.

$\{m_j\}$	$\{\rho_i(j)\}$								
	Age $j=A$			Age $j=B$			Age $j=C$		
	$i=1$	$i=2$	$i=3$	$i=1$	$i=2$	$i=3$	$i=1$	$i=2$	$i=3$
$(.01,.5,.49)$	0.3730	0.2955	0.3315	0.3139	0.3229	0.3632	0.3147	0.3225	0.3628
$(.1,.5,.4)$	0.2415	0.3569	0.4015	0.2810	0.3384	0.3806	0.2707	0.3432	0.3861
$(.3,.5,.2)$	0.2665	0.3452	0.3883	0.2914	0.3335	0.3751	0.2548	0.3507	0.3945
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	0.2943	0.3321	0.3736	0.2943	0.3321	0.3736	0.2943	0.3321	0.3736
$(.6,.3,.1)$	0.4025	0.2812	0.3163	0.3078	0.3258	0.3664	0.2557	0.3503	0.3940
$(.1,.3,.6)$	0.2557	0.3503	0.3940	0.3078	0.3258	0.3664	0.4025	0.2812	0.3163
$(.3,.1,.6)$	0.3078	0.3258	0.3664	0.2557	0.3503	0.3940	0.4025	0.2812	0.3163
$(.5,.25,.25)$	0.2735	0.4095	0.3170	0.3812	0.2357	0.3831	0.3812	0.2357	0.3831
$(.25,.5,.25)$	0.3812	0.2357	0.3812	0.2735	0.4095	0.3831	0.3812	0.2357	0.3831

3 SIMULATION

In the previous section, we found robust optimal assignment probabilities for three and four groups. Next, we want to know their effectiveness and appropriateness in actual clinical trials. I am going to use computer simulations based on a particular model to compare the robust allocation with that of Atkinson's method.

The model chosen is as follows:

- There are 10 patients initially in the trial with generated prognostic factor (Age) and treatments $\langle 1, 2, 3, 1, 2, 3, 1, 2, 3, 1\rangle$ respectively. Here, I started with a balanced situation, which is easier for the further operation.

- The final number of patients entered into the clinical trial each time is fixed at $n = 60$. I ran the simulation 1000 times. This should be large enough to illustrate the general properties of the methods.

- There is one prognostic factor - "age of the patient" - with levels A, B, C respectively. Balanced treatment numbers are desired for this prognostic factor.

- Patients enter the trial sequentially in purely random manner, factor levels of any one patient being independent of those for any other.

- The regression model used in the simulation is

$$E(Y) = \theta_1 u_1 + \theta_2 u_2 + \theta_2 u_3 + \gamma_1 x_1 + \gamma_2 x_2$$

  Here, $(u_1, u_2, u_3)^T = (1, 0, 0)^T$ for treatment one, $= (0, 1, 0)^T$ for treatment two, $= (0, 0, 1)^T$ for treatment three and $(x_1, x_2)^T = (0, 1)^T$ for category A , $= (1, 0)$ for category B.

- In this simulation, we used the multinomial distribution with probabilities of going into the three categories as m_A, m_B, m_C respectively. We choose $m_A = 0.3, m_B = 0.5, m_C = 0.2$ and so Age was generated from a multi(50, m_A, m_B, m_C).

- The same age sequence was used for Atkinson's method and the robust method in each run.

3.1 Measure of imbalance between treatments

One purpose of all the allocation methods is to avoid *imbalance* for the treatment. For the two methods, we are supposed to balance treatments across prognostic factors. Thus the measure of imbalance should include the imbalance of treatment for each factor combination. A measure of imbalance in cases is given by Wiens (2004):

$$D_n = \sum_{i=1}^{p} \frac{1}{n} \sum_{l=1}^{L} n_{.l} (\frac{n_{i,l}}{n.l} - \frac{n_{i.}}{n})^2 \tag{9}$$

where:

i represent the treatments, in our case $i = 1, 2, 3$.

l is the different combination of the prognostic factors. Since there is only one prognostic factor, $L = 3$ and the three "combinations" are merely the three age groups.

$n_{i,l}$ is the number of times in the n assignment that a patient at level l receives treatment i. I will calculate the average for the imbalance after each assignment for the 1000 simulations.

With reference to our example in §1, we illustrate how the imbalance is found. Suppose that initially we have the following table.

Treatment	Age
1	3
2	3
3	2
1	3
2	1
3	3
1	2
2	2
3	1
1	1

.

Using the above definitions for n_{il}, we have

$$\begin{aligned} n_{1\cdot} &= 4, n_{2\cdot} = 3, n_{3\cdot} = 3, n_{\cdot 1} = 3, n_{\cdot 2} = 3, n_{\cdot 3} = 4 \\ n_{11} &= 1, n_{12,} = 1, n_{13} = 2, n_{21} = 1, n_{22} = 1, n_{23} = 1 \\ n_{31} &= 1, n_{32} = 1, n_{33} = 1, \end{aligned}$$

and then calculate

$$\begin{aligned} D_{10} &= \frac{1}{10}\left[3*(1/3-4/10)^2+3*(1/3-4/10)^2+4*(2/4-4/10)^2\right]+ \\ &\quad \frac{1}{10}\left[3*(1/3-3/10)^2+3*(1/3-3/10)^2+4*(1/4-3/10)^2\right]+ \\ &\quad \frac{1}{10}\left[3*(1/3-3/10)^2+3*(1/3-3/10)^2+4*(1/4-3/10)^2\right] \\ &= 0.0101. \end{aligned}$$

Also when the 40^{th} patient with age C is allocated into treatment two, then for Atkinson's method we have

$$\begin{aligned} n_{1\cdot} &= 13, n_{2\cdot} = 15, n_{3\cdot} = 12, n_{\cdot 1} = 10, n_{\cdot 2} = 20, n_{\cdot 3} = 10 \\ n_{11} &= 4, n_{12,} = 6, n_{13} = 3, n_{21} = 4, n_{22} = 7, n_{23} = 4 \\ n_{31} &= 2, n_{32} = 7, n_{33} = 3, \end{aligned}$$

and for the robust method

$$\begin{aligned} n_{1\cdot} &= 10, n_{2\cdot} = 18, n_{3\cdot} = 12, n_{\cdot 1} = 10, n_{\cdot 2} = 9, n_{\cdot 3} = 21 \\ n_{11} &= 3, n_{12,} = 1, n_{13} = 6, n_{21} = 5, n_{22} = 1, n_{23} = 12 \\ n_{31} &= 2, n_{32} = 7, n_{33} = 3. \end{aligned}$$

Then the imbalance for Atkinson's method is

$$\begin{aligned} D_{40} &= \frac{1}{40}\left[10*(4/10-13/40)^2+20*(6/20-13/40)^2+10*(3/10-13/40)^2\right]+ \\ &\quad \frac{1}{40}\left[10*(4/10-15/40)^2+20*(7/20-15/40)^2+10*(4/10-15/40)^2\right]+ \\ &\quad \frac{1}{40}\left[10*(2/10-12/40)^2+20*(7/20-12/40)^2+10*(3/10-12/40)^2\right] \\ &= 0.0063. \end{aligned}$$

and the corresponding imbalance for the Robust method is

$$\begin{aligned} D_{40} &= \frac{1}{40}\left[10*(3/10-10/40)^2+9*(1/9-10/40)^2+21*(6/21-10/40)^2\right]+ \\ &\quad \frac{1}{40}\left[10*(5/10-18/40)^2+9*(1/9-18/40)^2+21*(12/21-18/40)^2\right]+ \\ &\quad \frac{1}{40}\left[10*(2/10-12/40)^2+9*(7/9-12/40)^2+21*(3/21-12/40)^2\right] \\ &= 0.1067. \end{aligned}$$

3.2 The measure of loss between the methods

We consider the efficiency and robustness of these methods, by evaluating the loss functions $\sum_{i=1}^{3} d_A(i)$ (measuring efficiency) and L' (incorporating robustness) after a new patient is allocated. The efficiency and robustness of a particular method will be assessed on how the new allocation affects the loss functions. Assuming the $(n+1)^{th}$ patient arrives sequentially to our trial, we will evaluate each loss for the two method after this new patient has been assigned a treatment.

Considering our example in §1, we compute the new probability as

$$\begin{aligned}
\rho_i(j) &= P(Treatment = i | Age = j), \\
\hat{\rho}_i(j) &= \frac{\text{number of Age in category } j \text{ who got treatment } i}{\text{Total number of Age in category } j}, \\
\hat{\rho}_i(j) &= \frac{n_{ij}}{n_{\cdot j}}.
\end{aligned}$$

Then

$$\begin{aligned}
\widehat{s_i} &= \sum_{j=1\hat{\rho}i}^{3} (j) m_j \\
\hat{g}_i &= \sum_{j=1}^{3} \hat{\rho}_i^2(j) m_j^2 - \frac{1}{3}\hat{s}_i^2, \ i = 1, 2, 3
\end{aligned}$$

and these values are substituted in (8) to obtain $L'(\widehat{\rho_1}, \widehat{\rho_2}, \widehat{\rho_3})$ after each assignment is made.

Using the example in the previous section, we have

$$\begin{aligned}
\hat{\rho}_1(1) &= 1/3, \hat{\rho}_1(2) = 1/3, \hat{\rho}_1(3) = 2/4, \hat{\rho}_2(1) = 1/3, \hat{\rho}_2(2) = 1/3, \\
\hat{\rho}_2(3) &= 1/4, \hat{\rho}_3(1) = 1/3, \hat{\rho}_3(2) = 1/3, \hat{\rho}_3(3) = 1/4,
\end{aligned}$$

and

$$\begin{aligned}
\hat{s}_1 &= 0.3667, \hat{s}_2 = 0.3167, \hat{s}_2 = 0.3167, \\
\hat{g}_1 &= 0.0030, \hat{g}_2 = 0.0069, \hat{g}_3 = 0.0069, \\
|D| &= 0.00016342, \ trD = 0.4145.
\end{aligned}$$

Using (8), the loss functions for both methods after the 10^{th} patient has been assigned is $L'(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3) = 2790$.

Similarly, when the 12^{th} patient is assigned a treatment, then for Atkinson's method we have

$$
\begin{aligned}
n_{11} &= 2, n_{12} = 1, n_{13} = 2, \\
n_{21} &= 1, n_{22} = 2, n_{23} = 1, \\
n_{31} &= 1, n_{32} = 1, n_{33} = 1, \\
n_{\cdot 1} &= 4, n_{\cdot 2} = 4, n_{\cdot 3} = 4, \\
\hat{\rho}_1(1) &= 2/4, \hat{\rho}_1(2) = 1/4, \hat{\rho}_1(3) = 2/4, \\
\hat{\rho}_2(1) &= 1/4, \hat{\rho}_2(2) = 2/4, \hat{\rho}_2(3) = 1/4, \\
\hat{\rho}_3(1) &= 1/4, \hat{\rho}_3(2) = 1/4, \hat{\rho}_3(3) = 1/4, \\
\hat{s}_1 &= 0.3750, \hat{s}_2 = 0.3750, \hat{s}_3 = 0.25, \\
\hat{g}_1 &= 0.0013, \hat{g}_2 = 0.0238, \hat{g}_3 = 0.0029, \\
|D| &= 1.5625e-004, \ trD = 0.4267
\end{aligned}
$$

and $L'(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3) = 3096.2$.
For the robust method, we have

$$
\begin{aligned}
n_{11} &= 2, n_{12} = 1, n_{13} = 2, \\
n_{21} &= 1, n_{22} = 1, n_{23} = 2, \\
n_{31} &= 1, n_{32} = 1, n_{33} = 1, \\
n_{\cdot 1} &= 4, n_{\cdot 2} = 3, n_{\cdot 3} = 5, \\
\hat{\rho}_1(1) &= 2/4, \hat{\rho}_1(2) = 1/3, \hat{\rho}_1(3) = 2/3, \\
\hat{\rho}_2(1) &= 1/4, \hat{\rho}_2(2) = 1/3, \hat{\rho}_2(3) = 1/5, \\
\hat{\rho}_3(1) &= 1/4, \hat{\rho}_3(2) = 1/3, \hat{\rho}_3(3) = 1/5, \\
\hat{s}_1 &= 0.3967, \hat{s}_2 = 0.3217, \hat{s}_3 = 0.2817, \\
\hat{g}_1 &= 0.0042, \hat{g}_2 = 0.0053, \hat{g}_3 = 0.0086, \\
|D| &= 1.5973e-004, \ trD = 0.4437
\end{aligned}
$$

and $L'(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3) = 3078.2$.

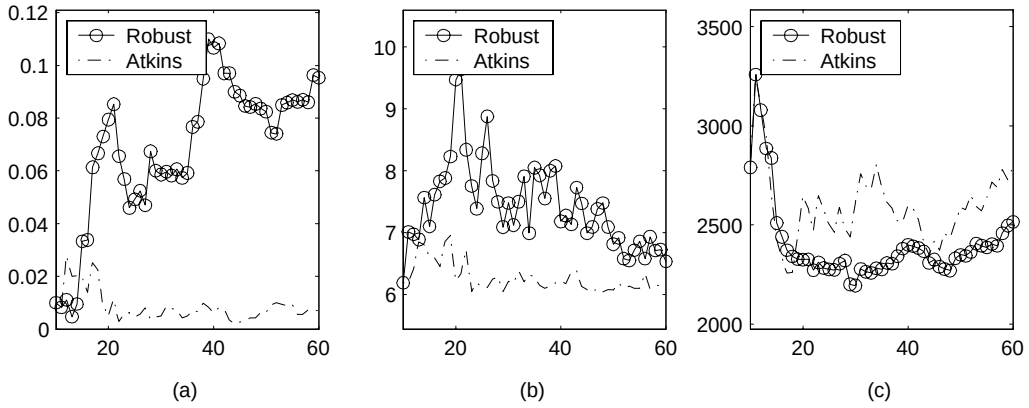

Figure 1: (a) Imbalance, (b) $\sum d_A(i)$ and (c) Loss L' vs. n for the Robust and Atkinson's methods (One run)

3.3 Result of Simulation

Using MATLAB, I generated 50 patients with prognostic factor 'age of the patient' in each simulation. The first ten of them already have been assigned a treatment. Assuming patients 11 to 60 are new arrivals who will come sequentially to our trial, I use the above two different allocation methods to assign the new patient a treatment. After each assignment, I calculated the imbalance and the loss for the trial. Then I repeated this procedure 1000 times. Each time, I get 50 data for each allocation method, which represents the imbalance and the loss after an assignment. I calculated the average for these 50 data and used the result to compare the effectiveness of the different methods.

Figure 1a) shows the imbalance for the two methods where the simulation was run once. The average imbalance for the robust allocation and the Atkinson's methods are 0.06862 and 0.00864 respectively. The range for Robust is from 0.004734 to 0.11 while that of Atkinson's is from 0.002135 to 0.02778. In this plot, we observe that for the first two patients the imbalance was approximately the same. Imbalance for the Atkinson's method was small and more stable as the number of patients increased in the trial. In particular,

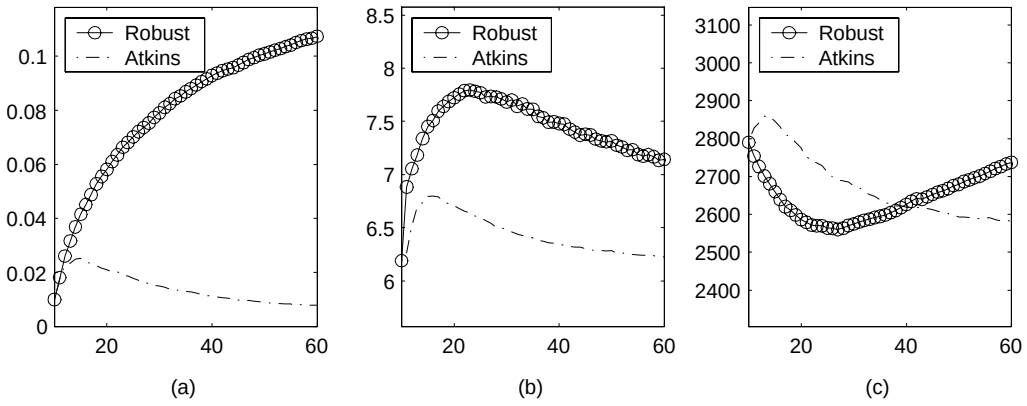

Figure 2: Average over 1000 runs of (a) Imbalance, (b) $\sum d_A(i)$ and (c) Loss L' vs. n for the Robust and Atkinson's methods

from patient 30 to 50, the imbalance between the groups was very small and stable as well. The situation is a bit different for the robust method, there is increase in the imbalance as the number of patients increase. There is a bit of stability from patient 38 to 50. In general, the imbalance for the Atkinson's method is small and more stable than the Robust Method.

Figure 2a) shows the average imbalance for the two methods. The average imbalance for the robust allocation and the Atkinson's methods for this plot are 0.07892 and 0.01422 respectively. The range for Robust is from 0.01 to 0.1075 while that of Atkinson's is from 0.007803 to 0.0252. From this plot, we could see that the averaged imbalance for the Atkinson's method decreases from the 17^{th} patent to 50^{th} patient. It is more stable and does not vary much. In the Robust method, averaged imbalance increases as the number of patient increased.

Figures 1b) and 2b) shows the plot of $\sum d_A(i)$ against the patients being assigned for the single and the 1000 runs respectively. We find that the $\sum d_A(i)$ for Atkinson's D_A-optimum design is small in both plots compared to the robust method. The mean $\sum d_A(i)$ for the Robust and Atkinson's methods are 7.427, 6.267 for Figure 1b) and $7.436, 6.437$ for Figure 2b) respectively. From Figure 2b), $\sum d_A(i)$ for the Atkinson's method decreases from the 15^{th} patient while that of the Robust method starts decreasing from the 25^{th} patient.

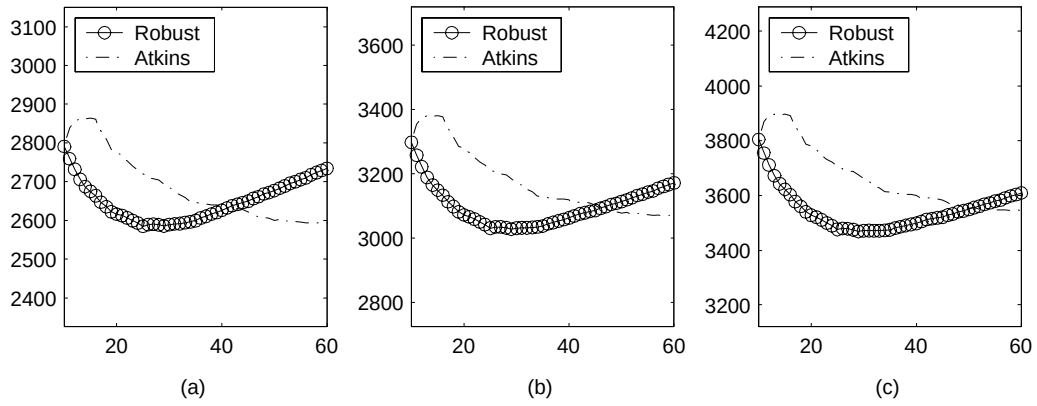

Figure 3: Average Loss L' vs. n for the Robust and Atkinson's methods. (a) $\nu = 1$ (b) $\nu = 1.2$ (c) $\nu = 1.4$.

Figures 1c) and 2c) shows the plot of loss against the patients being assigned for the single and the 1000 runs respectively. From Figure 1c), the average loss for the Robust and Atkinson's methods are 2401, 2585 respectively. The losses for both methods was decreasing until the 20^{th} patient where loss for Atkinson's method increased as the patients increased. From Figure 1c), we observe that the average loss of Atkinson's method increases as the number of patients increases with the Robust loss decreasing as the number of patients increases.

The situation of Figure 2c) is not quite different from that of Figure 1c). The average loss or the Robust and Atkinson's methods are $2642, 2677$ respectively. Robust method ranges from 2560 to 2790, a range of 229.8 while Atkinson's method ranges from 2591 to 2872, a range of 277.9. In general, the Robust method has a more stable loss function compared to Atkinson's method.

Figures 3a), 3b), and 3c) are the loss functions of Atkinson's and robust methods when $v = 1, 1.2, 1.4$ respectively. Surprisingly, the loss of Atkinson's methods increases as v increases with the robust method decreasing with v increasing. All other v values gave the same conclusion.

4 Conclusion

In this project we have found robust assignment probabilities by minimizing the upper bound of the loss derived in Wiens (2000). The assignment probabilities are robust against incorrectly specified regression responses.

We have assessed in a simulation study this robust allocation scheme and that of Atkinson's. In comparing these two methods we found that imbalance for the Atkinson's method were small and more stable as the number of patients arriving increased, while that of the robust method increased as the number of patients increased. Also, the averaged loss for the robust method were more resistant to change than the Atkinson's method.

4.1 REFERENCES

Atkinson, A.C (2002), "The comparison of designs for sequential clinical trials with covariate information," *Journal of the Royal Statistical Society,* 165**,** Part2, pp. 349-373.

Atkinson, A.C (1982), "Optimum Biased Coin Designs for Sequential Clinical Trials With Prognostic Factors," *Biometrika*, 69, 61-67.

Begg, C.B., and Iglewicz, B. (1980), "A Treatment Allocation Procedure for Sequential Clinical Trials," *Biometrics*, 36, 81-90.

Box,G.E.P., and Draper N.R. (1959), "A Basis for the selection of a Response Surface Design," *Journal of the American Statistical Association,* 54, 622-654.

Chalmers, T.C. (1990), "Discussion on Biostatistical collaboration in medical research" (by J.H Ellenberg). *Biometrics,* 46, 20-22.

Efron, B. (1971), "Forcing a Sequential Experiment to be Balanced," *Biometrika*, 58, 403-417.

Heckman, N.E (1987), "Robust Design in Two Treatment Comparison in the Presence of a Covariate," *Journal of Statistical Planning and Inference*, 16, 75-81.

Huber, P.J. (1981), *Robust Statistics.* Wiley, New York.

Pocock, S.J., and Simon, R. (1975), "Sequential Treatment Assignment With Balancing for Prognostic Factors in the Controlled Clinical Trial, " *Biometrics, 31, 103-115.*

Sibson, R. (1974), D_A-optimality and duality. In *Progress in Statistics* (Proc. 9th European Meeting of Statisticians, Budapest, 1972), Ed. J. Gani, K. Sarkadi and I . Vincze, Vol. 2. pp. 677-92. Amsterdam: North Holland.

Student (1931), The Lanarkshire milk experiment. *Biometrika,* **23,** 398-406.

Traves, D.R. (1974), "Minimization: A New Method of Assigning Patients to Treatment and Control Groups," *Clinical Pharmacology and Therapeutics,* 15, 443-453.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.

Wiens, D.P. (2000), "Designs for Clinical Trials," University of Alberta Statistics Centre Technical Report Series #00.06.

Wiens, D.P. (2004), "Robust allocation schemes for clinical trials with prognostic factors," *Journal of Statistical Planning and Inference*, to appear.

X. Wu, (2000), *Comparative Study of Various Allocation Schemes in Clinical Trial*, MSc Project, University of Alberta, Dept. of Mathematical and Statistical Sciences.

Zelen, M. (1974), "The randomization and stratification of patients to clinical trials," *Journal of Chronic Diseases* **27**, 365-375.