CLINICAL SCIENCE

SEQUENTIAL ALLOCATION TO BALANCE PROGNOSTIC FACTORS IN A PSYCHIATRIC CLINICAL TRIAL

Victor Fossaluza,[I,II] Juliana Belo Diniz,[I] Basilio de Bragança Pereira,[III] Eurípedes Constantino Miguel,[I] Carlos Alberto de Bragança Pereira[I,II]

OBJECTIVE: This paper aims to describe and discuss a minimization procedure specifically designed for a clinical trial that evaluates treatment efficacy for OCD patients.

METHOD: Aitchison's compositional distance was used to calculate vectors for each possibility of allocation in a covariate adaptive method. Two different procedures were designed to allocate patients in small blocks or sequentially one-by-one.

RESULTS: We present partial results of this allocation procedure as well as simulated data. In the clinical trial for which this procedure was developed, successful balancing between treatment arms was achieved. Separately, in an exploratory analysis, we found that if the arrival order of patients was altered, most patients were allocated to a different treatment arm than their original assignment.

CONCLUSION: Our results show that the random arrival order of patients determine different assignments and therefore maintains the unpredictability of the allocation method. We conclude that our proposed procedure allows for the use of a large number of prognostic factors in a given allocation decision. Our method seems adequate for the design of the psychiatric trials used as models. Trial registrations are available at clinicaltrials.gov NCT00466609 and NCT00680602.

KEYWORDS: Clinical research; Randomization; Aitchison's compositional distance.

INTRODUCTION

For certain specific psychiatric disorders, large trials are fairly rare. Obsessive-compulsive disorder (OCD), for instance, has only been studied in small trials. First-line treatments such as clomipramine and selective serotonin reuptake inhibitors are typically studied in trials with no more than two hundred patients.[1] For second-line treatments such as pharmacological augmentation strategies, the situation is worse, albeit understandably so.[2,3] To our knowledge, no trials have investigated these strategies using more than one hundred patients. Indeed, most such studies have presented final sample sizes of no more than 15 patients per arm.[3] This is in stark contrast to studies of common clinical diseases such as hypertension and diabetes. Therefore, certain specific aspects of trial design need to be taken into consideration when implementing clinical trials involving psychiatric patients, especially those with OCD.

In small trials, such as those typically used to study psychiatric disorders, an imbalance in prognostic factors between treatment arms can affect the interpretation of the results. One of the first methods created to address this problem was stratified allocation of individuals.[4] The use of this method also reduces the need to adjust statistical data analyses for covariates that serve as prognostic factors. The limitation of stratified allocation is that it can only be applied when a small number of covariates are involved. Otherwise,

[I] Institute of Psychiatry, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo - São Paulo/SP, Brazil.

[II] Institute of Mathematics and Statistics, Universidade de São Paulo - São Paulo/SP, Brazil.

[III] Preventive Medicine Department, Federal University of Rio de Janeiro - Rio de Janeiro/RJ, Brazil.

Email: victor.ime@gmail.com

Tel: 55 11 3069-6972

Received for publication on January 26, 2009

Accepted for publication on March 10, 2009

the sample has to be divided into a large number of strata with only a very few patients in each.[5] In order to overcome this limitation, it is necessary to devise alternative methods of minimizing imbalance among treatment arms.

Pocock and Simon[6] devised a general procedure for treatment assignment that minimizes imbalance among individual prognostic factors. Independently, Taves[7] also developed a minimization procedure that will not be discussed in this article, as it is encompassed by the Pocock and Simon method.

Since 1975, various other authors have developed procedures to minimize the risk of imbalance between arms, as reviewed elsewhere.[8,9] Each method has specific limitations and complexities. For example, the optimal allocation technique allows the use of continuous variables in their non-categorical format as covariates.[10] However, the complexity and associated difficulty has limited its practical implementation. No allocation procedure is ideal - accordingly, novel methods are still required in specific situations.

In response to the need for specific minimization procedures in clinical trials involving individuals with OCD, we developed an allocation method that uses the Aitchison distance,[11-13] which is a measure for compositional data (data that contains quantitative descriptions of the parts of some whole). The Aitchison distance was chosen because all prognostic factors used are better classified as categorical data with relative frequencies that can be taken as compositional data in nature.

We present allocation results for a partial sample collected during the first year of the study for which this procedure was developed, as well as results for simulated samples that feature variable patient's order of inclusion.

METHOD

The trials reported in this manuscript received prior approval from our local ethics committee.

Allocation

For clarity, let us first consider the simplest design. Consider a clinical trial in which patients are enrolled sequentially, according to the order in which they commence treatment at the clinic. Each patient is to be assigned to one, and only one, of two alternative treatments. Imagine that a new patient arrives after the study already included a considerable number of patients in each of the two arms: n_1 and n_2. In addition, consider that age, denoted a, is a factor for which we think adjustment should be made. Possible ages are divided into three different categories: a_1 if $a \leq$ 30; a_2 if $30 < a \leq 45$; and a_3 if $a > 45$. Consider now the following notation:

1. Represent the sample absolute frequencies of n_1 and n_2, respectively, as (n_{11}, n_{12}, n_{13}) and (n_{21}, n_{22}, n_{23}), i.e., for $i = 1$ or 2 and $j = 1$, 2 or 3, n_{ij} is the number of patients of age group a_j in arm i.

2. Represent the compositional vectors (relative frequency vectors), respectively, as $A_1 = (a_{11}, a_{12}, a_{13})$ and $A_2 = (a_{21}, a_{22}, a_{23})$. Thus, $a_{ij} = n_{ij}/n_i$ represents the relative frequency of patients in age group a_j in arm i given the proportion of age group a_j in arm i.

The allocation of a new incoming patient consists of the following steps:

i. Consider including the new patient in the first arm, in which case we represent the new composition vector by $(A_1)^*$, and we calculate the distance between $(A_1)^*$ and A_2 by evaluating $d_1 = \Delta[(A_1)^*; A_2]$.

ii. Consider including the new patient in the second arm, in which case we represent the new compositional vector by $(A_2)^*$, and compute the distance $d_2 = \Delta[A_1; (A_2)^*]$.

iii. Compare the values of d_1 and d_2. If $d_1 < d_2$, we allocate the new patient to n_1; if $d_1 > d_2$, we allocate the patient to n_2; and if $d_1 = d_2$, the patient may be allocated to either arm (at random).

Note that we make our decision based on the values of the distances between the compositional vectors. We use the term "compositional vectors" to designate vectors for which the sum of their components is fixed and known. In our case, with relative frequencies, the values of the components are numbers in the interval [0;1], and the sum of these components is one. Aitchison[11,12] argued that, for compositional data, the correct distance measure is not the usual standard Euclidian measure but is instead as described below.

In general, consider factor a having k (>1) possible alternatives, i.e., $A_i = (a_{i1}, a_{i2}, …, a_{ik})$ is the relative frequency vector of arm i (=1,2). For category j, consider the natural logarithm of the between-arm relative frequency ratio, $r_j = (a_{1j} \div a_{2j})$, denoted by $ln(r_j)$, and the mean of these logarithms, denoted by L:

$$ln(r_j) = ln(a_{1j}) - ln(a_{2j}) \text{ and } L = \{ln(r_1) + ln(r_2) + … + ln(r_k)\} \div k$$

The Aitchison distance measure between the two compositional vectors A_1 and A_2 is defined as follows:

$$\Delta[A_1, A_2] = \sqrt{\sum_{j=1}^{k} \{ln(r_j) - L\}^2}$$

To demonstrate the allocation procedure, let us consider the case of the three age categories a_1, a_2, and a_3, as before.

Suppose that, in one stage of the process, we encountered the following vectors of absolute frequencies: $n_1 = (3;7;5)$ and $n_2 = (5;6;6)$. A new patient enrolls and falls under age category a_2. Table 1 presents the vectors used to calculate the required distances.

Table 1 - Vectors of relative frequencies

Aging	A_1	A_2	A_1*	A_2*
a_1	0.2000	0.2941	0.1875	0.2778
a_2	0.4667	0.3529	0.5000	0.3889
a_3	0.3333	0.3529	0.3125	0.3333
Sum	1	1	1	1

The following three expressions represent the intragroup distances, (not) including the new patient:
A) Before the arrival of the new patient, $\Delta[A_1;A_2] = 0.4702$
B) With the new patient in n_1, $\Delta[(A_1)*;A_2] = 0.5676$
C) With the new patient in n_2, $\Delta[A_1;(A_2)*] = 0.3661$

The figures above indicate that a new patient in age group a_2 should best be allocated to n_2. This conclusion might appear obvious to the decision-maker, as it is desirable to increase the frequency of a_2 in n_2. Note that this choice increases the n_2 sample size. We are also interested to control the arm sample sizes. Let us transform the sample size (s) into a compositional factor. Let us consider the vectors $S_1 = (s_1;s_2) \div s$ and $S_2 = (s_2;s_1) \div s$ as the relative frequencies of n_1 and n_2, respectively. In our example we would have the following:

$S_1 = (15;17) \div 32 = (.47;.53)$, $S_2 = (.53;.47)$, $(S_1)* = (.48;.52)$ and $(S_2)* = (.55;.45)$

Here the * indicates that the new patient entered a specific arm. The new distance values are $\Delta[S_1;S_2] = 0.1770$, $\Delta[(S_1)*;S_2] = 0.1314$, and $\Delta[S_1;(S_2)*] = 0.2174$. Clearly, the best allocation would be to the first arm, which features a smaller sample size. Bear in mind that, given the age categories above, our best choice would have been to allocate the new patient to the second arm. However, considering the sample size factor, the first arm is the more appropriate allocation for the new patient. If the decision-maker feels that age is more important than sample size, a larger weight could be assigned to age than to sample size. For example, consider our decision to use a weight of 2 for age and a weight of 1 for sample size. The weights reflect the importance of covariates and should be prospectively made in agreement with the investigators. We may use the weighted average of the distances to guide this decision. The results for our example are as follows:

i. new patient in n_1 produces $D_1 = \{2\Delta[(A_1)*;A_2] + \Delta[(S_1)*;S_2]\} \div 3 = 0.4222$
ii. new patient in n_2 produces $D_2 = \{2\Delta[A_1;(A_2)*] + \Delta[S_1;(S_2)*]\} \div 3 = 0.3165$

The overall distance using the two factors indicates that the new patient should be allocated to n_2 in order to bring the arms closer in terms of age and sample size. It may also be possible to assign equal weights to the two factors. Regarding the constructed factor for sample size, to appropriately compare factors, we need to adopt the same range for all factor distance measures. Hence, the definitions of S_1 and S_2 seem quite adequate.

The examples discussed below are more specific and demonstrate several interesting particularities. The first, related to phase one of the trial described, deals with two arms, four prognostic factors, and a simultaneous allocation of three patients each time. The second example, related to phase two, deals with seven prognostic factors and allocation of patients one at a time into the three groups.

Two-phase clinical trial

A study group specializing in OCD wished to conduct a clinical trial consisting of two phases. The objective of phase one was to compare patient responses to pharmacological and psychotherapeutic treatments. They expected that 360 OCD patients would be enrolled in phase one over a 3-year period. However, for logistical reasons, they did not anticipate that the numbers of patients allocated to the individual treatment arms would be similar in phase one. They anticipated that the pharmacological arm may receive more patients than the psychotherapeutic arm at certain times, and vice-versa at other times. These variations are required because, for practical reasons, each therapeutic arm must accommodate a different number of patients at different times in the study. As the psychotherapeutic arm involves group psychotherapy, when groups are first created, a rapid influx of patients is necessary, but when groups are well-established, the speed of inclusion should diminish as only two groups can be conducted simultaneously.

From previous experience, we expected that 60% of the patients treated in either of the two arms of phase one would report less than adequate symptom improvement. Patients who participated in the pharmacological arm (n_1) and did not respond to treatment would be invited to participate in phase two. Patients who were non-responders in the psychotherapeutic arm (n_2) of phase one would be invited to participate in the pharmacological arm (n_1), and, if treatment resistance were to persist in n_1, they would also be invited to participate in phase two. Phase

two consists of three arms, the objective being to compare three different pharmacological augmentation strategies. In phase two, unlike in phase one, it is expected that a similar number of patients will be allocated to each of the three arms. Considering expected response rates, drop-out rates, and frequency of refusals to participate in a study using a placebo, it is likely that 30 to 40 patients will be included in each arm. This sample size seems adequate according to the investigator's hypothesis. The entire clinical trial design is illustrated in Figure 1.

The allocation strategies for the two phases of the trial are different. To accommodate the logistical characteristics of phase one (that require a different pace of patient enrollment for each arm at different study time points), the researcher will receive groups of 3 patients to be allocated into the two groups. On some occasions, it might be necessary to allocate 2 patients to n_1 and 1 to n_2. On other, well-defined occasions, the situation is reversed, and n_1 will receive 1 patient while the other 2 patients will be allocated to n_2.

Prognostic factors to be balanced across arms

When the allocation program was developed for phase one, a smaller number of variables was chosen, as we did not know how many covariates this new procedure could accommodate without compromising its efficiency in minimizing arm imbalance. After the initial results of allocation were analyzed for phase one, it became clear that the procedure could accommodate a greater number of covariates. Therefore, the program for phase two was designed with a greater number of hypothesized prognostic factors that could also have been covariates in phase one. The factors used to establish the allocating strategy were chosen based on previous studies. Although the appropriateness of the factors used might be obvious, discussion of these choices is outside the scope of this paper. Our sole intention is to present the allocation strategy used in the study described. The factors and their categories were as follows:

1. Current age (age) was categorized into three classes: under 30 years of age (a_0); between 31 and 45 (a_1); and over 45 (a_2).
2. OCD symptom severity (OCD), as measured using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS), was categorized under one of eight classes based on ordered scores for each of the two symptom types (obsession and compulsion): ≤ 8 = low (L); 8-15 = moderate (M); and > 15 = high (H). Therefore, the eight possible classes were as follows: ($L;L$), ($L;M$), ($L;H$), ($M;L$), ($M;M$), ($M;H$), ($H;L$), ($H;M$), and ($H;H$). The first and second notations are based on the scores for obsession and compulsion, respectively. Clearly, ($L;L$) is not considered, as it would

indicate no OCD symptoms, and we had no patients in this class.
3. Treatment history (his) was divided into three categories in each phase. In phase one, h_0 indicates no previous appropriate treatment, h_1 indicates one previous course of appropriate treatment without response, and h_2 indicates two or more previous courses of appropriate treatment without response. In phase two, h_0 indicates no Y-BOCS score reduction or Y-BOCS score increase after n_1 in phase one, h_1 indicates a 1-20% reduction in Y-BOCS score after n_1 in phase one, and h_2 indicates a 20-35% reduction in Y-BOCS score after n_1 in phase one.
4. Level of education (sch) was divided into four categories: sc_0 indicates no schooling, sc_1 indicates ≤ 8 years of schooling, sc_2 indicates 9-12 years of schooling, and sc_3 indicates higher education (undergraduate or graduate work).
5. Marital status (mar) was categorized as si_0 (married, divorced or widowed) or si_1 (single, never married).
6. Genders (gen) are indicated by m for male and f for female.
7. Sample sizes (sam) are denoted as s_1, s_2, s_3, and $s = s_1 + s_2 + s_3$ represents the total sample size. Note that there are three arms only in phase two. The compositional vector of sample size in arm i ($i = 1,2,3$) is the vector (p_i;1-p_i) where $p_i = s_i \div s$.

For each of these factors, the distance between arms can be computed. The distance values for the seven variables listed above are given by the following symbols: Δ_{age}, Δ_{OCD}, Δ_{his}, Δ_{sch}, Δ_{mar}, Δ_{gen}, and Δ_{sam}. The global distances for phases one and two are, respectively, Δ_1 and Δ_2, defined as follows:

$$\Delta_1 = (2\Delta_{age} + 3\Delta_{OCD} + 3\Delta_{his} + \Delta_{gen}) \div 9$$
$$\Delta_2 = (2\Delta_{age} + 4\Delta_{OCD} + 5\Delta_{his} + 2\Delta_{sch} + 3\Delta_{mar} + \Delta_{gen} + 4\Delta_{sam}) \div 21$$

As there are three arms in phase two, there are also three vectors. Hence the global distance for phase two should be the average of the three values of Δ_2, i.e., the distance for phase two should be as follows:

$$\Delta_2 = \{\Delta_2(n_1;n_2) + \Delta_2(n_1;n_3) + \Delta_2(n_2;n_3)\} \div 3$$

Phase one allocation strategy

Based on previous studies, factors measurable before treatment initiation may influence the treatment response. Therefore, it is desirable that certain prognostic factors be distributed as homogeneously as possible between the two arms. We assume that previous treatment response (assessed through patient interviews) and severity of the

disorder (assessed using a specific scale) at the time of inclusion in the study are the most important prognostic factors for clinical response. In addition, although of lesser importance to treatment response, gender and age should also be homogeneously distributed between arms. For logistical reasons, it was not possible to enroll the same number of patients in each phase one treatment arm. The healthcare professionals who administer the treatments can assist more patients in the pharmacological arm than in the psychotherapeutic arm. Over the course of the study, there are times at which the pace of patient enrollment into the psychotherapeutic arm is, of necessity, more rapid than that of patient enrollment into the pharmacological arm. We expect that the pharmacological arm will account for a higher percentage of the final sample. Due to the time needed to perform the clinical evaluations, there is a two-week gap between inclusion and allocation. Consequently, it is possible to allocate patients simultaneously in groups of three, one to one arm and the remaining two to the other arm (the arm that receives two patients at certain time points will only receive one at other time points). This simultaneous inclusion of patients guarantees that provider blinding is not compromised by allocation in small blocks.

The choice of prognostic factors to be balanced between arms is based on reports in the literature that suggest that these prognostic factors may influence the results considerably. In addition, having information about these factors at the time of the initial evaluation is feasible and provides useful clues for balancing between intervention groups. As previous treatment response and initial severity are known to be associated with treatment response in clinical trials evaluating OCD patients,[14-18] they were included as prognostic factors in order to balance the arms.

Phase two allocation strategy

Patients entering phase two were those that failed to completely respond to the treatment in n_1 of phase one (regardless of treatment history). The investigators assumed that treatment response and current severity were highly predictive of phase two treatment response, and that these patients should therefore be distributed homogeneously among the three groups. In addition, it is understood from previous studies that level of education and marital status are significantly associated with poor treatment response and should be included in the phase two strategy model. Gender and current age should also be distributed homogeneously among the three arms. Another important factor is sample size, which should be as similar as possible among the three arms.

For each patient in phase two, the allocation to one of the three arms needs to be determined during the evaluation conducted at week 12 of treatment. The strategy at this point involves a sequential, one-by-one allocation of patients who had participated in phase one. Here, in addition to having more than two factors to consider in our attempts to balance the arms, we are dealing with three arms. Bear in mind that though the distance measures are defined for two vectors in order to facilitate decisions regarding patient allocation, we consider the average of the three distances between pairs to represent the overall distance of the three arms. The proposed solutions for the two phases are described below.

RESULTS

Partial results for phase one

To perform stratified allocation of patients, a Microsoft Excel macro was created. This macro divided patients into homogenous intervention subgroups consistent with the factors chosen by investigators. To avoid empty categories in the allocation process, for a factor with k (>1) categories, we added to each category the fraction $1/k$. Therefore, no category had a frequency of 0 at the time of Δ distance calculation.

The partition that provides the best degree of homogeneity between groups is defined as the division that minimized the difference between factor vectors of categorized relative frequencies. To calculate this measure, the Aitchison distance between two vectors was used, as previously discussed. This distance was chosen based on the assumption that it preserves the sub-compositional coherence of the simplex space. After the values of Δ_1 were calculated for each possible allocation, the allocation that resulted in the least intragroup distance was chosen as the optimal allocation.

At the time of writing, 152 individuals had been allocated to n_1 and 107 had been allocated to n_2. The partial allocation results are shown in Table 2. The results of the reverse order allocation of patients appear in parentheses.

Simulated results for phase two

The phase two allocation system has the same objectives as that of phase one, namely to bring the arms as close together as possible in terms of the relevant prognostic factors. However, in phase two, the allocation logistics are different than those in phase one. We now have three arms, and the patients are allocated sequentially one-by-one. In addition, the balancing factors in this phase are different from those considered in phase one. As well as now accounting for level of education, marital status, and sample size, the previous treatment history factor, unlike in phase

Table 2 - Partial (reverse order) allocation of 259 patients in phase one

Prognostic factor	Compositional category	Arm 1 frequencies		Arm 2 frequencies	
		Absolute	Relative	Absolute	Relative
Current age	a_0	71 (71)	0.467	49 (49)	0.458
	a_1	53 (52)	0.349	38 (38)	0.355
	a_2	28 (29)	0.184	20 (20)	0.187
OCD symptom severity	(L;M)	2 (2)	0.013	1 (1)	0.009
	(M;L)	2 (2)	0.013	1 (1)	0.009
	(L;H)	3 (3)	0.020	2 (2)	0.019
	(H;L)	3 (2)	0.020	0 (1)	0.000
	(M;M)	59 (59)	0.388	42 (42)	0.393
	(M;H)	19 (18)	0.125	12 (13)	0.112
	(H;M)	17 (16)	0.112	11 (12)	0.103
	(H;H)	47 (50)	0.309	38 (35)	0.355
Treatment history	h_0	78 (82)	0.513	61 (57)	0.570
	h_1	49 (44)	0.322	30 (35)	0.280
	h_2	25 (26)	0.165	16 (15)	0.150
Gender	m	63 (63)	0.414	44 (44)	0.411
	f	89 (89)	0.586	63 (63)	0.589
Sample size	Arm size	152	0.587	107	0.413
	Complement	107	0.413	152	0.587

Table 3 - Sequential (reverse order) allocation of 90 patients in phase two: patients from phase one

Prognostic factor	Compositional Category	Arm 1 frequencies		Arm 2 frequencies		Arm 3 frequencies	
		Absolute	Relative	Absolute	Relative	Absolute	Relative
Current age	a_0	13 (14)	0.433	13 (12)	0.433	14 (14)	0.467
	a_1	13 (12)	0.433	12 (12)	0.400	11 (12)	0.367
	a2	4 (4)	0.133	5 (5)	0.167	5 (5)	0.167
OCD symptom severity	(L;M)	0 (1)	0.000	0 (0)	0.000	1 (0)	0.033
	(M;L)	0 (0)	0.000	0 (0)	0.000	0 (0)	0.000
	(L;H)	0 (0)	0.000	0 (0)	0.000	0 (0)	0.000
	(H;L)	1 (1)	0.033	0 (1)	0.000	1 (0)	0.033
	(M;M)	12 (11)	0.400	14 (12)	0.467	10 (13)	0.333
	(M;H)	2 (2)	0.067	2 (2)	0.067	3 (3)	0.100
	(H;M)	4 (3)	0.133	3 (4)	0.100	4 (4)	0.133
	(H;H)	11 (12)	0.367	11 (10)	0.367	11 (11)	0.367
Treatment history	h_0	15 (15)	0.500	15 (15)	0.500	15 (15)	0.500
	h_1	9 (10)	0.300	10 (9)	0.333	10 (10)	0.333
	h_2	6 (5)	0.200	5 (5)	0.167	5 (6)	0.167
Level of education	sc_0	14 (15)	0.467	15 (14)	0.500	15 (15)	0.500
	sc_1	0 (0)	0.000	0 (0)	0.000	0 (0)	0.000
	sc_2	0 (0)	0.000	0 (0)	0.000	0 (0)	0.000
	sc_3	16 (15)	0.533	15 (15)	0.500	15 (16)	0.500
Marital status	si_0	16 (15)	0.533	15 (15)	0.500	15 (16)	0.500
	si_1	14 (15)	0.467	15 (14)	0.500	15 (15)	0.500
Gender	m	13 (13)	0.433	13 (12)	0.433	13 (14)	0.433
	f	17 (17)	0.567	17 (17)	0.567	17 (17)	0.567
Sample size	Arm size	30 (30)	0.333	30 (29)	0.333	30 (31)	0.333
	Complement	60 (60)	0.667	60 (61)	0.667	60 (59)	0.667

one, is based on the relative reduction, during phase one, in the Y-BOCS score. Patients who present a > 35% reduction in Y-BOCS score are not included in this phase. In order to consider sample size as a factor, the proportion of patients in each group will also be included in the distance calculation. Table 3 presents the allocation results for 90 patients selected from those in n_1 of phase one.

DISCUSSION

The objective of using this strategy of random allocation of patients in different treatment arms is to avoid intentional bias in group allocation. However, treatment group imbalance for prognostic factors can still occur.[9] These unfortunate consequences of randomization can only be prevented by using techniques that produce an intentional optimal allocation of patients to each treatment arm.

Another interesting point is the choice of the Aitchison distance to treat compositional data. To illustrate the difference between the Aitchison distance and the standard Euclidian distance, we calculate the two distances for the pairs $(A_1;A_2)$ and $(B_1;B_2)$ of relative frequency vectors defined below:

$A_1 = (0.1;0.2;0.7)$, $A_2 = (0.2;0.1;0.7)$, $B_1 = (0.2;0.4;0.4)$, and $B_2 = (0.4;0.2;0.4)$

The standard Euclidian measure would produce $D[A_1;A_2]$ = 0.1414 and $D[B_1;B_2]$ = 0.2828. Note that the Aitchison distance will produce the same value for both pairs: $\Delta[A_1;A_2]$ = $\Delta[B_1;B_2]$ = 0.9803, as expected for an adequate distance. Note that, for any category j (= 1, 2, 3), the ratio between frequencies is the same for the two pairs:

$a_{1j} \div a_{2j} = b_{1j} \div b_{2j}$

We also note that problems occur if a frequency of 0 is listed for a particular class. To address this, we recommend that prior correction be used, commencing the allocation process at a small real number for each category that should be added to the observed absolute frequency. As an analogy to Bayesian categorical data analysis, we adopt an equally constant for each category, as with the standard strategy. Our recommendation is to add $1/k$ to the absolute frequencies for each category, thereby avoiding the occurrence of a relative frequency of 0 in any category.

The "random" aspect of treatment arm allocation is preserved due to the random sequence in which patients arrive for inclusion in clinical trials. The most important property of the allocation strategy defined here is that we can verify *post hoc* adherence to the desired strategy. The use

of a purely randomized strategy cannot be proven after the trial has been concluded. To verify that the order in which the patients entered the experiment was indeed random, we considered the reverse order of the first 259 patients (phase one) and 90 patients (phase two). In Tables 2 and 3, it can be seen how close the results from the reverse order are to those obtained for the original order. An interesting fact is illustrated in Tables 4 and 5. We can see that more than half of the patients would change arms by reversing the order in which they enter into a phase. In phase one, 130 of the 259 patients would have changed arms if the reverse order had been considered. For phase two, under the reverse order, 48 of the 90 patients would have changed arms.
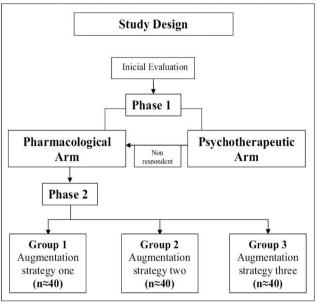
Table 4 - Phase one patient allocation

Original order	Reverse order		Sample Size
	arm 1	arm 2	
arm 1	87	65	152
arm 2	65	42	107
Sample size	152	107	259

Table 5 - Phase two patient allocation

Original order	Reverse order			Sample Size
	arm 1	arm 2	arm 3	
arm 1	15	9	6	30
arm 2	10	11	9	30
arm 3	5	9	16	30
Sample size	30	29	31	90

This article describes a method of computer-based intentional allocation that has the advantage of being flexible, effective, and feasible for the allocation of each consecutive patient, or groups of three patients, in terms of multiple covariates. Future studies are still needed to evaluate the effectiveness of this procedure when a large number of covariates are included. The procedure presented here is reliable for the inclusion of up to six compositional covariates with twenty-one possible alternative results. When using this procedure, researchers could include any allocation strategy variables suspected of influencing treatment response.

One limitation of the proposed procedure is the need to know the value of each prognostic factor, for each patient entering the study, prior to allocation. Consequently, values of prognostic factors that can only be obtained through long-term evaluation cannot be included in the model. Another issue is the weights given to the allocation factors. For the trial in question, the second author (the psychiatrist responsible for conducting the trial) based the choice of variable weights on the evidence levels of the prognostic

factors that influence treatment response. Although this was an arbitrary decision, it does not favor any allocation tendency that might result in arm imbalance. Thus, there is no imbalance that might favor a particular result in terms of treatment response.

Figure 1 - Clinical study design

CONCLUSION

In conclusion, the procedure described here can be easily adapted to different study designs. The results of the trial, used as illustration, show that the methodology was reliable for the sample tested. There are, however, some reliability issues that require further investigation, such as the inclusion of cases presenting a large number of factors or factors that feature a large number of possible alternative categories. Nevertheless, we believe that this paper presents a procedure that will be useful in avoiding imbalance among treatment arms when patient allocation is performed sequentially.

ACKNOWLEDGEMENTS

REFERENCES

1. Eddy KT, Dutra L, Bradley R, Westen D. A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. Clin Psychol Rev. 2004;241011-30.
2. Bloch MH, Landeros-Weisenberger A, Kelmendi B, Coric V, Bracken MB, Leckman JF. A systematic review: antipsychotic augmentation with treatment refractory obsessive-compulsive disorder. Mol Psychiatry. 2006;11:622-32.
3. Fontenelle LF, Nascimento AL, Mendlowicz MV, Shavitt RG, Versiani M. An update on the pharmacological treatment of obsessive-compulsive disorder. Expert Opin Pharmacother. 2007;8:563-83.
4. Lachin JM. Statistical properties of randomization in clinical trials. Controlled Clin Trials. 1988;9:289-311.
5. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz R. Stratified randomization for clinical trials. J Clin Epidemiol. 1999;52:19-26.
6. Pocock SJ, Simon R. Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. Biometrics. 1975;31:103-15.
7. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther. 1974;15:443-53.
8. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. Controlled Clinical Trials. 2002;23:662-74.
9. Pocock SJ, Assmann SE, Enos LE, Kasten, LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statist Med. 2002;21:2917-30.
10. Atkinson AC. Optimum Biased Coin Designs for Sequential Clinical-Trials with Prognostic Factors. Biometrika. 1982;69:61-7.
11. Aitchison J. Principal component analysis of compositional data. Biometrika. 1983;70:57-65.
12. Aitchison J. The Statistical Analysis of Compositional Data. London: Chapman & Hall;1986.
13. Aitchison J, Egozcue J. Compositional Data Analysis: Where Are We and Where Should We Be Heading. Mathematical Geology. 2005;37:829-50.
14. Alarcon RD, Libb JW, Spitler D. A predictive study of obsessive-compulsive disorder response to clomipramine. J Clin Psychopharmacol. 1993;13:210-3.
15. De Haan E, Van Oppen P, Van Balkom AJ, Spinhoven P, Hooguin L, Van Dyck R. Prediction of outcome and early versus late improvement in OCD patients treated with cognitive-behaviour therapy and pharmacotherapy. Acta Psychiatr Scand. 1987;96:354-61.
16. Mataix-Cols D, Rauch S, Manzo PA, Jenike MA, Bayer L. Use of factor-analyzed symptom dimensions to predict outcome with serotonin reuptake inhibitors and placebo in the treatment of obsessive-compulsive disorder. Am J Psychiatry. 1999;156:1409-16.
17. Steketee G, Eisen J, Dyck I, Warshaw M, Rasmussen S. Predictors of course in Obsessive-Compulsive Disorder. Psychiatry Res. 1999;89:229-38.
18. Thomsen PH, Mikkelsen HU. Course of obsessive-compulsive disorder in children and adolescents: a prospective follow-up study of 23 Danish cases. J Am Acad Child Adolesc Psychiatry. 1995;34:1432-40.