ORIGINAL ARTICLE

Randomization in clinical trials: stratification or minimization? The HERMES free simulation software

Hélène Fron Chabouis · Francis Chabouis ·
Florence Gillaizeau · Pierre Durieux · Gilles Chatellier ·
N. Dorin Ruse · Jean-Pierre Attal

Abstract

Objectives Operative clinical trials are often small and open-label. Randomization is therefore very important. Stratification and minimization are two randomization options in such trials. The first aim of this study was to compare stratification and minimization in terms of predictability and balance in order to help investigators choose the most appropriate allocation method. Our second aim was to evaluate the influence of various parameters on the performance of these techniques.

Materials and methods The created software generated patients according to chosen trial parameters (e.g., number of important prognostic factors, number of operators or centers, etc.) and computed predictability and balance indicators for several stratification and minimization methods over a given number of simulations. Block size and proportion of random allocations could be chosen. A reference trial was chosen (50 patients, 1 prognostic factor, and 2 operators) and eight other trials derived from this reference trial were modeled. Predictability and balance indicators were calculated from 10,000 simulations per trial.

Results Minimization performed better with complex trials (e.g., smaller sample size, increasing number of prognostic factors, and operators); stratification imbalance increased when the number of strata increased. An inverse correlation between imbalance and predictability was observed.

Conclusions A compromise between predictability and imbalance still has to be found by the investigator but our software (HERMES) gives concrete reasons for choosing between stratification and minimization; it can be downloaded free of charge.

Clinical relevance This software will help investigators choose the appropriate randomization method in future two-arm trials.

Keywords Random allocation · Minimization · Stratified randomization · Randomized controlled trials · Simulations · Predictability

H. Fron Chabouis (✉) · J.-P. Attal
Sorbonne Paris Cité, Faculté de chirurgie dentaire,
Biomaterials department (URB2i, EA4462), Clinical Research
Unit, Université Paris Descartes, 1 rue Maurice Arnoux,
Montrouge 92120, France
e-mail: helene.fron@parisdescartes.fr

F. Chabouis
Master MASEF, Université Paris Dauphine,
Place du Maréchal de Lattre de Tassigny,
Paris 75775, France

H. Fron Chabouis · J.-P. Attal
Service d'odontologie, APHP, Hopital Charles Foix,
Ivry-sur-Seine 94200, France

H. Fron Chabouis
Sorbonne Paris Cité, Ecole doctorale Galillée, Université Paris 13,
Villetaneuse 93430, France

F. Gillaizeau · P. Durieux · G. Chatellier
APHP, HEGP; INSERM, UMR S872/20, Paris 75015, France

N. D. Ruse
Biomaterials, University of British Columbia, Vancouver, Canada

Introduction

A good randomized controlled trial is needed to support the correct prescription of a treatment or a medical procedure to a patient [1]. However, in operative dentistry, as in many other surgical specialties like orthopedics, many technical difficulties complicate the design and conduct of clinical trials [2, 3]. Operative protocols are often long, and therefore expensive, making it difficult to include many patients; operative parameters (such as the instruments used, the

operator himself, etc.) can vary, and it is difficult to carry out a double-blind trial, as the operator generally cannot be blinded during the procedure.

When a clinical trial cannot be blinded, the protocol and randomization have to be faultless and well documented [4]. Randomization should be carried out as late as possible so that treatment knowledge does not influence the operator's actions upstream [5]. It should take into account low recruitment, absence of blinding, and operative parameters. When there is limited recruitment, the first aim of the randomization is to obtain a good balance between the treatment or procedure groups being compared [6] in order to optimize estimation of treatment effect and power. Secondly, to overcome absence of blinding, the allocations should not be predictable [7]. Lastly, the parameters which may influence the treatment effect estimate (which are also called "prognostic factors") should be considered. Only a limited number of main prognostic factors should be accounted for at the time of randomization [8–13]. If these factors are well distributed between groups, it is possible to attribute the effect observed to the evaluated treatment or procedure rather than to these factors [14, 15]. Taking the main prognostic factors into account is especially important when few patients are included (<200 patients per trial arm) [16, 17], when the trial is open-label, when subgroup or intermediate analyses are planned, or when the trial is aiming to demonstrate equivalence [18–20].

Most dental trials fall within the above-mentioned types of trials and therefore, stratified blocked randomization or minimization should be implemented; these are the two traditional techniques which achieve the randomization aims described above [21]. Other techniques have been described, but they are more sophisticated and thus more difficult to implement (e.g., Efron's biased coin design [22], Wei's urn design [23], Soares and Wu's big stick design [24], Signorini's dynamic balanced allocation [25]).

Stratified blocked randomization consists of generating blocks of treatment allocation (e.g., a block of 4: "ABBA", meaning the first patient receives treatment A, the second treatment B, etc.). Blocks can be of varying size, but one block contains an equal number of treatments A and B in order to achieve balance between groups. The order of treatments within a block is randomly generated. A randomization list (which is a block series) is generated for each stratum of patients which contains patients whose prognostic factors are all identical [21]. Stratification is the procedure recommended by regulatory bodies [15]. It works quite well when only a few prognostic factors must be taken into account.

Minimization is a dynamic method that minimizes the imbalance between the number of patients in each treatment group over a number of prognostic factors [26]. The treatment allocated to the next participant depends on the characteristics of the patients already enrolled [5]. Minimization can take into account many factors [27], but it is not recommended by the regulatory bodies. The first method described (by Taves [28] and by Pocock and Simon [29]) was completely deterministic. Thus, the method was considered to be a potential source of bias since operators might be able to predict all the allocations, especially in monocenter and industry-supported trials. Although some authors have proposed introducing some randomness to the minimization [29, 30], authorities still require the use of this method to be justified [10, 15].

Different suggestions have been put forward to help in the effective choice between the two methods. For minimization, the proportion of random allocations recommended ranges from 5 to 30 % depending on the author or the recommendation [30, 31]. It has been generally considered that minimization could easily deal with 10 prognostic factors [27] and Rovers et al. stated that the expected number of patients in each subcategory should be greater than five, to prevent empty cells [17]. For stratification, two recommendations have defined the maximum number of strata: Therneau suggests that the number of strata be less than half the sample size [27], while Kernan et al. suggests keeping the number of strata S below $N/(B \times 4)$, where N is the sample size and B is the block size [20]. However, the choice of the block size and the proportion of random allocations included depend on other clinical trial parameters. Statisticians have therefore suggested using computer simulations to choose the allocation method best suited for the proposed trial [32, 33].

Some authors have used simulated data to compare both methods in terms of balance between treatment arms [27, 29], balance within a factor and within strata [25, 34], and statistical analysis (e.g., performance of conventional model-based statistical inference [33], estimated treatment effect, size of the rejection region, and power [35]). Recently, Zhao et al. compared many designs in terms of imbalance and correct guess probability, but stratification and minimization were not among the designs compared [36]. Real clinical data have also been used to conduct a posteriori simulations and compare randomization methods in terms of balance [37], statistical power [32], and nominal significance level [38]. Brown et al. compared deterministic minimization to minimization incorporating various random elements in terms of prediction rates and balance [30]. However, comparisons between stratified blocked randomization and minimization in terms of predictability and balance are missing.

The aim of this study and of the HERMES software we created was to compare stratification methods (with various block sizes) and minimization methods (including more or less randomness), to analyze the effect of various parameters of clinical trials on this choice (sample size, number of prognostic factors, and operators), and to therefore, provide guidance for future investigators. To compare the methods, we computed an indicator of balance between treatments and an indicator of predictability for operators. Specifically,

we wanted to choose the optimal method to randomize 358 patients between two groups in a trial comparing ceramic and composite to make inlays/onlays (CECOIA trial, NCT01724827, www.cecoia.fr), while taking into account four prognostic factors (pulp vitality, inlay or onlay, premolar or molar, and operator). We wrote an initial computer simulation program that answered this question. We then wrote a second program (HERMES)[1] that can be applied to other studies, in order to help future clinical trial investigators choose the most suitable randomization method.

Materials and methods

A computer simulation program was coded with Visual Basic for Applications and Excel software. It allocated one of two treatments to patients simulated according to their expected characteristics. Various randomization methods could then be compared in terms of balance between groups and predictability for the operator.

The "Simulation" tab of the Excel interface allowed us to enter the following clinical trial parameters: number of patients to be included; number of simulations to be performed; number of prognostic factors to take into account, associated with the number of levels or values that could be taken by each factor; proportion of patients expected in each prognostic level; number of operators; and the parameters of the allocation methods to be tested.

In order to clarify the minimization and stratification, consider the following simple example of a trial including 10 patients. Two factors were identified: factor 1, with levels, a and b (e.g., for a surgical trial this could be "smoker" or "non-smoker"); and factor 2, with three values, a', b', and c' (e.g., the three operators in the trial). The study investigators were expecting 50 % a and 50 % b for factor 1, and 20 % a', 30 % b', and 50 % c' for factor 2. The software randomly generated the patients listed in Table 1, taking into account these expected proportions.

Minimization was coded as described by Pocock and Simon [29]. In our example, the allocation of the first patient was completely random. Suppose treatment A was assigned. The numbers then obtained for treatments A and B are shown in Fig. 1. If treatment A was assigned to the second patient, the sum of squared imbalances would be $2^2+1^2+1^2=6$, whereas it would be $1^2+1^2=2$ if treatment B was assigned (see Fig. 1). Therefore treatment B, which minimized the sum of squares, was assigned to the second patient by minimization.[2] This is a fully-deterministic minimization, where allocations depend

solely on the characteristics of already-included patients. The allocations for the 10 patients are listed in Table 1.

It is also possible to introduce some randomness so that the minimization is less predictable. Thus, according to the percentage $X \in [0-50 \text{ %}]$ of randomness chosen, the treatment allocated will be the one dictated by minimization in $(100-X)$ % of cases, and the other treatment will be allocated in X% of cases. This was programmed in a simple way by generating a random number U according to the uniform distribution on [0, 1]. The allocation was then made depending on $U \leq X$ or $U > X$. For example, in the case of randomization with 30 % randomness ($X=30$ %), B was the treatment that minimized the sum of squared imbalances for patient 5. The random number generated was 0.23. As this is less than 0.3, treatment A was finally assigned (see Table 1). Finally, it was also possible to introduce the treatment (A or B) as a factor in the minimization, in order to keep a better balance overall.

For the stratified blocked randomization, a blocked randomization list was generated for each stratum. The strata number was equal to the product of the number of levels of each factor (here, 2 levels for factor 1×3 levels for factor 2=6 strata). Initially, one block was generated by the software for each stratum (see Table 2). Then, a new block was generated for all strata each time a randomization list had been entirely allocated. Take the example of randomization in blocks of two: the program assigned the first patient (a, b') the first treatment on the corresponding randomization list (a, b'), here treatment B (see Table 2), and so on (see results in Table 1). For this 10-patient trial, it was not necessary to generate additional blocks. Similarly for randomization in blocks of four, the first patient (a, b') was assigned treatment A (see Table 2), and the following allocations are shown in Table 1.

Indicators used

For each set of patients, an imbalance indicator was calculated, which corresponded to the absolute value of the difference in numbers between treatments A and B, divided by the number of patients included (to allow comparison of the balance of one trial with another). The imbalance indicators of the various randomization methods, using our example, are listed in Table 1. An indicator of the within-factor imbalance was computed; it was the mean imbalance within prognostic factor levels. Each prognostic level imbalance was the absolute value of the difference in numbers between treatments A and B, divided by the number of patients included in the prognostic level.

Predictability was computed by operator (or center). We adopted four different methods to mimic how operators may predict treatment allocation: 1, prediction based upon knowledge of the last allocation only; 3, the last 3; 5, the last 5; or all, the operator had written down or had access to all his allocations [30, 40, 41]. Consider our example again, with a', b', and c' indicating the three operators, who have remembered only

---

[1] This software can be downloaded at the following address: chabouis.fr/helene/hermes

[2] We could also do a simple sum of the absolute imbalance values, but this method allowed us to penalize more serious imbalances. [39]

Table 1 Characteristics of the 10 patients randomly generated by the program and treatments allocated to them by the various randomization methods, final numbers, and balance indicators

Patient	Factor 1	Factor 2	Deterministic minimization	Minimization with 30 % randomness	(Random number generated)	Stratified randomization (block size: 2)	Stratified randomization (block size: 4)
1	a	b'	A[a]	A[a]		B	A
2	a	c'	B	B	0.61	B	A
3	b	c'	A	A	0.32	A	A
4	a	b'	B	B	0.52	A	B
5	b	a'	B	A	0.23[b]	B	A
6	a	a'	A	B[a]		A	B
7	b	b'	B[a]	B	0.68	B	B
8	b	c'	A	B	0.92	B	A
9	b	b'	A	A	0.41	A	B
10	a	c'	B	A	0.56	A	A
Number of treatments A			5	5		5	6
Number of treatments B			5	5		5	4
Imbalance indicator			0	0		0	0.2[c]

[a] Allocated at random because the sums of squared imbalances were the same for this patient

[b] This patient did not receive the treatment allocated by minimization, because the random number drawn was less than 0.3

[c] $|6–4|/10=0.2$

their last allocations, which were generated by deterministic minimization. Predictability was calculated from the second inclusion. The operator predicted that his new patient would receive the treatment other than the previous allocated one. Thus operator b' predicted his second patient would receive treatment B, since his first had received treatment A (see Table 3). His second patient did receive treatment B.

Numbers after inclusion of patient 1 (a, b')

Factor	Class	Treatment A	Treatment B
1	a	1	0
	b	0	0
2	a'	0	0
	b'	1	0
	c'	0	0

Numbers if treatment A is allocated to patient 2 (a, c')

Factor	Class	Treatment A	Treatment B	Imbalance
1	a	2	0	2
	b	0	0	0
2	a'	0	0	0
	b'	1	0	1
	c'	1	0	1

[a]imbalance =| treatment A –treatment B|

Numbers if treatment B is allocated to patient 2 (a, c')

Factor	Class	Treatment A	Treatment B	Imbalance
1	a	1	1	0
	b	0	0	0
2	a'	0	0	0
	b'	1	0	1
	c'	0	1	1

Fig. 1 Treatment allocation to patient 2 by minimization. Numbers after inclusion of patient 1 (a, b'), Numbers if treatment A is allocated to patient 2 (a, c'), Numbers if treatment B is allocated to patient 2 (a, c')

However, he predicted his third patient would receive treatment A whereas he was allocated treatment B, and so on.

The predictability indicator was calculated as the sum of cases where operators correctly guessed the treatment allocated to their patients, divided by the number of guesses. For our example, predictability indicators of each allocation method are shown in Table 3. If the operator remembered 3, 5, or all of his inclusions, he predicted his next patient would receive the treatment that was least affected among the latter.

The number of simulations was the number of samples (like the one in Table 1) simulated. Several minimization and/or stratification methods can be used on each sample. The simulations were activated by the "Launch" button on the "Simulation" tab of the Excel interface. The mean and standard deviation of the indicators of balance and predictability were calculated and these results appeared after a few seconds on the "Results" tab of the Excel interface.

Table 2 Randomization lists generated for each stratum for the stratified randomization in blocks of 2 (upper side) and blocks of 4 (lower side)

	a	b
a'	AB	BA
b'	BA	BA
c'	BA	AB
	a	b
a'	BABA	ABBA
b'	ABAB	BBAA
c'	AABB	AABB

Table 3 Calculation of the predictability indicator if the operator remembered his last allocation only

	Patients	Deterministic minimization	Minimization with 30% randomness	Stratified randomization (block size : 2)	Stratified randomization (block size : 4)
Operator a'	5, 6	BA	AB	BA	AB
Operator b'	1, 4, 7, 9	ABBA	ABBA	BABA	ABBB
Operator c'	2, 3, 8, 10	BAAB	BABA	BABA	AAAA
Predictability indicator		5/7 = 71%	6/7 = 86%	7/7 = 100%	2/7 = 29%

In green: operator predicted the treatment actually allocated. In red: operator guessed wrong and the treatment other than the one he had predicted was allocated

Results

The software developed provides access to the predictability and balance of a given randomization method. To present the results, we chose a reference situation (Trial 0), which we varied depending on the number of patients included, the number of selected prognostic factors, the number of operators, and the distribution of subjects within factors (Table 4). We performed 10,000 simulations for each trial. (Appendix 1).

Trial 0

Predictability and balance results for Trial 0, according to the method of randomization selected, are shown in Table 5. Overall imbalance and within-factor imbalance increased when minimization included more randomness (i.e., when X increased), and when the block size increased for stratification. Predictability decreased as minimization included more random allocations or as the block size increased for stratification. Predictability increased when the operator remembered more allocations

and it was maximal if he had written down all his allocations (Table 5). We then considered the situation of the operator remembering his last five allocations, because this seemed to be the most likely situation to occur in real-life practice [40].

Influence of parameter variation on predictability and imbalance in trial 0

Sample size

The graphs in Fig. 2 show the effect of the number of patients. For a given randomization method, when the number of patients included increased, the imbalance decreased, as did predictability. For a small trial (cf. Trial 1), the properties of the various methods differed greatly.

Number of prognostic factors

The graphs in Fig. 3 show the effect of the number of prognostic factors. For stratification, the imbalance increased greatly

Table 4 Characteristics of simulated trials (patients, factors, frequencies, operators, and strata)

Simulated trial	Number of patients (N)	Number of factors	Distribution of subjects within factors	Number of operators	Percentage of patients per operator	Number of strata (S)	Maximum block size (B)[a]
0	50	1	40/60 %	2	50/50 %	4	3.7
1	30	1	40/60 %	2	50/50 %	4	1.9
2	100	1	40/60 %	2	50/50 %	4	6.2
3	50	2	40/60 %, 30/70 %	2	50/50 %	8	1.6
4	50	5	[b]	2	50/50 %	192	0.06
5	50	1	40/60 %	1	100 %	2	6.2
6	50	1	40/60 %	4	[c]	8	1.6
7	50	1	10/90 %	2	50/50 %	4	3.7
8	50	1	10/90 %	2	10/90 %	4	3.7

[a] Kernan rule: $B=N/(S\times4)$

[b] Factor 1, 40/60 %; factor 2, 20/80 %; factor 3, 10/90 %; factor 4, 20/30/50 %; factor 5, 20/20/30/30 %

[c] 15/25/30/40 %

Table 5 Imbalance and predictability indicators of Trial 0 for various randomization methods

Randomization method	Predictability						Imbalance		Within factor imbalance	
	Operator remembered last allocation		Operator remembered last 5 allocations		Operator remembered all previous allocations					
	Mean %	SD %	Mean %	SD %	Mean %	SD %	Mean %	SD %	Mean %	(Min, Max) %
Minimization										
- Deterministic	73.8	5.5	75.3	5.7	95.4	4.5	0.6	1.4	2.3	(0.0, 14.3)
- Deterministic+T[a]	73.8	6.5	72.5	6.5	89.9	7.2	0.1	0.6	2.5	(0.0, 14.3)
- $X= 0$ %	66.8	6.9	69.7	6.3	84.8	7.3	1.3	2.0	3.0	(0.0, 23.1)
- $X=20$ %	60.5	7.3	63.9	7.0	75.1	8.6	2.0	2.4	4.2	(0.0, 41.2)
- $X=30$ %	56.4	7.4	59.1	7.1	65.9	8.7	3.4	3.5	6.3	(0.0, 64.7)
- $X=30$ %+T[a]	56.5	7.5	58.5	7.3	64.2	9.3	2.9	3.2	6.9	(0.0, 57.9)
Stratified randomization										
- Block size: 2	62.6	5.9	67.4	5.6	79.7	5.7	1.9	2.1	3.2	(0.0, 14.3)
- Block size: 4	56.1	7.0	61.1	6.1	71.0	6.0	2.7	2.6	4.2	(0.0, 25.0)
- Block size: 6	54.1	7.0	58.2	6.2	66.5	5.8	3.4	3.0	5.1	(0.0, 33.3)

[a] $+T$ minimization including treatment as a minimization factor

when the number of prognostic factors increased, as the number of strata increased more quickly. However, imbalance increased slightly for minimization as its predictability decreased.

Number of operators

The graphs in Fig. 4 show the effect of the number of operators. Imbalance increased with the number of operators. An increasing number of operators had a negative impact on both methods, but affected stratification more.

Subject distribution between prognostic factors

The graphs in Fig. 5 show the effect of a more or less unequal distribution of subjects between prognostic levels. Unequal distribution of subjects favored minimization (cf. Trial 0 vs. Trial 7).

Note that including treatment as a minimization factor almost always improved the two indicators (balance and predictability) simultaneously, except when there was only one operator (Trial 5) or when the number of subjects was poorly distributed between operators (Trial 8). Finally, note that in our examples, imbalance was always minimal for a deterministic minimization including treatment as a minimization factor (Table 5, Figs. 2, 3, 4, and 5).

Discussion

In this study, we applied computer simulation of stratification and minimization randomization techniques to a balanced two-arm study model, with the aim of discerning which technique was most appropriate. We found that there
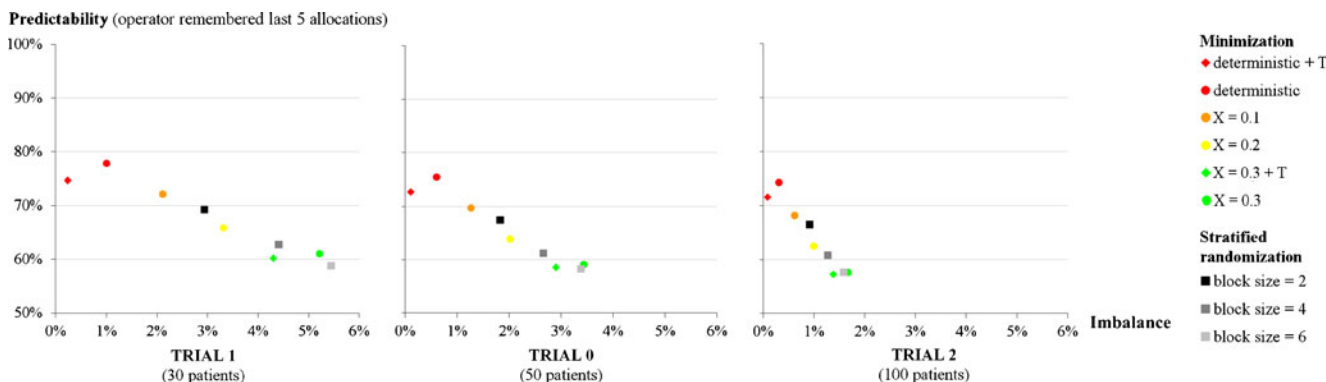

Fig. 2 Effect of change in sample size on imbalance and predictability indicators of various randomization methods (operator remembered his last five allocations)
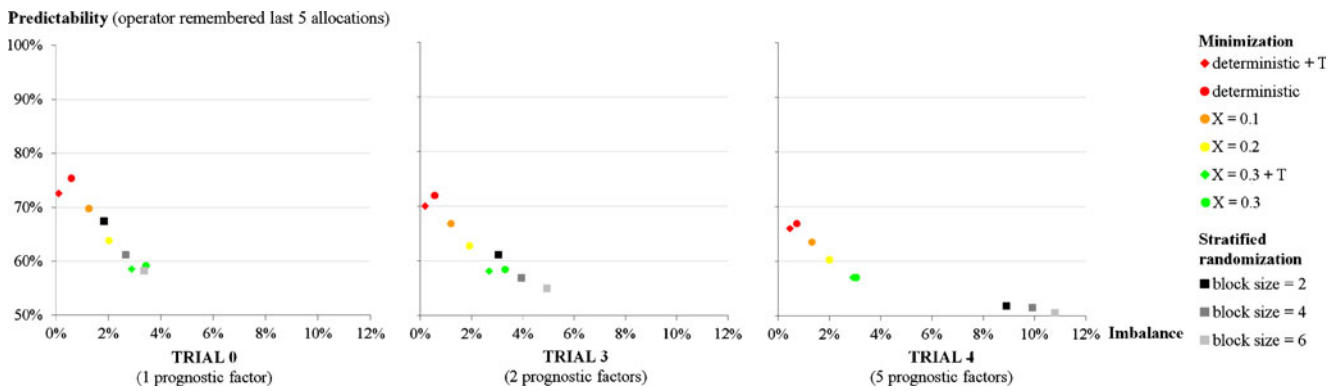
Fig. 3 Effect of the number of prognostic factors on imbalance and predictability indicators of various randomization methods (operator remembered his last five allocations)

was remarkable adaptability in the minimization, and inverse correlation between predictability and balance.

Choosing the right randomization method is important because it will affect the results of the clinical trial in terms of balance between groups, of predictability for the operator, and in terms of statistical analysis. Some rules, such as Kernan's rule [20], already exist to help investigators make that choice. However, we saw, for example on Fig. 5, that Trial 0 was more favorable to stratification than Trial 8, although Kernan's index was the same for both methods (3.7, Table 4). These rules, thus, lack subtlety in the choice of the randomization method because many other parameters can vary within a trial and influence the decision (e.g., the number of subjects, the number of prognostic factors to consider, the number of operators, and the distribution of subjects between factor levels). For this reason, statisticians have suggested performing simulations. However, up until now, investigators have not had freely-available simulation tools; this situation prompted us to create the HERMES software.

The first main result emerging from our study is that minimization was more adaptable and worked better in complex cases. This result confirms those of previous studies [9, 27, 35, 37, 38]. When the number of patients decreased or when the number of prognostic factors or the number of operators increased, the imbalance of stratification increased much more than that of minimization.

The second main result is that there was generally an inverse correlation between predictability and imbalance: when a less predictable method was wanted, its imbalance increased and vice versa. This idea has been mentioned or suggested by some authors [33, 38] but not as clearly demonstrated [30] until recently [36]. A trade-off between predictability and imbalance is therefore necessary as the perfect method (i.e., 0 % imbalance and 50 % predictability) does not exist.

This trade-off highlights a limitation of our study: although the software computed the predictability and imbalance of various methods, choosing the best method would still often be difficult. The imbalance–predictability graphs are helpful because various methods can be compared in terms of these two properties simultaneously; a method whose marker is located at the lower left of another is better. However, choosing the best method is less obvious. Analysis approaching that
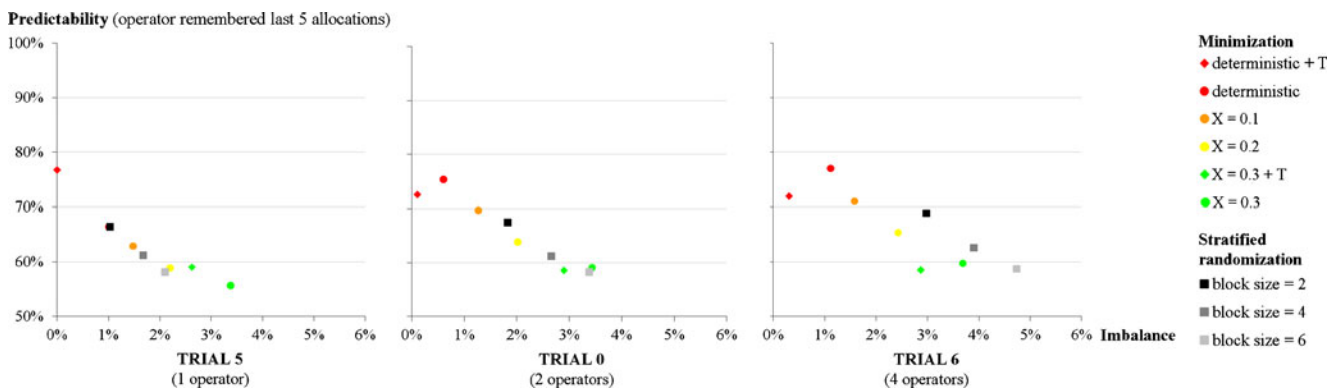

Fig. 4 Effect of the number of operators on imbalance and predictability indicators of various randomization methods (operator remembered his last five allocations)
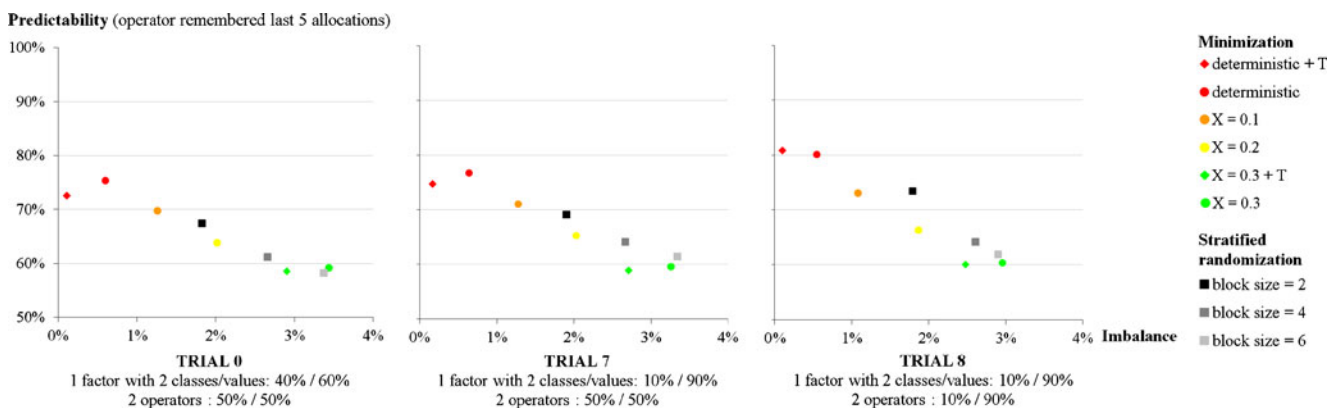
Predictability (operator remembered last 5 allocations)

Fig. 5 Effect of subject distribution between classes/values of prognostic factors on imbalance and predictability indicators of various randomization methods (operator remembered his last five allocations

of receiver operating characteristic curves would make us choose the method closest to the origin of the graph [42]. However, the graph is modified by the scale attributed to imbalance and predictability. Ultimately, the choice of the best method on the graph will depend on the trial; predictability is a less fundamental criterion for a double-blind trial than it is for an open-label trial, whereas balance will be less critical if a large difference in effect size between treatments or procedures being compared is expected.

For example, for Trial 0, in the case of a double-blind trial, we may decide to perform a deterministic minimization with the treatment as a minimization factor. If it were an open-label trial, we would prefer a minimization with a random factor of $X = 10$ or 20 % or even more. Note that for this trial, stratification by blocks of two was not far below these minimizations. We could therefore, decide in this case to adopt block stratification because it provides several advantages: it is easier to implement [20, 43] (the sequence can be generated in advance [44]), it is recommended by the authorities, and it allows all interactions that may exist between factors to be taken into account. This latter factor weakens it when many factors must be considered, because some intersections have very few patients. However, it is also a strong point if interactions between prognostic factors are suspected or if subgroup analyzes are planned [34].

The generalizability of our program is limited in two ways. Firstly, it does not take into account possible interactions between prognostic factors at the time of patient simulation. However, these interactions can rarely be quantified before the start of a dental trial. If significant interactions between factors are suspected, the results of stratification should be compared to those of a minimization, taking into account these interactions. For example, if an interaction is suspected between factor 1 with 2 levels (a and b) and factor

2 with 3 levels (a', b', and c'), patients can be simulated (as if the interaction existed) by 6 levels of minimization (aa', ab', ac', ba', bb', and bc'), instead of by the 5 levels which a minimization without interaction would have included (a, b, a', b', and c').

Secondly, we restricted the comparison of the different randomization methods to the case of a trial with two balanced arms. However, we wrote the program in Visual Basic for Applications so that the code is accessible and can be changed, if necessary, to adapt it to a wide variety of clinical trials.

A final limitation of our work is that it compared different methods of stratification and minimization on criteria of balance and predictability, but not on the statistical analysis of results. However, to do so would require making assumptions about trial outcomes. This would not be straightforward in our field, and we believe that existing data are preferable for such comparisons. Conclusions regarding the consequences of the randomization method on the statistical analysis obtained on real datasets can be found in Appendix 2.

In conclusion, the HERMES software does compare stratification and minimization in terms of predictability and balance, but it does not entirely solve the choice of the most suitable method for a trial. The right compromise between predictability and imbalance remains to be found, but the software helps to justify this choice based on concrete reasoning. It is available for free download at this internet address: chabouis.fr/helene/hermes.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1

Justification for the number of simulations (10,000) carried out in this study

We aimed to estimate the expected value of a random variable (predictability or imbalance) with n independent realizations (simulations were independent) and identical distribution (patients were simulated in the same proportions). Our indicator was therefore the estimator of the random variable's expected value. According to the central limit theorem, we know that the standard deviation of our estimator was $\sqrt{\frac{Var(X)}{n}}$, where n is the number of simulations (here 10,000).

Taking Trial 0 as an example, the precision of our imbalance indicator for deterministic minimization (whose estimated expected value was 0.597 %) was $1.42\ \%/\sqrt{10,000}$, which is 0.0142 %.

The 95 % confidence interval of our imbalance indicator was therefore $0.597 \pm 1.96 \times 0.0142$, or [0.569 to 0.625]. This level of precision seemed acceptable to compare these methods, and is the reason why we presented the results with one digit after the decimal point.

Appendix 2

How the allocation method impacts the analysis of resulting data

Type of analysis

Tests of statistical inference are based on the assumption of random assignment to treatment and control groups. Only simple randomization has this property, so that distorted p values and concerns over the validity of the analysis surround not just minimization but all other allocation methods [39]. However, the disadvantage of adaptive methods like minimization is that the correct analysis is complex and not clearly worked out [39, 44]. In fact, minimization achieves balance only among the marginal distribution of the strata [25]. Where the outcome measure is a continuous variable, various authors recommend that adjustment should be made for factors in the minimization using analysis of covariance [28, 32, 45]. Several other authors recommend using permutation tests to analyze trials where minimization has been used [43, 46]; however, because these tests are not straightforward in practice and seem to make little difference to the results obtained, some authors believe permutation tests are unnecessary and that a classical analysis will usually yield satisfactory conclusions provided that minimization factors

are used as covariates in the analysis [8, 9, 11, 12]. Some authors consider that the nominal significance level in that case should be adjusted [32, 35, 45], whereas Hagino et al. believe this is unnecessary [38]. For both stratified randomization and minimization, the collapsibility of the data should be checked, and if it depends on the statistic used, the tenuousness of the conclusions should be noted [47].

Inclusion of randomization covariates in the analysis

Although scientists may find the results of simple, unadjusted treatment comparisons with demonstration of good balance of important factors more convincing than the results of a covariate analysis, there is a consensus that the prognostic factors included in the randomization scheme should be taken into account in the analysis, not just for minimization [12, 32, 46, 48, 49] but also for stratification [50]. The p value for a difference between endpoint rates in treatment groups would be otherwise overestimated [18, 38, 51].

Effect on the nominal level and the power of the test

Authors in favor of stratified randomization put forward its value in reducing the risk of type I error [20] and increasing power [52], but this advantage is controversial. Some authors believe that both stratification and minimization procedures produce comparable improvements in reducing type I and II error [35]. Tu et al. found that minimization procedures were inferior to stratified allocation in reducing the two types of error, due to existing interactions between covariates [48]. However, other simulations have given opposite results [32, 35, 38].

References

1. Altman DG, Bland JM (1999) Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ 318(7192):1209
2. Pihlstrom BL, Barnett ML (2010) Design, operation, and interpretation of clinical trials. J Dent Res 89(8):759–772. doi:10.1177/0022034510374737
3. Pandis N, Polychronopoulou A, Eliades T (2010) An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 38(9):713–721. doi:10.1016/j.jdent.2010.05.014
4. D'Agostino RB, Sr. and Massaro JM (2004) New developments in medical clinical trials. J Dent Res 83 Spec No C:C18-24
5. Pandis N, Polychronopoulou A, Eliades T (2011) Randomization in clinical trials in orthodontics: its significance in research design and methods to achieve it. Eur J Orthod. doi:10.1093/ejo/cjq141
6. Pocock SJ, Assmann SE, Enos LE, Kasten LE (2002) Subgroup analysis, covariate adjustment, and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 21(19):2917–2930. doi:10.1002/sim.1296
7. Altman DG (1991) Randomisation. BMJ 302(6791):1481–1482

8. Halpern J, Brown BW Jr (1986) Sequential treatment allocation procedures in clinical trials—with particular attention to the analysis of results for the biased coin design. Stat Med 5(3):211–229

9. Watson HR, Pearce AC (1990) Treatment allocation in clinical trials: randomisation and minimisation compared in three test cases. Pharmaceutical Medicine 4(3):207–212

10. ICH Harmonised Tripartite Guideline, Statistical principles for clinical trials, International Conference on Harmonisation E9 Expert Working Group (1999) Stat Med 18(15):1905–1942

11. Buyse M (2000) Centralized treatment allocation in comparative clinical trials. Applied Clinical Trials 9:32–37

12. Green H, McEntegart DJ, Byrom B, Ghani S, Shepherd S (2001) Minimization in crossover trials with non-prognostic strata: theory and practical application. J Clin Pharm Ther 26(2):121–128

13. McEntegart D (2003) The pursuit of balance using stratified and dynamic randomization techniques: an overview. Drug Information Journal 37:293–308

14. Assmann SF, Pocock SJ, Enos LE, Kasten LE (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 355(9209):1064–1069. doi:10.1016/S0140-6736(00)02039-0

15. Committee for Proprietary Medicinal Products (CPMP), Points to consider on adjustment for baseline covariates (2004) Stat Med 23(5):701–709. doi:10.1002/sim.1647

16. Treasure T, MacRae KD (1998) Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. BMJ 317(7155):362–363

17. Rovers MM, Straatman H, Zielhuis GA, Ingels K, van der Wilt GJ (2000) Using a balancing procedure in multicenter clinical trials. Simulation of patient allocation based on a trial of ventilation tubes for otitis media with effusion in infants. Int J Technol Assess Health Care 16(1):276–281

18. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG (1976) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 34(6):585–612

19. Nam JM (1995) Sample size determination in stratified trials to establish the equivalence of two treatments. Stat Med 14(18):2037–2049

20. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI (1999) Stratified randomization for clinical trials. J Clin Epidemiol 52(1):19–26

21. Altman DG, Bland JM (1999) How to randomise. BMJ 319(7211):703–704

22. Efron B (1971) Forcing a sequential experiment to be balanced. Biometrika 58(1971):403–417

23. Wei LJ (1977) A class of designs for sequential clinical trials. J Am Stat Assoc 72:382–386

24. Soares JF, Wu CF (1983) Some restricted randomization rules in sequential designs. Communications in statistics—theory and methods 12:2017–2034

25. Signorini DF, Leung O, Simes RJ, Beller E, Gebski VJ, Callaghan T (1993) Dynamic balanced randomization for clinical trials. Stat Med 12(24):2343–2350

26. Altman DG, Bland JM (2005) Treatment allocation by minimisation. BMJ 330(7495):843. doi:10.1136/bmj.330.7495.843

27. Therneau TM (1993) How many stratification factors are "too many" to use in a randomization plan? Control Clin Trials 14(2):98–108

28. Taves DR (1974) Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 15(5):443–453

29. Pocock SJ, Simon R (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31(1):103–115

30. Brown S, Thorpe H, Hawkins K, Brown J (2005) Minimization—reducing predictability for multi-centre trials whilst retaining balance within centre. Stat Med 24(24):3715–3727. doi:10.1002/sim.2391

31. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134(8):663–694

32. Weir CJ, Lees KR (2003) Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. Stat Med 22(5):705–726. doi:10.1002/sim.1366

33. Toorawa R, Adena M, Donovan M, Jones S, Conlon J (2009) Use of simulation to compare the performance of minimization with stratified blocked randomization. Pharm Stat 8(4):264–278. doi:10.1002/pst.346

34. Kundt G (2009) Comparative evaluation of balancing properties of stratified randomization procedures. Methods Inf Med 48(2):129–134. doi:10.3414/ME0538

35. Birkett NJ (1985) Adaptive allocation in randomized controlled trials. Control Clin Trials 6(2):146–155

36. Zhao W, Weng Y, Wu Q, Palesch Y (2012) Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. Pharm Stat 11(1):39–48. doi:10.1002/pst.493

37. Verberk WJ, Kroon AA, Kessels AG, Nelemans PJ, Van Ree JW, Lenders JW, Thien T, Bakx JC, Van Montfrans GA, Smit AJ, Beltman FW, De Leeuw PW (2005) Comparison of randomization techniques for clinical trials with data from the HOMERUS-trial. Blood Press 14(5):306–314. doi:10.1080/08037050500331538

38. Hagino A, Hamada C, Yoshimura I, Ohashi Y, Sakamoto J, Nakazato H (2004) Statistical comparison of random allocation methods in cancer clinical trials. Control Clin Trials 25(6):572–584. doi:10.1016/j.cct.2004.08.004

39. Scott NW, McPherson GC, Ramsay CR, Campbell MK (2002) The method of minimization for allocation to clinical trials. A review. Control Clin Trials 23(6):662–674

40. Hills R, Gray R, Wheatley K (2003) High probability of guessing next treatment allocation with minimisation by clinician. Control Clin Trials 24(suppl 3S):70S

41. McPherson G, Campbell M, Elbourne D (2003) Minimisation: predictability versus imbalance. Control Clin Trials 24:133S

42. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36

43. Simon R (1979) Restricted randomization designs in clinical trials. Biometrics 35(2):503–512

44. Lachin JM, Matts JP, Wei LJ (1988) Randomization in clinical trials: conclusions and recommendations. Control Clin Trials 9(4):365–374

45. Forsythe AB (1987) Validity and power of tests when groups have been balanced for prognostic factors. Computational Statistics and Data Analysis 5:193–200

46. Kalish LA, Begg CB (1985) Treatment allocation methods in clinical trials: a review. Stat Med 4(2):129–144

47. Bracken MB (2001) On stratification, minimization, and protection against types 1 and 2 error. J Clin Epidemiol 54(1):104–105

48. Tu D, Shalay K, Pater J (2000) Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. Drug Inf J 34:511–523

49. Senn S (1995) A personal view of some controversies in allocating treatment to patients in clinical trials. Stat Med 14(24):2661–2674

50. Senn SJ (1989) Covariate imbalance and random allocation in clinical trials. Stat Med 8(4):467–475

51. Ohashi Y (1990) Randomization in cancer clinical trials: permutation test and development of a computer program. Environ Health Perspect 87:13–17

52. Ross N (1999) Randomised block design is more powerful than minimisation. BMJ 318(7178):263–264