E

TWC/21/6
ORIGINAL: English
DATE: May 26, 2003

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-First Session
Tjele, Denmark, June 10 to 13, 2003

THE USE OF INCOMPLETE BLOCK DESIGN IN DUS TRIALS

Document prepared by experts from Denmark and Poland

THE USE OF INCOMPLETE BLOCK DESIGN IN DUS TRIALS

Kristian Kristensen[1], *Wiesław Pilarczyk*[2]

Introduction

1. The increasing number of varieties to be tested in many crops makes complete replicates of all varieties large. As the influence of soil variability within a block usually increases as the block size increases, the increased number of varieties to be tested means larger variability and thus makes it more difficult to discriminate between candidate varieties and reference varieties. Several remedies exist for decreasing the block sizes.

2. First of all, the total number of varieties may be decreased by using grouping characteristics to separate the varieties and carry out an experiment for each group. This requires that the groups are non-overlapping at that all varieties can be assigned to the correct group with high probability. If those two requirements are not fulfilled it may be risky to carry out an experiment for each group, because this means that varieties in different groups cannot be compared using a standard approach. They can be compared only within so-called inter block stratum which usually leads to higher variance of comparisons as usually variance of blocks is much higher than the variance of plots. So only very big differences can be effectively detected, see e.g. Caliński and Kageyama (2002). See also the discussion.

3. Another possibility is to split the total number of plants per variety into more complete blocks, i.e. 6 blocks with 10 plants per plot in each instead of 3 blocks with 20 plants per plot or maybe even further down to 10 blocks with 6 plants per plot or 12 blocks with 5 plants per block. However, the use of many blocks instead of just a few increases the work needed for sowing/planting and other tasks needed to maintain the trial. Also the area needed for the whole trial increases as the number of blocks are increased because more area is needed for guard areas and for gangways etc. In the extreme cases the block sizes will only be decreased very slightly as the most of the area will be used for guard areas and gangways etc.

4. The last solution to be mentioned here is to use designs where each block does not contain all varieties (called incomplete block designs). For instance, there may be 80 varieties, which are to be compared using 3 plots with 20 plants per variety. Instead of having 3 blocks each with 80 plots (one with each variety) it could be arranged in 30 blocks with 8 plots each. This of course means that each block only can contain a subset of the varieties. If this assignment of varieties to blocks is done randomly the design may turn out to be a disaster where it may be impossible to compare some pairs of varieties and where many of the varieties may be compared with very low precision (or high variability). However, if the assignment of varieties to blocks are carried out in a sensible way (based on statistical methods) it is possible to construct a design where all pairs of varieties can be compared with almost the same precision - and a precision that may be considerably better than the one that can be obtained using complete blocks because of the smaller variance between plots in small blocks (here 8 blocks each) compared to the variance between plots in large blocks (here 80 plots each).

[1] Danish Institute of Agricultural Sciences, Denmark
[2] Agricultural University of Poznań and The Research Centre of Cultivar Testing, Poland

5. Such types of design have for many years been used in other trials. The ideas originate back to Yates (1939) and such types of design have long been used when comparing the performance of large number of varieties or lines (see, e.g. Le Clerg, 1966 or Patterson and Hunter, 1983). During the last few years examples of such designs have also been used in DUS trials (e.g. Pilarczyk, 2001 and Kristensen, 2000). The remaining part of this paper will focus on such designs, e.g. the principles, the availability and the benefits/draw backs of incomplete block designs.

Description of incomplete block designs

6. Several types of incomplete block design exist, but here we will only mention one such type, the ones called α-designs (or generalized lattice designs). The α-designs were proposed by Patterson and Williams (1976) as a generalization of lattice designs (Yates, 1939). The α-designs are very flexible as they are available for almost any number of varieties and for almost any number of replicates (at least two replicates) and for a wide range of block sizes. Another feature of these designs is that the incomplete blocks may be collected so that they form complete replicates. This means that the designs may be laid out in the field such that from a practical point of view look like randomized complete block designs.

7. The construction of optimal designs is most easily done using a computer program. One such program is Alpha+, which was written by CSIRO, Australia, and Biomathematics & Statistics Scotland, and is available from any of those institutes. Near optimal designs can also by constructed manually from the generating arrays given by Patterson and Williams (1976). Very briefly, this design type gives designs where 1) the number of times pairs of varieties come together in an incomplete block is as equal as possible and 2) all pairs of varieties can be compared indirectly through as many other varieties as possible.

8. The analysis of data can be done using the same statistical model as for designs with complete blocks, i.e. assuming additive variety and block effects (model 1 below). However, because the incomplete blocks can be and usually are grouped to form complete replicates the block effects are usually subdivided into a replicate effect and an effect of incomplete blocks within replicates (model 2). The effect of incomplete (and complete) blocks can be either systematic or random. When it is reasonable to assume that the effect of incomplete blocks are random (model 3) the generalized lattice will give a precision that is at least as good as if a randomized complete block design were used.

(1) $Y_{vb} = \mu + \alpha_v + \gamma_b + E_{vb}$ Effect of blocks are assumed systematic

(2) $Y_{vrb} = \mu + \alpha_v + \beta_r + \gamma_{rb} + E_{vrb}$ Effect of blocks within reps. are assumed systematic

(3) $Y_{vrb} = \mu + \alpha_v + \beta_r + C_{rb} + E_{vrb}$ Effect of blocks within reps. are assumed random

E_{vb}, E_{vrb} and C_{rb} asumed independent and normal distributed with constant varians, σ_E^2 and σ_C^2

9. When using incomplete blocks the user first has to choose an appropriate block size and then later to decide how to place those in the field. In the following we try to formulate the authors' experience on those matters. The block size is usually chosen in the neighbourhood of the square root of number of varieties - e.g. with 80 varieties we would usually chose a block size around 9. If the field to be used is heterogeneous we would decrease the block size and use a design with 8, 7 or may be even 6 plots per block. In the field we would first subdivide the area into complete replicates such that the area covering each of those are as homogeneous as

possible and then subdivide each of the complete replicates into a number of incomplete blocks such that the plots within each of those are as homogeneous as possible. In addition to those principles we try also to fulfil the following: an incomplete block should cover only one row of plots in the field and should be compact. In large trials (with many varieties) we also try to ensure that field operations such as sowing, planting etc. can be stopped at a border between two complete replicates.

Efficiency factors and effectiveness of block designs

10. Every block design is characterized by a so-called efficiency factor (often denoted by ε) that denotes the amount of information contained in within-block stratum (it is a theoretical characterization of the design). Two extreme cases are:

a) complete blocks – all information contained in within block stratum (efficiency factor ε=1), and

b) design with all blocks of one plots (efficiency factor ε=0), all information contained in between block stratum.

11. General dependence is: bigger blocks provide higher efficiency factors. So, on the assumption that the variance of error is not dependent on the block size, the randomized complete block design is the best one as a design possessing the highest efficiency factor. However, in practice, in plot experiments, we observe that the variance of error is increasing when the block size increase. So looking for the best designs means in practice looking for the mid-course between the loss in efficiency factor and gain in variance of error - both caused by reduction of block size. So, in practice, every design can be better characterized by so-called effectiveness. It is the ratio of average variance of treatment comparisons provided by randomized complete block design to the same variance provided by design under consideration. The best design is the one with the highest effectiveness. Let $MS_e(k)$ stand for mean square for error (it is the estimate of the error variance) in design with blocks of k plots and let ε_k mean its efficiency factor. Then the effectiveness of block design with blocks of *k* plots can be calculated as:

a) $MS_e(v)/(MS_e(k)/\varepsilon_k) = \gamma\ \varepsilon_k$ under model 2 considered here,

b) $\gamma\left\{\varepsilon_k + \dfrac{(1-\varepsilon_k)(s-1)}{\gamma(v-1)-(v\text{-}s)}\right\}$ under model 3, see Patterson and Hunter (1983).

12. Here $\gamma = MS_e(v)/MS_e(k)$, v is the total number of varieties and s denotes the number of incomplete blocks contained within one complete replicate.

Examples of designs

13. An example of an α-design with 71 varieties is shown in appendix 1**.** Each column of plots forms a complete replicate, which is then subdivided into 9 incomplete blocks of which one contains only 7 plots and the remaining contain 8 plots each (two block sizes are used in order to let the number of plots per replicate total 71)**.** The records for UPOV characteristics nos. 6, 8, 9, 10, 13, 14, 16, 17, 18, 19 and 20 were analyzed using the model where the effect of

incomplete blocks within replicates are assumed random (model 3 above). This model reduced the LSD-values for these characters by 0 to 24%. In other years the reduction may be different. In Table 1 the reduction in LSD-values is shown for each character in each of the years 1997-1999 (based on Kristensen, 1999 and Kristensen, 2000). In these years there were 55, 66 and 71 varieties and the block sizes were 11, 11 and 8(7), respectively. In the last column the reduction in COYD LSD-values is shown. The reductions in COYD LSD-values are calculated by comparing the COYD LSD values when the COY-D is based on simple variety-by year means and when it is based on yearly estimates using a model with random incomplete-block effects. The largest reductions were found for characteristic numbers 16 and 17.

Table 1: Reduction of LSD-values, %, for some characters of Yellow Mustard when α-designs are used instead of randomised complete block designs. The last column shows the reduction in COYD LSD values.

No	Character name	1997	1998	1999	COY-D
06	Leaf : Number of lobes	4.4	0.1	0.2	1.2
08	Leaf : Length	0.5	0.0	3.0	1.0
09	Leaf : Width	1.7	0.0	1.0	2.0
10	Leaf: Length of petiole	7.7	1.0	0.5	2.5
13	Flower: Length of petals	0.3	0.2	0.1	0.0
14	Flower: Width of petals	0.1	0.5	0.1	0.2
16	Plant: Height (at full flowering)	9.0	8.1	23.8	8.2
17	Plant: Total length incl. side branches	25.6	22.9	9.7	14.9
18	Siliqua: Length	4.8	3.4	3.9	4.2
19	Siliqua: Length of beak	3.3	3.5	3.3	2.6
20	Siliqua: Length of peduncle	0.6	0.0	0.6	0.0

14. A similar investigation was performed in Poland. In 1995-1996 three experiments on maize were conducted. The number of varieties involved was 200, 212 and 98 respectively. The four characteristics for which calculations were performed was::

c_1 – plant length,
c_2 – height of insertion of upper ear,
c_3 – leaf width,
c_4 - length of peduncle.

15. The effectiveness of incomplete blocks in these three trials is reported in Table 2.

Table 2: Reduction of LSD-values, %, for some characters of maize when α-designs are used instead of randomized complete block designs.

Year	Place	V	k	ε	Reduction of LSD-values			
					c_1	c_2	c_3	c_4
1995	Słupia W.	200	13(12)	0.870	18.4	20.8	2.6	1.7
1996	Słupia W.	212	15(14)	0.889	32.8	18.1	5.0	6.1
1996	Przecław	98	10(9)	0.846	12.7	10.1	15.3	14.1

16. As one can see, the effectiveness is dependent on the year and characteristic. A similar calculation performed for a trial on French bean conducted in 1998 at Słupia Wielka showed high effectiveness of applied incomplete block design (v=40, k=10, ε=0.8540) for plant height. For three characteristics concerning the pod, incomplete block design had no advantage over complete block design.

<u>Discussion and conclusion</u>

17. The benefit of using incomplete block designs is first of all the possibility to compare many varieties in one design with a precision that is about as good as if only a few varieties were compared in a design with complete blocks. Compared to randomized complete block designs with many varieties the incomplete block design is superior for characteristics that are sensible to soil fertility. For characteristics that are less sensitive to soil fertility the gain in precision will be smaller or zero (but never negative). As the gain from using incomplete block designs instead of randomized complete block designs is usually largest in trials where the randomized complete block design would have yielded large LSD-values the LSD-values from trial to trial are expected to be more equal when using incomplete block designs.

18. The construction, layout in the field and analysis of incomplete block designs are slightly more complicated than for randomized complete block designs. However, today efficient programs for PC's are available that can ease these processes.

19. The use of incomplete block designs - as described above- prevents the user imposing "restricted randomizations" in order to let groups of varieties be located close to each other. Such grouping may be imposed by using designs similar to how α-designs may be used to construct incomplete split plots (Kristensen, 2003). As an example we describe a possible plan for a crop where 40 varieties are to be compared. Of those 40 varieties 11 are in group 1, 12 are in group 2, 12 are in group 3 and 4 are in group 4. An example of how such a trial can be laid out in the field is given in appendix 2. The layout shown there is based on three α-designs with a block size of 4(3) plots per block and 11, 12 and 12 varieties each together with one complete randomized block design with 4 varieties. The layout allows the varieties within groups to be compared with a precision that is expected to be similar (or slightly better) than if 4 separate trials were used - and to compare varieties in different groups with a lower precision.

References

CALIŃSKI T., KAGEYAMA S. 2002. Block Designs: A Randomization Approach. Springer Verlag.

KRISTENSEN, K. 1999. Efficiency of incomplete block designs in Spring Rape and Yellow Mustard. TWC/17/8. UPOV, Geneva. 7 pp.

KRISTENSEN, K. 2000. Efficiency of incomplete block designs in Winter Rape Spring Rape and Yellow Mustard. TWC/18/4, Geneva, 6 pp.

KRISTENSEN, K. 2003. Incomplete split-plots in variety trials - based on α-designs. Cultivar news. To appear.

LE CLERG, E.L. 1966. Significance of Experimental Design in Plant Breeding. Plant Breeding. Pp. 243-313

PATTERSON, H.D., WILLIAMS, E.R. 1976. A new class of resolvable incomplete block designs, Biometrika 63, pp. 83-92

PATTERSON H.D., HUNTER E.A. 1983. The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. Jour. Agric. Sci. Camb. 101, pp. 422-433.

PILARCZYK, W. 2001. The efficiency of incomplete block designs in DUS trial on pea varieties, TWC/19/3, UPOV Geneva, pp.1-5.

YATES, F. 1939. The recovery of inter-block information in variety trials arranged in three-dimensional lattices. Annals of Eugen. Pp. 9, 136-156.

YATES, F., 1940. The recovery of inter-block information in balanced incomplete block designs. Annals of Eugenies. Pp. 10, 317-325.

APPENDIX 1

1 1 05	2 1 16	3 1 29
1 1 60	2 1 60	3 1 53
1 1 57	2 1 20	3 1 10
1 1 09	2 1 64	3 1 64
1 1 26	2 1 55	3 1 65
1 1 47	2 1 25	3 1 04
1 1 61	2 1 28	3 1 14
1 1 33	2 1 62	3 1 23
1 2 48	2 2 48	3 2 08
1 2 53	2 2 42	3 2 68
1 2 63	2 2 56	3 2 46
1 2 15	2 2 57	3 2 47
1 2 31	2 2 02	3 2 34
1 2 37	2 2 04	3 2 07
1 2 25	2 2 46	3 2 03
1 2 68	2 2 44	3 2 66
1 3 42	2 3 14	3 3 20
1 3 01	2 3 24	3 3 15
1 3 65	2 3 33	3 3 36
1 3 07	2 3 35	3 3 38
1 3 17	2 3 15	3 3 27
1 3 32	2 3 19	3 3 56
1 3 16	2 3 01	3 3 69
1 3 27	2 3 34	3 3 26
1 4 28	2 4 52	3 4 62
1 4 36	2 4 67	3 4 11
1 4 10	2 4 07	3 4 17
1 4 59	2 4 49	3 4 24
1 4 03	2 4 10	3 4 44
1 4 19	2 4 05	3 4 54
1 4 54	2 4 63	3 4 61
1 4 12	2 5 13	3 4 49
1 5 04	2 5 39	3 5 21
1 5 66	2 5 68	3 5 39
1 5 18	2 5 23	3 5 51
1 5 55	2 5 58	3 5 28
1 5 35	2 5 61	3 5 45
1 5 58	2 5 43	3 5 63
1 5 21	2 5 36	3 5 42
1 5 11	2 6 41	3 5 33
1 6 70	2 6 38	3 6 41
1 6 51	2 6 03	3 6 67
1 6 43	2 6 11	3 6 58
1 6 52	2 6 09	3 6 60
1 6 56	2 6 70	3 6 30
1 6 08	2 6 31	3 6 48
1 6 29	2 6 65	3 6 32
1 6 62	2 7 59	3 6 12
1 7 50	2 7 17	3 7 02
1 7 64	2 7 30	3 7 43
1 7 69	2 7 66	3 7 40
1 7 39	2 7 22	3 7 05
1 7 44	2 7 51	3 7 01
1 7 34	2 7 69	3 7 31
1 7 41	2 7 37	3 7 59
1 8 46	2 8 45	3 7 55
1 8 20	2 8 18	3 8 18
1 8 40	2 8 53	3 8 19
1 8 22	2 8 54	3 8 25
1 8 23	2 8 08	3 8 06
1 8 67	2 8 32	3 8 22
1 8 71	2 8 71	3 8 57
1 8 24	2 8 26	3 8 70
1 9 45	2 9 40	3 9 52
1 9 06	2 9 29	3 9 09
1 9 14	2 9 50	3 9 37
1 9 13	2 9 47	3 9 50
1 9 02	2 9 21	3 9 71
1 9 38	2 9 06	3 9 35
1 9 30	2 9 27	3 9 13
1 9 49	2 9 12	3 9 16

Layout of trial with 71 varieties of Yellow Mustard in a DUS trial in Denmark.

The trial has 3 complete replicates that are subdivided into 9 blocks of 8 (7) plots each.

Each cell is a plot. The first digit in each cell gives the number of the whole replicate; the next one gives the incomplete block number (numbered from 1 to 9 in each whole replicate) and the last two digits show the variety number (numbered from 01 to 71).

APPENDIX 2

Layout of trial with 39 varieties assigned to 4 groups with 11, 12, 12 and 4 varieties, respectively

The trial has 3 complete replicates that are subdivided into 10 blocks of 4 (3) plots each.

Each cell is a plot. The first digit in each cell denotes the group number (different color), the second number gives the number of the whole replicate; the next one gives the incomplete block number (numbered from 01 to 10 in each whole replicate) and the last two digits show the variety number (numbered from 01 to 39).

4 1 01 37	2 2 01 20	3 3 01 31
4 1 01 38	2 2 01 18	3 3 01 24
4 1 01 36	2 2 01 13	3 3 01 29
4 1 01 39	2 2 01 15	3 3 01 30
1 1 02 02	3 2 02 35	1 3 02 03
1 1 02 10	3 2 02 31	1 3 02 10
1 1 02 05	3 2 02 24	1 3 02 07
1 1 02 09	3 2 02 26	1 3 02 11
2 1 03 20	4 2 03 36	1 3 03 06
2 1 03 18	4 2 03 37	1 3 03 02
2 1 03 17	4 2 03 38	1 3 03 04
2 1 03 22	4 2 03 39	1 3 04 08
3 1 04 34	1 2 04 08	1 3 04 09
3 1 04 28	1 2 04 11	1 3 04 05
3 1 04 31	1 2 04 02	1 3 04 01
3 1 04 33	2 2 05 21	2 3 05 18
3 1 05 29	2 2 05 14	2 3 05 19
3 1 05 26	2 2 05 19	2 3 05 21
3 1 05 24	2 2 05 22	2 3 05 12
3 1 05 25	1 2 06 09	3 3 06 25
2 1 06 23	1 2 06 07	3 3 06 27
2 1 06 19	1 2 06 06	3 3 06 35
2 1 06 13	1 2 06 03	3 3 06 34
2 1 06 16	1 2 07 01	2 3 07 22
1 1 07 07	1 2 07 10	2 3 07 20
1 1 07 08	1 2 07 04	2 3 07 15
1 1 07 04	1 2 07 05	2 3 07 16
3 1 08 35	2 2 08 23	3 3 08 33
3 1 08 32	2 2 08 12	3 3 08 28
3 1 08 27	2 2 08 16	3 3 08 26
3 1 08 30	2 2 08 17	3 3 08 32
1 1 09 01	3 2 09 33	2 3 09 17
1 1 09 11	3 2 09 29	2 3 09 14
1 1 09 06	3 2 09 34	2 3 09 13
1 1 09 03	3 2 09 32	2 3 09 23
2 1 10 21	3 2 10 28	4 3 10 38
2 1 10 14	3 2 10 25	4 3 10 36
2 1 10 12	3 2 10 30	4 3 10 37

[End of document]