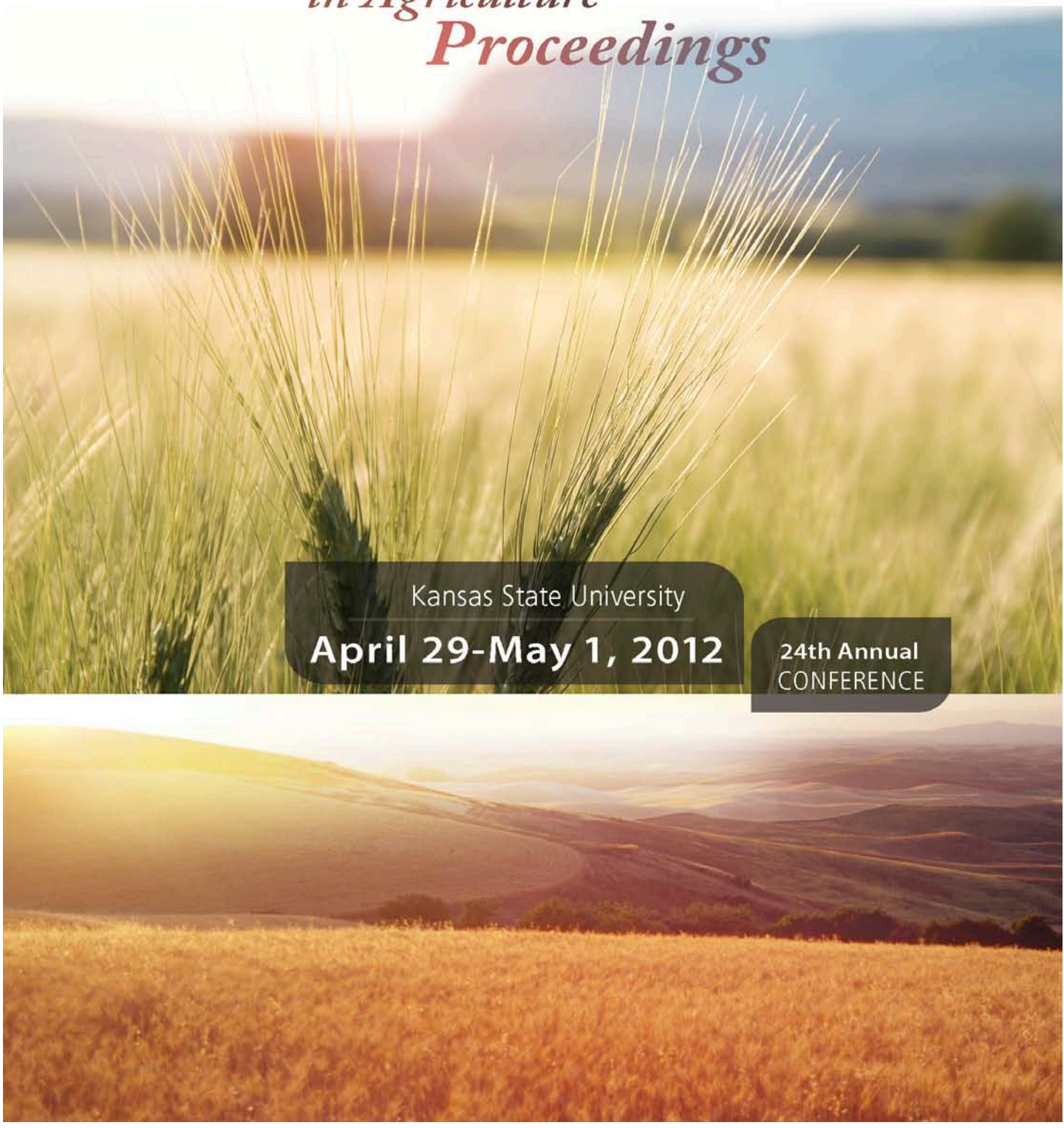
Conference on Applied Statistics in Agriculture Proceedings



Kansas State University

April 29-May 1, 2012

24th Annual CONFERENCE

ARMedAND DANGEROUS:
THE CONSEQUENCES OF NOT RANDOMIZING THE FIRST BLOCK

Edzard van Santen and Mark West

Edzard van Santen, Ph.D., Dept. of Agronomy and Soils, Auburn University, AL 36849-6248. USA.Email: vanedza@auburn.edu

Mark West, Ph.D., Northern Plains Area Statistician, USDA Agricultural Research Service, 2150 Centre Ave., Fort Collins, Co 80526-8119. USA. Email: Mark.West@ars.usda.gov

ABSTRACT

Replication and randomization and are the keys for statistically valid experiments. Both are necessary components for statistically valid experimentation. Yet it is an industry wide practicein weed science research to assign treatment in the first block of a randomized complete block design in a systematic order for reasons of convenience. We investigated this practice by comparing four randomization/analysis scenarios: (**i**) complete randomization in all blocks, (**ii**) systematic assignment of treatmentsin block 1, where the best treatment was assigned to the best plot, (**iii**) systematic assignment of treatmentsin block 1, where the best treatment was assigned to the worst plot,and (**iv**) systematic assignment of treatments in block 1 but not using it in the analysis. We created 1000 simulated datasets for three levels of experimental precision and two group sizes (t=3 and t=9). Results indicate that dropping block 1 from the analysis resulted in a loss of power, as did the best to worst assignment scenario. The best to best assignment resulted in increased power that would lead to an inflated Type I error. Differences between the drop block 1 and best to worst scenarios tended to become smaller as the experiment size increased and the experimental precision decreased. The recommendation for the practice would be (1) to follow proper randomization procedures, and (2) to add an extra block to the experiment for demonstration purposes only.

INTRODUCTION

Replication and randomization and are the keys for statistically valid experiments. Assigning treatments to at least two experimental units enables the estimation of experimental error, the variation among experimental units treated alike. Randomization is defined as the process of assigning experimental units to treatments under the assumption that each experimental unit has an equal chance of being assigned to a given treatment (Lentner and Bishop, 1993). One of the earliest if not the earliest references to randomization is Fisher's (1926) publication. It became more widely known in his now classic book *The Design of Experiments*published in 1935 (Fisher, 1966). The concept of randomization has greatly contributed to the advances of research in every field (Harville, 1975).As Hinkelmann and Kempthorne (2007) pointed out, "An experimenter who does not use randomization with variable material is widely regarded as incompetent"; use of randomization in experiments is now common practice.

Randomization means applying a well-understood random procedure to assign experimental units to different treatment groups. This procedure can be done by flipping a coin, rolling a dice, or using computer software with a good random number generator. Randomization ensuresthat factors not explicitly controlled for in the design do not exert an undue influence on the outcome of an experiment. Randomizationalso allowsus to make causal inferences and providesa probability model for drawing inferences. Randomization, and not the treatments, as a source for differences, is a measure of uncertainty associated with the confidence level (or *P*-value) (Ramsey and Shafer, 2002). Therefore, the application of randomization into an experimental design makes an objective assessment of treatments possible.

Yet, common practice often tends to ignore this idealized process. One such instance is aweed science industry-wide habit of not randomizing the first block in randomized complete block (RCB) field trials. The argument for such an approach is a practical one; treatments can be demonstrated more easily at field days with a systematic arrangement. This would not be a problem if the treatment list itself were randomized but field trial management software such as ARM (Gylling Data Management, Inc., Brookings, SD, USA) and others use a time-saving approach to creating treatment list such as given in Table 1. Not only will the first block of such an experiment not have a random assignment of treatments to plots (= experimental units) but a split-plot restriction on the randomization of the underlying RCB design is induced through such action. This is not the fault of the software designersas ARM offers a radio button that will enable a randomized assignment of treatments in all blocks. The non-randomized first block default feature is due to customer demands.

Table 1. Treatment table for a hypothetical herbicide experiment with three rates for each of two herbicides.

Trt_N	Herbicide	Rate
1	A	0.5 x
2	A	1.0 x
3	A	1.5 x
4	B	0.5 x
5	B	1.0 x
6	B	1.5 x

The potential statistical consequences of such an approach would probably be small if the number of complete blocks were quite large. But a standard agronomic trial typically has no more than four complete blocks. Having a non-random assignment of experimental units for 25% of the total experimental units could have severe consequences. Furthermore, blocks will likely not be homogeneous as field trial designsoften represent ‘convenience blocking’, i.e. the total experimental area is subdivided to arrive at a convenient blocking pattern, not to maximize the differences among blocks and minimize the differences within blocks. In many cases it is likely each blocks would consist of a single tier of plots with plots lined up like pearls on a string even though it has been know for decades that equilateral blocks would minimize within block variation.

The objectives of this study were to assess the consequences of not-randomizing the first block on statistical power in simulated experiments.

SIMULATION

The underlying linear additive model for these simulated experiments was

$$Y_{ij} = \mu + \alpha_i + B_j + e_{ij},$$

whereμ is the overall mean of the experiment, α_i is the effect of the i^{th} treatment, $B_j$is the effect of the j^{th} block and e_{ij} is the corresponding residual; i ranged from 2 to 9, and j from 3 to 10. A random block effect was created for each block of the simulated dataset. We then created a random interaction (= residual) and a random plot quality effect for each plot. To the sum of the block, interaction, and plot quality effects we added a fixed treatment effect to arrive at the "observed" value for Y. The magnitude of interaction and plot quality effects was set to 50, 75, and 100% of the maximum fixed treatment effect to represent a range of experimental conditionsfrom high precision to low. We generated 1000 simulated datasets for each treatment number x interaction magnitude x number of blocks combination, calculated the *P*-value for treatments and from these the power based on those 1000 datasets.

We investigated four scenarios:(**1**) the first block of each basic dataset was either left as is (**All random**);(**2**) fixed treatments were added to the random components in block 1 only by rank, i.e. the best plot received the best treatment **(Best to Best**); (**3**) fixed treatments were added to the random components in block 1 only by reversed rank, i.e., the worst plot received the best treatment (**Best to Worst**); and (**4**) first block was dropped from the dataset. Our expectation was that compared to the all random arrangement, "**Best to Best**" would show increased power because treatment differences would be magnified, "**Best to Worst**" should show drastically reduced power because treatment differences would be minimized, and "**Block 1 deleted**" should have somewhat reduced power.

RESULTS

Effect on the overall F-test

In very precise experiments there is a considerable loss of power when the number of blocks is low for either treatment number for **Best to Worst** assignment scenario (Fig. 1, Interaction 50%, green line) when compared to **All Random**. It took approximately twice the number of blocks to achieve 80% power under the **Best to Worst**scenario compared to a the **All Random** scenario. For a standard four-block RCB experiment the loss of power was a minimum of 53% (data not shown). The loss of power incurred under the **Block 1 deleted** scenario (blue dashed line in Fig. 1) compared to the **All Random**was less than half (24%) that number. As the residual error increased the penalty incurred for these two scenarios decreased, particularly with an increase in the number of treatments (Fig.1, lower right hand panel). However, under these conditions the **Best to Best** scenario (Fig. 1, red solid lineclearly showed a bias for a significant increased treatment effect when compared to **All Random**.

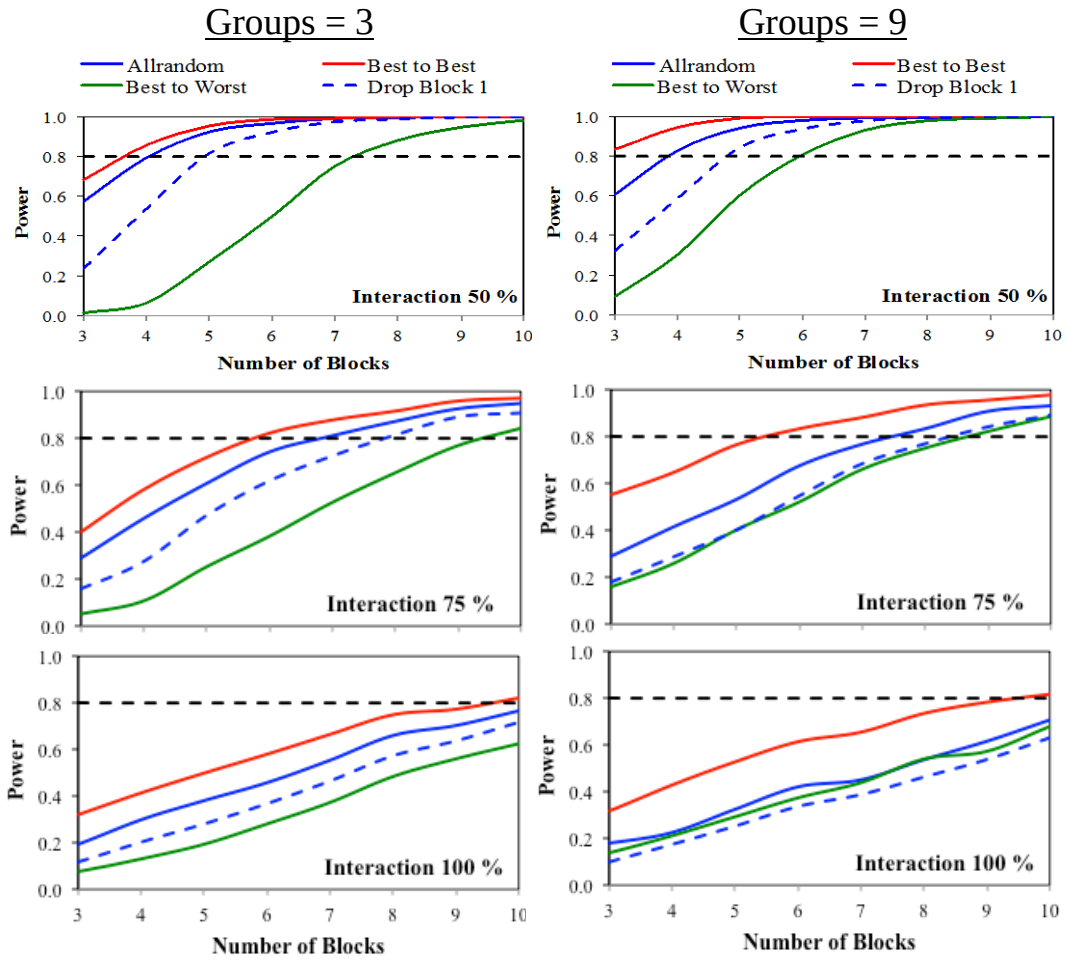


Fig. 1. Effect of block number on the power of the overall F-test for three and nine treatment groups based on the following experimental conditions: (i) **All Random** (solid blue line), where treatments were randomly assigned to plots in all blocks); (ii) **Best to Best** (red solid line), where the best treatment was assigned to the plot with the best inherent quality in Block 1; (iii) **Best to Worst** (green solid line), where the best treatment was assigned to the plot with the worst inherent quality in Block 1;and (iv) **Drop Block 1** (dashed blue line), where there was some systematic assignment of treatments to plots in block 1 but that block was dropped from the analysis. The interaction term = residual was set to either 100, 75, or 50% of the maximum absolute treatment effect to simulate experiments of increasing precision in 1000 simulated datasets for each condition.

The effect of **Best to Best**increased as the number of treatments increased. For a standard 4-block RCB experiment with nine treatments the difference in power to a regular randomization was 27%, which essentially amounts to a Type I error rate of 27%, declaring far more tests significant than the underlying experiment would have warranted.

Effect on the maximum contrast

We also investigated the effect on the contrast between the best and worst treatment for t = 3 and 9, i.e. a theoretical difference of 20 units. As expected, the results magnify the results obtained for the overall F-test as this contrast contributes the majority to the treatment variance

(Fig. 2). The effect of treatment number was also magnified. One aspect that warrants further investigation is the effect of dropping block 1 from the analysis. It appears that for larger trials with a large residual error this scenario performed worse than the best to worst assignment.

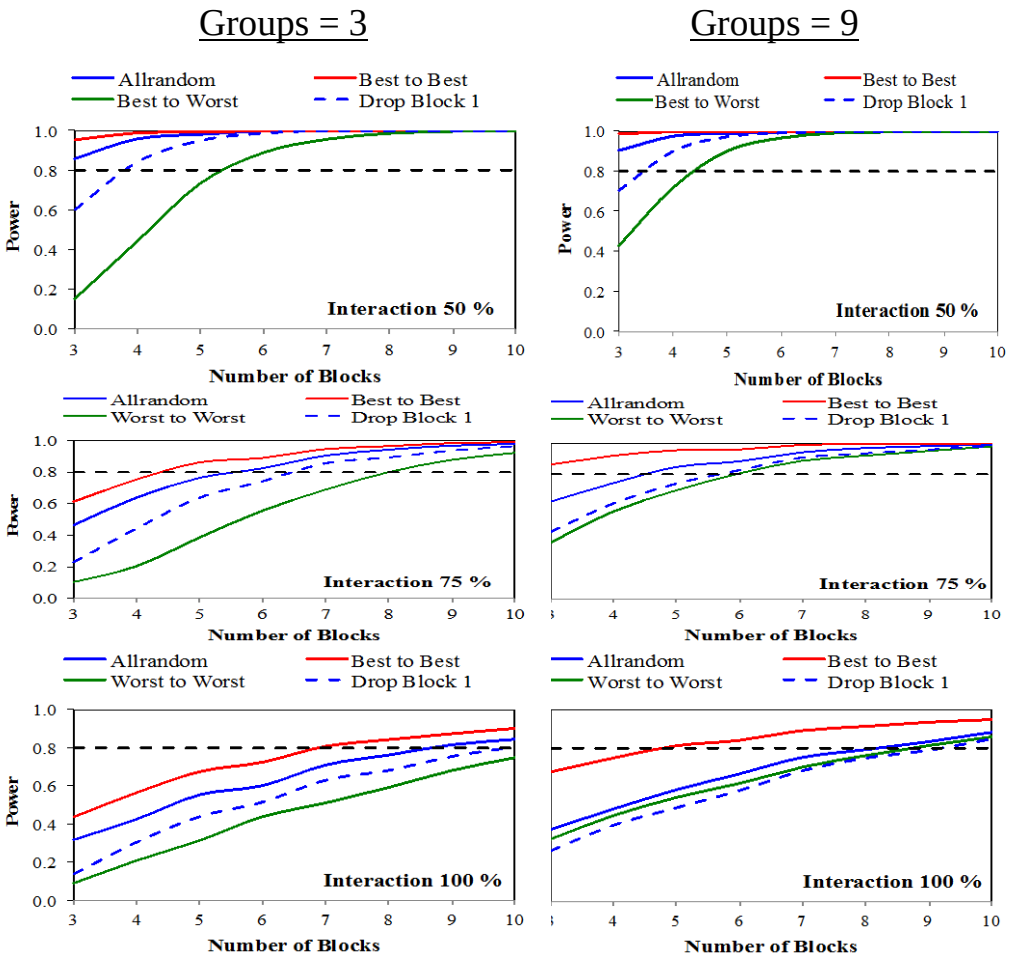


Fig 2. Effect of block number on the power of the pairwise comparisons between the best and worst treatment (maximum treatment difference) for three and nine treatment groups based on the following experimental conditions: (i) **All Random** (solid blue line), where treatments were randomly assigned to plots in all blocks); (ii) **Best to Best** (red solid line), where the best treatment was assigned to the plot with the best inherent quality in Block 1; (iii) **Best to Worst** (green solid line), where the best treatment was assigned to the plot with the worst inherent quality in Block 1; and (iv) **Drop Block 1** (dashed blue line), where there was some systematic assignment of treatments to plots in block 1 but that block was dropped from the analysis. The interaction term = residual was set to either 100, 75, or 50% of the maximum absolute treatment effect to simulate experiments of increasing precision in 1000 simulated datasets for each condition.

Effect on the intermediate contrast

The power of a test decreases as the expected difference between two means decreases, as was the case for the intermediate contrast where the expected difference was half the maximum treatment difference (Fig. 3). In very precise experiments, which every experimenter strives for, the best to worst assignment in block 1 led to a drastic reduction in power. As the

precision of the overall experiment decreased the differences among the four scenarios all but disappeared.

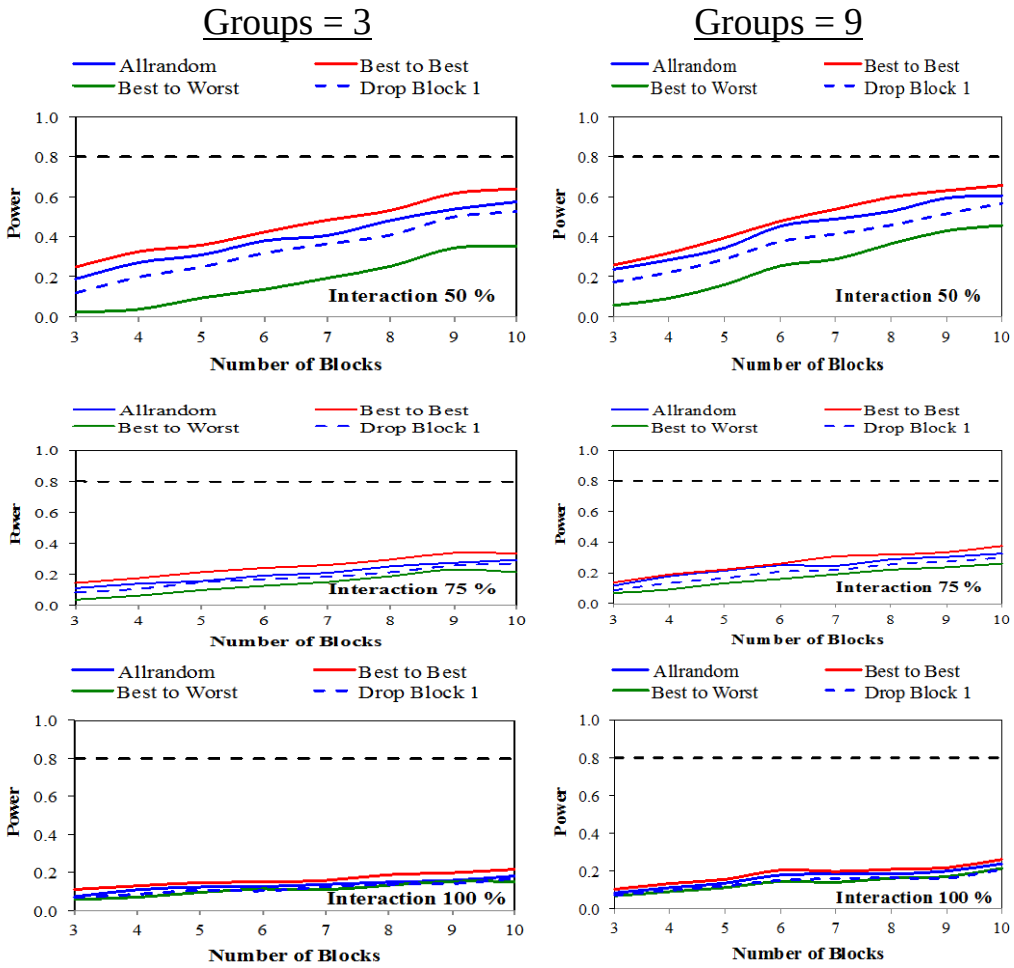


Fig 3. Effect of block number on the power of the pairwise comparisons between the best and the center treatment (half maximum treatment difference) for three and nine treatment groups based on the following experimental conditions: (i) **All Random** (solid blue line), where treatments were randomly assigned to plots in all blocks); (ii) **Best to Best** (red solid line), where the best treatment was assigned to the plot with the best inherent quality in Block 1; (iii) **Best to Worst** (green solid line), where the best treatment was assigned to the plot with the worst inherent quality in Block 1; and (iv) **Drop Block 1** (dashed blue line), where there was some systematic assignment of treatments to plots in block 1 but that block was dropped from the analysis. The interaction term = residual was set to either 100, 75, or 50% of the maximum absolute treatment effect to simulate experiments of increasing precision in 1000 simulated datasets for each condition.

SUMMARY

Not randomizing treatments in the first block of a field study conducted as an RCB seems to be a risky proposition. Under a **Best to Best** scenario there is an increased risk of committing a Type I error, i.e. declaring significance that are not warranted based on the underlying

experiment. Under a **Best to Worst** scenario, there is a tremendous loss of power, particularly in small but precise experiments. Not utilizing the first block (**Drop Block 1**) results in a loss of power that might exceed the damage incurred under a **Best to Worst** scenario in large.

The problem is that the experimenter rarely knows the true state of nature. The best course of action for the practitioner would seem to be to follow proper randomization procedures and to add an extra block to the experiment just for demonstration purposes.

ACKNOWLEDGEMENT

The first author thanks his current and former students of**AGRN 7080 Experimental Methods** for fruitful discussions. I continue to be amazed at your creativity and thoughtfulness. Your questions forced me to consider issues that I had not paid much attention to.

REFERENCES

Fisher R.A. 1926. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain 33:503-513.

Fisher R.A. 1966. The design of experiments. 8th edition. Hafner Publishing Company, New York.

Harville, D. A. 1975. Experimental Randomization: Who needs it? The American Statistician 29:27-31.

Hinkelmann, K. and O. Kempthorne. 2007. Design and Analysis of Experiments. Vol.Introduction to Experimental Design. 2nd edition. John Wiley & Sons, Hoboken, NJ, USA.

Lentner, M. and T. Bishop. 1993. Experimental design and analysis. Valley Book Company, Blacksburg, Va.

Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis. 2nd edition. Duxbury Press.