Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 1, 2021 (New Series), pp 95-106

# A Cost Effective Approach to the Design and Analysis of Multi–group Experiments

**Satya Prakash Singh**[1], **Shyamal D. Peddada**[2], **Ori Davidov**[3]
[1]*Department of Mathematics, Indian Institute of Technology Hyderabad, Telangana 502285, India*
[2]*Department of Biostatistics, University of Pittsburgh School of Public Health, 130 DeSoto Street, Pittsburgh, PA 15261, USA*
[3]*Department of Statistics, University of Haifa, Mount Carmel, Haifa 3498838, Israel*

Received: 15 August 2020; Revised: 29 October 2020; Accepted: 01 November 2020

**Abstract**

In this communication it is shown that employing statistical methods which account for constraints, inherent in some scientific problems, will often lead to a substantial reduction in the sample size required while simultaneously maintaining the power of the study and its scientific validity. In fact a 40%, or even higher, reduction in the required sample size is possible. These savings have the potential to impact individual labs and researchers and will translate to saving of millions of dollars annually for granting authorities and federal agencies such as the NIH.

*Key words*: Maxi-min designs; Order restricted inference; Power; Sample size.

**AMS Subject Classifications**: 62F30, 62K05

## 1. Introduction

Scientific research often requires testing of hypotheses comparing two or more experimental groups. The successful conduct of such investigations requires a study design appropriate for the scientific question at hand, a valid testing procedure for the hypothesis of interest, and an adequate sample size which guarantees suitable power. Sample size determination, or equivalently power calculations, are usually based on two sample and two–sided alternative hypotheses designed to test whether the mean response of the treatment group is different from that of the control group, *cf.*, Ryan (2013). Such calculations are simple and very widely used and numerous software packages, such as SAS and SPSS, have built–in routines for such tasks.

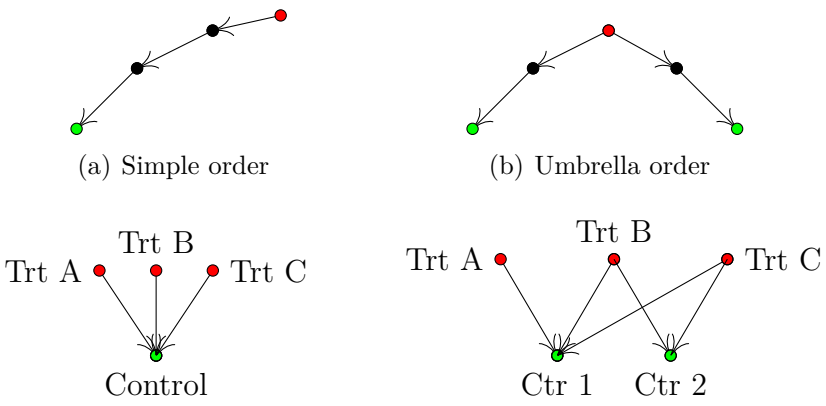
In many applications, such as dose–response studies or multi–drug trials, researchers may have a priori beliefs about the experimental groups. Such prior beliefs are usually based on earlier studies or an understanding of the underlying scientific phenomenon and are often formulated as mathematical inequalities or constraints, known as *order restrictions*.

Correponding Author: Satya Prakash Singh
Email: spsingh@math.iith.ac.in

For example, in a dose–response studies toxicologists may expect that the mean response increases (or decreases) with the dose of a chemical. This constraint is known as the *simple order*. Observational data are also often of this form. For example, in Spiegelhalter *et al.* (1999) the length of the ramus bone of 20 boys was measured at three equally spaced time points from ages 8 to 9. The question of interest was to know whether there was a significant growth spurt during the observed time period. In a time–course gene expression study, the mean expression of a gene may increase up to a certain point, reflecting its biological activity [Peddada *et al.* (2003)] and then decrease. This constraint is known as the *umbrella order*. In clinical trials, a researcher may be interested in demonstrating that the standard treatment is inferior to one of the new treatments, or, that a new treatment is at least as efficacious as the existing ones. This constraint is called the *tree order*. For example, Igari *et al.* (2014) compared the effect of various doses of cytisine on a dysporic–like state in rats. In some cases, the study design may include multiple control and multiple treatment groups. For example, the US National Toxicology Program (NTP) evaluates toxicity and carcinogenicity of chemicals using the concurrent control group as well as historical controls (which are controls collected from similar studies conducted by the NTP). This set up leads naturally to a *bipartite order* restriction [Kanno *et al.* (2003) and Peddada *et al.* (2007)].

The above mentioned order relations are represented graphically in Figure 1 by their corresponding *order graphs*. In each of the Figures, a circle represents a group mean, or more generally any other statistical parameter, and a pointed arrows implies an inequality among the two means or parameters. The *roots* of the order graph are the nodes with the largest means, whereas the *leaves* are the nodes with the smallest means. A variety of other constraints, or order restrictions, arise in applications. There exists over six decades of literature on this subject starting with the pioneering papers of Ayer *et al.* (1955), van Eden (1956) and Bartholomew (1959). Several books summarizing the work done in this field have also been published, e.g., Barlow *et al.* (1972), Robertson *et al.* (1988) and Silvapulle and Sen (2005).

In this article we highlight some important consequences of incorporating order restrictions in both the design and the analysis of experiments. Doing so addresses the scientific questions motivating the study in a principled manner. For if, for example, a standard two–sided test is applied in Figure 1(c), then a significant result tells us that there are differences among the treatments, it does not tell us that one of the treatments is superior to the control. Such inferences, however, are built–in into the procedures of constrained inference. Thus incorporating constraints in the analysis provides more meaningful inferences about the existence of an ordering among the experimental groups. In addition, using the constraints substantially improves efficiency. This means that we can expect considerable improvement in power and therefore the required sample sizes are reduced. In other words, failing to properly incorporate the order restrictions may lead to inflated costs of conducting studies, loss of power and inadequate scientific conclusions.




(a) Simple order (b) Umbrella order

(c) Tree order (d) Bipartite order

**Figure 1: Order graphs for some common order restrictions. Circles represent group means and a pointed arrow indicates an inequality among the means. Green circles correspond to leaves of the order graph and red circles to their roots. We refer to the leaves and roots as the extreme groups. The intermediate groups are designated by a black circle.**

## 2. Power, Order and Scientific Discovery

It is well known that tests tailored to accommodate order restrictions, called restricted or constrained tests [Silvapulle and Sen (2005)], are typically more powerful than their unconstrained counterparts. For example, consider the one–way analysis of variance (ANOVA) model

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where $Y_{ij}$ is the response of $j^{th}$ observation in $i^{th}$ treatment group, $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$ and the errors $\epsilon_{ij}$ are independent $\mathcal{N}(0, \sigma^2)$ random variables (RVs). For simplicity, and without any loss of generality, see Remark 2.2 in Singh and Davidov (2020), one may assume that $\sigma^2 = 1$ in which case the unconstrained likelihood ratio test (LRT) is of the form

$$T_n = \sum_{i=1}^{K} n_i (\bar{Y}_i - \hat{\mu}_i)^2$$

where $\bar{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$ for $i = 1, \ldots, K$ and $\hat{\mu}_i = \bar{Y} = N^{-1} \sum_{i=1}^{K} n_i \bar{Y}_i$ are the unrestricted estimators. Similarly the constrained LRT is given by

$$T_n = \sum_{i=1}^{K} n_i (\tilde{\mu}_i - \hat{\mu}_i)^2$$

where $\tilde{\mu}_i$ is the $i^{th}$ component of $\tilde{\boldsymbol{\mu}} = \text{argmax}\{\sum_{i=1}^{K} n_i (\bar{Y}_i - \mu_i)^2 : \mathbf{R}\boldsymbol{\mu} \geq \mathbf{0}\}$, the restricted maximum likelihood estimator of $\boldsymbol{\mu}$ which is assumed to satisfy a collection of linear inequalities $\mathbf{R}\boldsymbol{\mu} \geq \mathbf{0}$. It is well known that under the null the unconstrained LRT follows a chi–square distribution whereas the restricted LRT follows, what is known as, a chi–bar–square distribution [Silvapulle and Sen (2005)].

Figure 2 plots the power function of the standard (unconstrained) ANOVA test versus its constrained counterpart as a function of the per–group sample size under a balanced design. Clearly, the constrained test has higher power. Consequently the sample size required to guarantee a prespecified power is smaller when using a constrained test. At the 5% significance level and 80% power the unconstrained test requires 136 observations whereas the constrained test requires only 88 observations. It is evident that the reduction in sample sizes is a substantial 35%.

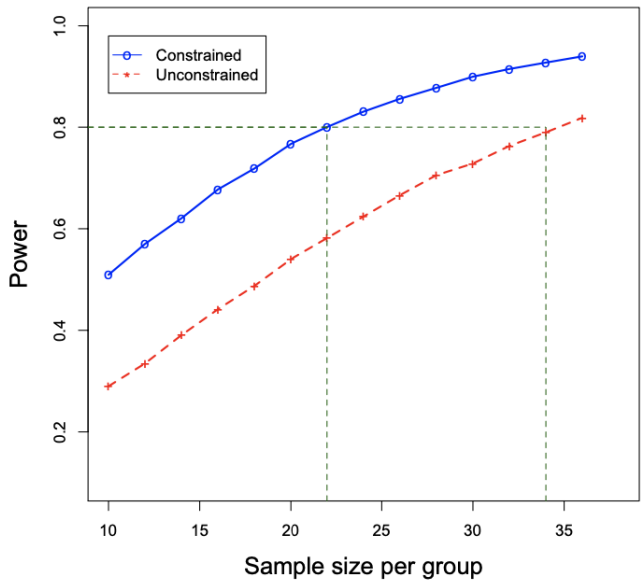



**Figure 2: The power of the constrained and unconstrained tests in the ANOVA setting. Data were simulated from normal populations with means 0, 0.25, 0.5 and 0.75, and unit standard deviation**

Even more dramatic examples are reported in the literature both in the context of ANOVA [Farnan *et al.* (2014)] as well as a variety of other settings [*e.g.*, Davidov and Herman (2012) and Rosen and Davidov (2017)]. A theoretical proof of the superiority of the restricted LRT is provided by Praestgaard (2012) and Davidov and Iliopoulos (2020). In the following we provide two examples from our own research which demonstrate that using methods which incorporate constraints helps to uncover clinically important features in the data which were missed by standard methods.

**Example 1:** Uterine fibroids, also known as uterine leiomyomata, are benign smooth muscle hormonally mediated tumors commonly found in pre–menopausal women. Nearly 70% of all women have these tumors. They cause pain, bleeding, urinary incontinence and pregnancy complications. The total annual cost of treating these tumors in US is estimated to be between 4 to 9 billion US dollars. The NIH, [cf. Peddada *et al.* (2008)], conducted a large prospective study of 72 pre–menopausal women (38 black and 34 white). Fibroid volumes were measured by MRI taken at baseline and at 3, 6, and 12 months, with at least two measurements per woman. African American women are known to have greater tumor

burden so a standard ANOVA–based analysis with an interaction between race and was performed. The interaction was found to be barely significant at $p = 0.05$. Since these tumors are known to be estrogen dependent, it is reasonable to hypothesize that tumor growth rates would decrease with age. This hypothesis was investigated in a recent re–analysis of these data [Peddada and Jelsema (2016)] using methods which account for order restrictions. A statistically significant decreasing trend in mean growth rates among whites ($p = 0.015$) but not among blacks ($p = 0.1880$) (Figure 3) was formally discovered. Thus, testing for order restrictions allows us to make a clinically important discovery that was not discovered by the standard ANOVA based methodology.

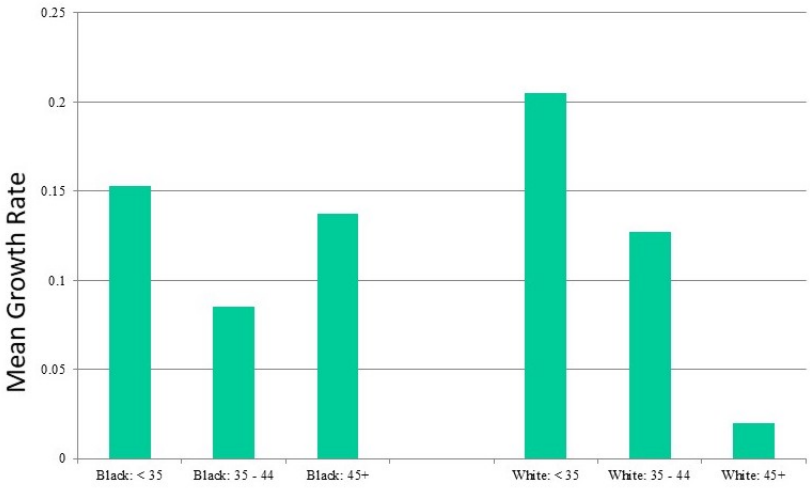



**Figure 3: Mammary gland fibroadenoma incidence in female rats**

**Example 2**: The Fish industry uses Malachite Green Chloride as an antifungal agent. The US National Toxicology Program (NTP) conducted a two year cancer bioassay with 48 female rats assigned to each of four dose groups of Malachite Green Chloride, namely, 0, 100, 300 or 600 parts per million. The incidence of mammary gland adenomas and pituitary gland adenoma–carcinomas are reported in Table 1. It is well–known that pituitary gland tumors may be associated with mammary gland tumors via the prolactin pathway [cf. McComb *et al.* (1984) and TR-527 (2005)]. Although these tumors are biologically dependent, the NTP analyzed them separately. The $p$–values for the corresponding trend tests were not significant, 0.113 for mammary gland adenoma and 0.162 for the pituitary gland adenoma–carcinomas. Davidov and Peddada (2011) developed a nonparametric multivariate ordered test that exploited the underlying dependence among the binary variables to test for trends in multivariate data. Using this constrained trend test Davidov and Peddada (2011) reanalyzed the NTP's Malachite Green Chloride data and discovered a significant increasing trend in both mammary gland adenomas as well as pituitary gland adeno–carcinomas, with a joint $p$–value of 0.025, suggesting a carcinogenic effect of Malachite Green Chloride on both tumors in a dose–related fashion. This finding reinforce the fact that the methods of constrained inference may discover finding not detected by standard methods.

Another advantage of using the methods of order restricted inference is that it relaxes

**Table 1: Tumor incidence rates of control and Malachite Green Chloride treated animals in the NTP study**

| Tumor type | Estimator | Control | 100 ppm | 300 ppm | 600 ppm |
|---|---|---|---|---|---|
| Mammary Gland | Unconstrained | 0.050 | 0.052 | 0.023 | 0.130 |
| | Constrained | 0.042 | 0.042 | 0.042 | 0.130 |
| Pituitary Gland | Unconstrained | 0.607 | 0.822 | 0.696 | 0.756 |
| | Constrained | 0.609 | 0.758 | 0.758 | 0.758 |

parametric assumptions. For example, suppose one is interested in the effect of an allele on a phenotype $Y$. It is very common to test for "trend" over the alleles $aa, Aa, AA$ by assigning scores $X = 0, 1$ and 2, respectively and performing a linear regression of $Y$ on $X$. The basic assumption, when using such a modelling framework, is that the change in the mean response from $aa$ to $aA$ is same as from $Aa$ to $AA$. Such assumptions may not be supported by the data and preclude the possibility of some non–linear but monotonic response such as in Figure 1(a). Such non–parametric curves are easily accommodated by constrained methods. In toxicology, it is also very common to perform linear regression–based tests such as the Cochran–Armitage trend test [Cochran (1954) and Armitage (1955)]. Some investigators use the exact dose as the explanatory variable and others use scores such as 1, 2, 3 and 4. When linearity is not justifiable considerable loss of power is to be expected [Peddada *et al.* (2005a)].

To summarize, incorporating the constraints in the analysis does not only lead to a beautiful and less restrictive statistical theory with improved operating characteristics, it may, much more importantly, help uncover biologically and clinically important results which standard methods fail to detect.

## 3. Optimal Design: Sample Size and Cost Efficiency

Smucker *et al.* (2018) emphasized that one should *customize the experiment for the setting instead of adjusting the setting to fit a classical design*, a comment that underscores the importance of carefully planned experiments. Recently, Singh and Davidov (2019) developed a rigorous framework for constructing optimal experimental designs which incorporate order restrictions. Their designs, known as Max–Min (MM) designs, maximize power under the worst possible (allowable) configuration in the alternative. They showed that the MM–design is of the form

$$\boldsymbol{\xi}_{\text{MM}} = |\mathcal{V}|^{-1} \sum_{(i,j)\in\mathcal{V}} \boldsymbol{\xi}_{ij}, \tag{1}$$

where $\boldsymbol{\xi}_{ij} = (\boldsymbol{e}_i + \boldsymbol{e}_j)/2$, $\boldsymbol{e}_l$ is the $l$th standard basis of $\mathbb{R}^K$ and $\mathcal{V}$ is the set of all maximal

pairs. A pair $(i, j)$ where $i \in \mathcal{R}$, the set of roots, and $j \in \mathcal{L}$, the set of leaves, is called a maximal pair if there is a path from $i$ to $j$. For more details see Singh and Davidov (2019).

The formula (1) is simple and easy to use. MM–designs for some common order restrictions such as the simple, tree, umbrella, and bipartite order (*cf.*, Figure 1) are given in Table 2 along with some other commonly used designs. It turns out that MM–designs allocate observations only to the leaves and roots of the order graph. In fact, if there are $N$ observations, then $N/2$ will be distributed among the leaves and $N/2$ among the roots. When there is more than one root the allocation among the roots is proportional to the degree of the root, *i.e.*, the number of paths to distinct leaves; and similarly for the leaves. Thus, the MM–design for the simple order will allocate $N/2$ observations to the two extreme groups. No observations are allocated to any of the intermediate groups. In the case of the umbrella order, the MM–design assigns $N/2$ observations to the peak of the umbrella and the remaining $N/2$ observations are equally divided among the extreme groups (first and last). Similar logic applies to the tree and bipartite order. We note that MM–designs do not allocate any observations to intermediate treatment groups, and thus do not allow any comparisons among them. This potential practical deficiency can be be addressed and rectified by using Singh and Davidov (2019)'s so called IUT–designs, which, for lack of space, we will not further discuss here.

**Table 2: The proportion of the observations allocated by the MM, Balanced, and Dunnetts' design are reported for the order relations depicted in Figure 1. The notation "-" indicates that there is no design to consider**

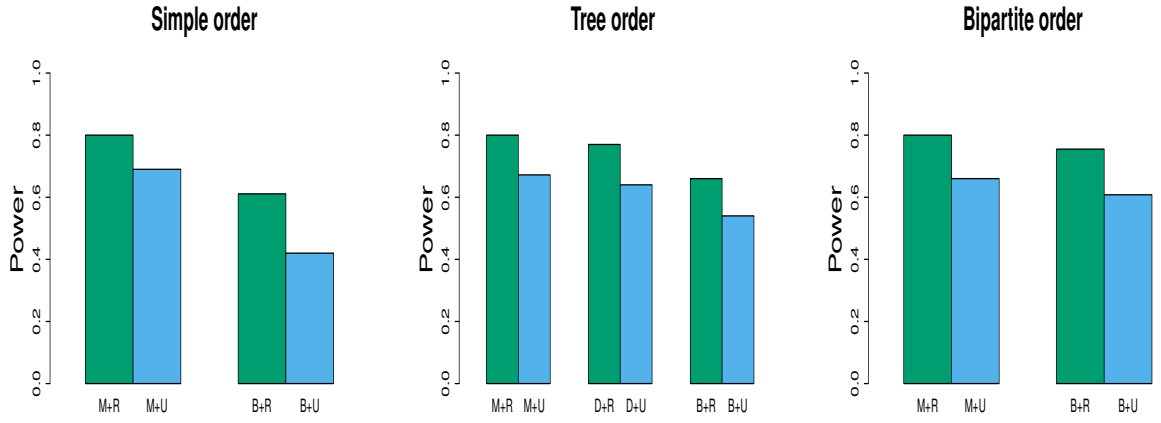
| | Order | | | |
|---|---|---|---|---|
| Design | Simple | Umbrella | Tree | Bipartite |
| MM | $(1/2, 0, 0, 1/2)$ | $(1/4, 0, 1/2, 0, 1/4)$ | $(1/2, 1/8, 1/8, 1/8, 1/8)$ | $(3/10, 2/10, 1/10, 2/10, 2/10)$ |
| Balanced | $(1/4, 1/4, 1/4, 1/4)$ | $(1/5, 1/5, 1/5, 1/5, 1/5)$ | $(1/5, 1/5, 1/5, 1/5, 1/5)$ | $(1/5, 1/5, 1/5, 1/5, 1/5)$ |
| Dunnett | - | - | $(1/3, 1/6, 1/6, 1/6, 1/6)$ | - |

## 4. Results

The benefits associated with MM–designs were assessed by simulations using data from the published scientific literature. Simulations under the simple order were based on the data of Spiegelhalter *et al.* (1999), whereas the simulations for the tree and bipartite orders were based on data from Igari *et al.* (2014) and Kanno *et al.* (2003), respectively. The substantive scientific problems investigated in these papers were already briefly described. For simplicity, the simulated data is normally distributed with mean values and standard deviations as reported in Table 3. For each ordered alternative, we performed an unconstrained and restricted likelihood ratio test. The results of the simulation study, based on $10^5$ simulation runs, are summarized in Figures 4 and 5 which display powers and sample sizes, respectively.

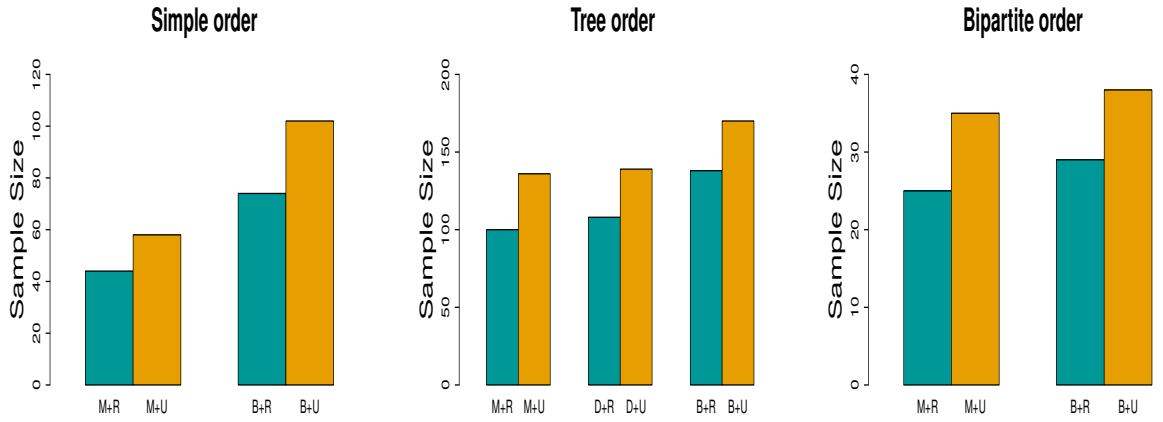
**Table 3: A brief summary of the results of Spiegelhalter *et al.* (1999), Igari *et al.* (2014) and Kanno *et al.* (2003). We report on the group size, sample mean and standard deviation as well as the pooled standard deviation (PSD). For the tree order, treatment 1, serves as the control and is compared to the remaining treatments. In the bipartite case, treatments 1 and 2 are the controls. Treatment 1 is compared to treatments 3, 4, and 5, whereas treatment 2 is compared to the 4 and 5**

| | Treatment Group | | | | | |
|---|---|---|---|---|---|---|
| Order | 1 | 2 | 3 | 4 | 5 | PSD |
| Simple | $48.66 \pm 2.52$<br>20 | $49.62 \pm 2.54$<br>20 | $50.57 \pm 2.63$<br>20 | | | 2.56 |
| Tree | $97.6 \pm 10.39$<br>12 | $101.6 \pm 8.66$<br>12 | $102.2 \pm 4.50$<br>12 | $103.4 \pm 10.04$<br>12 | $105.9 \pm 14.90$<br>12 | 10.26 |
| Bipartite | $29.5 \pm 2.95$<br>6 | $30.0 \pm 2.30$<br>6 | $32.2 \pm 3.13$<br>6 | $34.8 \pm 3.48$<br>6 | $31.8 \pm 4.34$<br>6 | 3.31 |




(a) (b) (c)

**Figure 4: Power comparisons between the Maxi-Min (MM), Balanced (B) and Dunnett's (D) designs when applied with both the unrestricted and restricted test. For example MM+R is the power of the MM design with a restricted test**




(a) (b) (c)

**Figure 5: Sample sizes required for 80% power under the Maxi-Min (MM), Balanced (B) and Dunnett's (D) designs when applied with both the unrestricted and restricted test. For example MM+R is the sample size required by the MM design with a restricted test**

Our simulation study shows, as previously noted, that using the restricted test is always better than using the unrestricted test. It is clear that the MM–design results in improved power relative to the balanced and other designs irrespective of the test being used. For example, Figure 5(b) shows that the MM design analyzed by a restricted test requires a sample of approximately 100 subjects whereas the balanced design with and standard test requires 170 subjects.

## 5. Summary

This communication shows that accounting for constraints, which occur naturally in a wide variety of scientific investigations, has a huge dividend. In particular it is shown, using examples from the literature, that a substantial reduction in the sample size is achieved when both designing and analyzing data using methods that account for constraints. It is emphasized that the largest benefits are achieved when an experiment is both designed and analyzed using order based methods. The reduction in the required sample sizes, or equivalently the increase in power [Singh and Davidov (2019)], is nothing but phenomenal suggesting that the routine use of order based methods, when appropriate, will result in much more economical and efficient designs. In fact, since in many experimental sciences a substantial portion of the budget is devoted to acquiring a large as possible sample, researchers, pharmaceuticals, granting agencies and others may save millions of dollars on data collection and do much more with a fixed budget. In addition, if the study involves biological samples from animal or human subjects, then these methods would require the participation of fewer animals or human subjects.

It is surprising that although the methodology we describe here traces its roots to the late 1950's it has not had a major impact on data collection and analysis in the sciences.

There are several reasons for that. The first is that the focus of statisticians working in this area had been largely theoretical with little concern for practical issues such as cost reductions. Secondly, appropriate software for analyzing data using these constrained inference based methods were not available until recently. Software such as ORIOGEN [Peddada *et al.* (2005b)] and CLME [Peddada and Jelsema (2016)] have taken the important first steps in this direction and are gaining popularity among users. Finally, the development of experimental designs [Singh and Davidov (2019)] which capitalize on scientific constraints is a recent development with potential far reaching consequences.

## Acknowledgements

The research leading to this paper was conducted (mostly) while the first author was a postdoctoral fellow at the University of Haifa, Israel. The work of Satya Prakash Singh was further supported by SEED grant (Indian Institute of Technology, Hyderabad) and is gratefully acknowledged. The work of Ori Davidov is supported by the Israeli Science Foundation Grant No. 456/17 and is gratefully acknowledged.

## References

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, **11**(**3**), 375–386.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, **26**(**4**), 641–647.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, **46**(**1-2**), 36–48.

Cochran, W. G. (1954). Some methods for strengthening the common $\chi^2$ tests. *Biometrics*, **10**(**4**), 417–451.

Davidov, O. and Herman, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *Journal of the Royal Statistical Society, **B***, **74**(**5**), 825–847.

Davidov, O. and Iliopoulos, G. (2020). Semiparametric ordered inference for conditional distributions. *Submitted*.

Davidov, O. and Peddada, S. D. (2011). Order restricted inference for multivariate binary data with application to toxicology. *Journal of the American Statistical Association*, **106**(**496**), 1394–1404.

Farnan, L., Ivanova, A. and Peddada, S. D. (2014). Constrained inference in biological sciences: Linear mixed effects models under constraints. *PLOS ONE*, **9**(**1**), e84778.

Igari, M., Alexander, J. C., Ji, Y., Xiaoli, Q., Papke, R. L. and Bruijnzeel, A. W. (2014). Varenicline and cytisine diminish the dysphoric-like state associated with spontaneous nicotine withdrawal in rats. *Neuropsychopharmacology*, **39**, 445–455.

Kanno, J., Onyon, L., Peddada, S., Ashby, J., Jacob, E. and Owens, W. (2003). The oecd program to validate the rat uterotrophic bioassay. phase 2: Dose-response studies. *Environmental Health Perspectives*, **111**(**12**), 1530–1549.

McComb, D. J., Kovacs, K., Beri, J. and Zak, F. (1984). Pituitary adenomas in old sprague-dawley rats: A histologic, ultrastructural, and immunocytochemical study. *Journal of the National Cancer Institute*, **73**(**5**), 1143–1166.

Peddada, S., Lobenhofer, E., Li, L., Afshari, C. and Umbach, D. (2003). Gene selection and clustering for time-course and dose response microarray experiments using order-restricted inference. *Bioinformatics*, **19**(**7**), 834–841.

Peddada, S. D., Dinse, G. and Haseman, J. (2005a). A survival-adjusted quantal response test for comparing tumor incidence rates. *Journal of the Royal Statistical Society, **C***, **54**(**1**), 51–61.

Peddada, S. D., Dinse, G. and Kissling, G. (2007). Incorporating historical control data when comparing tumor incidence rates. *Journal of the American Statistical Association*, **102**(**480**), 1212–1220.

Peddada, S. D., Harris, S., Zajd, J. and Harvey, E. (2005b). ORIOGEN: order restricted inference for ordered gene expression data. *Bioinformatics*, **21**(**20**), 3933–4.

Peddada, S. D. and Jelsema, C. M. (2016). CLME: An R package for linear mixed effects models under inequality constraints. *Journal of Statistical Software*, **75**(**1**), 1–32.

Peddada, S. D., Laughlin, S. K., Miner, K., Haneke, K., Vahdat, H. L., Semelka, R. C., Kowalik, A., Armao, D., Davis, B. and Baird, D. D. (2008). Growth of uterine leiomyomata among premenopausal black and white women. *Proceedings of the National Academy of Sciences*, **105**(**50**), 19887–19892.

Praestgaard, J. (2012). A note on the power superiority of the restricted likelihood ratio test. *Journal of Multivariate Analysis*, **104**(**1**), 1–15.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

Rosen, S. and Davidov, O. (2017). Ordered regressions. *Scandinavian Journal of Statistics*, **44**(**4**), 817–842.

Ryan, T. P. (2013). *Sample Size Determination and Power*. John Wiley and Sons, Hoboken, NJ.

Silvapulle, M. and Sen, P. (2005). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Hoboken, NJ, New York.

Singh, S. P. and Davidov, O. (2019). On the design of experiments with ordered treatments. *Journal of Royal Statistical Society, **B***, **81**(**5**), 881–900.

Singh, S. P. and Davidov, O. (2020). On Bayes and Nash experimental designs for hypothesis testing problems. *Electronic Journal of Statistics*, **14**(**2**), 3976–4003.

Smucker, B., Krzywinski, M. and Altman, N. (2018). Optimal experimental design. *Nature Methods*, **15**, 557–560.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1999). *BUGS 0.5 Examples Volume 2, MRC Biostatistics Unit*. Institute of Public Health. Robinson Way, Cambridge CB2 2SR.

TR-527, N. (2005). NTP technical report on the toxicology and carcinogenesis studies of malachite green chloride and leucomalachite green. Technical report.

van Eden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen A*, **59**, 444–455.