J. N. Srivastava and Experimental Design

J. P. Morgan *[,a], Subir Ghosh [b], and A. M. Dean [c,d]

[a] *Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439*

[b] *Department of Statistics, University of California, Riverside, Riverside, CA 92521, USA*

[c] *Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA*

[d] *Department of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK*

Abstract

J. N. Srivastava was a tremendously productive statistical researcher for five decades. He made significant contributions in many areas of statistics, including multivariate analysis and sampling theory. A constant throughout his career was the attention he gave to problems in discrete experimental design, where many of his best known publications are found. This paper focuses on his design work, tracing its progression, recounting his key contributions and ideas, and assessing its continuing impact. A synopsis of his design-related editorial and organizational roles is also included.

MSC 2000 subject classification. 62K15.

Key words: association scheme, balanced array, coherent configuration, fractional factorial design, row-column design, screening design, search design

* Corresponding author. Tel: +01-540-231-9701; fax: +01-540-231-3863
Email addresses: `jpmorgan@vt.edu` (J. P. Morgan), `ghosh@ucr.edu` (Subir Ghosh), `dean.9@osu.edu` (A. M. Dean).

1 Academic history

Jagdish Narain Srivastava was born in Lucknow, India, on June 20, 1933. His mother, Madhuri Devi, passed away when he was two years old; his father, Mahabir Prasad, married again and young Jagdish was raised by his stepmother, Lila. Though his family struggled financially, his father emphasized English and mathematics at an early age and enrolled Jagdish in the fifth grade when he was only eight years old. Always younger than his classmates, Jagdish was nonetheless a superior student. He completed Bachelor's and Master's degrees in Mathematical Statistics from Lucknow University, the latter in 1954 with highest standing in the Master's examination. Now a young man needing to support not just himself but family members as well, Jagdish immediately sought employment. He worked at institutions in Lucknow and Delhi in agricultural statistics from 1954 to 1959, during which time he continued his study of statistical theory. In 1958 he traveled to Calcutta for a battery of rigorous written and oral exams, leading to the Statistician's Diploma from the Indian Statistical Institute. His examiners included D. B. Lahiri, P. C. Mahalanobis and C. R. Rao, among others.

Srivastava was very proud of this diploma as well as the Indian Statistical Institute, which was thriving under Mahalanobis' dynamic leadership. Also inspired by the deeply creative contributions of Rao, then Director of the Research and Training School at ISI, Srivastava grew determined to make his own mark in statistics. His papers with K. Kishen (Kishen and Srivastava (1959a), Kishen and Srivastava (1959b)) developing a mathematical theory for confounding of fractional factorials were his first publications in statistical theory. R. C. Bose, who had earlier worked with Kishen (notably on the influential paper Bose and Kishen (1940)), and who was by then well known throughout the statistical world for many fundamental contributions in experimental design, had moved to the United States in 1947, accepting a faculty position at the University of North Carolina (UNC) at Chapel Hill in 1949. At Kishen's encouragement, and with financial constraints making it difficult for him to pursue a doctoral degree at ISI, Srivastava also made the leap to the U.S. in hopes of working with Bose. He gained admission to UNC as a graduate student in 1959 during the early golden period of its Statistics Department, bringing him in contact not just with Bose but with other giants such as H. Hotelling, W. Hoeffding, and S. N. Roy. He completed his thesis under Bose in late 1961 (the doctoral degree was officially awarded in 1962), then worked as a Post Doctoral Researcher with Bose and Roy until 1963. In 1963 he joined the faculty at the University of Nebraska (UN) as Associate Professor, was soon promoted to Full Professor, and from there moved in 1966 to the Department of Statistics at Colorado

State University (CSU). Srivastava was instrumental in bringing Bose to CSU in 1971 where, following his retirement from UNC, Bose was awarded a joint professorship in the Departments of Mathematics and of Statistics, cementing CSU as a world leader in combinatorial design. In 1991 Srivastava received a joint appointment in the CSU Philosophy Department, a result of his pursuit of interests in quantum mechanics, mathematical logic, and spirituality (see, for example, Srivastava (2001)). Srivastava retired in early 2000 and became a College of Natural Sciences Research Professor at CSU. He then moved to Northern California with his wife Usha, to be close to their youngest son Ashok and his family. He remained an active researcher up to his death on November 18, 2010.

Srivastava was an elected fellow of both ASA and IMS, an elected member of ISI, a foundation fellow of the Institute of Combinatorics and Applications, and a Fellow of the Third World Academy of Sciences. He was president of the Forum for Interdisciplinary Mathematics as well as the International Indian Statistical Association. Srivastava directed 16 doctoral dissertations during his career, 13 at CSU and three at UN (see The Mathematics Genealogy Project at *genealogy.math.ndsu.nodak.edu* for a complete list). A co-author of this paper, S. Ghosh, was one of his CSU doctoral students.

2 The Early Years: Fractions and Balanced Arrays

Srivastava's dissertation, under the direction of R. C. Bose as indicated above, was titled *Contributions to the Construction and Analysis of Designs*. From 1963 to 1966, Srivastava and Bose published five articles together, whose titles we list to aid exposition:

(1) Mathematical theory of factorial designs, Bose and Srivastava (1963)
(2) Analysis of irregular factorial fractions, Bose and Srivastava (1964a)
(3) Multidimensional partially balanced designs and their analysis, with applications to partially balanced factorial fractions, Bose and Srivastava (1964b)
(4) Some economic partially balanced 2^m factorial fractions, Srivastava and Bose (1966)
(5) On a bound useful in the theory of factorial designs and error correcting codes, Bose and Srivastava (1964c)

Srivastava made important contributions to coding theory beginning with paper (5), but our purpose here is to discuss experimental design. Papers (1)–(4) followed a common theme which was to be richly developed by Srivastava over the coming decade. Like much design work today, all assumed uncorrelated observations with common variance σ^2. Perspective on what was being done in those papers is gained from this quote, from (4):

> "In the beginning, the theory largely concerned itself with orthogonal fractions, in which the estimates of the effects of interest to us are all uncorrelated. However the number of assemblies required in such fractions is rather large, i.e. they are uneconomic. Attention has therefore lately shifted more to the consideration of fractions which are **economic** but which may give rise to correlated estimates. We shall call such fractions as **irregular**."

The bold type is ours. The concepts of run-size economy and irregular/nonregular fractions are still hotly pursued today, but their roots are fifty years in the past.

Here is another quote, from Harter (1963) in a review of (1): "In addition, these partial replicates can in general be analyzed without resort to an electronic computer." While not something we would give much thought to today, this quote plainly reveals how different was the mind set in design research 50 years ago. More directly, Harter's statement implies that the designs that Srivastava was exploring could be evaluated analytically for their statistical performance, a crucial fact for work to follow, as we will explain.

Much work was then being done on the notion of partial balance of designs in conjunction with association schemes as first introduced by Bose and Nair (1939). Indeed, this approach to discrete design would continue apace for another decade. All of the papers (1)–(4) were built on a partial balance framework, using what Srivastava and Bose called the *multidimensional partially balanced association scheme* (MDPB scheme) defined as follows:

- There are m sets of objects $S_1, S_2, \dots, S_m$
- Set i contains n_i objects: $|S_i| = n_i$
- Fix any (not necessarily distinct) $i, j, k \in \{1, 2, \dots, m\}$

(a) Given $x_{ia} \in S_i$, there is a partition of S_j into g_{ij} classes, where class q contains $n_q(i,j)$ objects termed the q^{th} associates of x_{ia} in S_j $(q = 1, \dots, g_{ij})$. The number g_{ij} does not depend on the particular $x_{ia} \in S_i$ selected.

(b) The relationship is symmetrical, that is, if $x_{jb} \in S_j$ is the q^{th} associate of $x_{ia} \in S_i$, then $x_{ia} \in S_i$ is the q^{th} associate of $x_{jb} \in S_j$. Consequently $g_{ij} = g_{ji}$.

(c) Given distinct objects $x_{ia} \in S_i$, $x_{jb} \in S_j$ which are q^{th} associates, let $\Omega(x_{ia}, u, k)$ denote the set of u^{th} associates of x_{ia} in S_k, and $\Omega(x_{jb}, w, k)$ the set of w^{th} associates of x_{jb} in S_k. Then

$$|\Omega(x_{ia}, u, k) \cap \Omega(x_{jb}, w, k)| = p^q_{u,w}(i, j, k)$$

does not depend on the particular objects x_{ia}, x_{jb} chosen.

In brief, partition a set of objects into m classes $S_1, S_2, \ldots, S_m$. There is an MDPB scheme on these objects if (i) there is an ordinary association scheme (as defined in Bose and Shimamoto (1952)) for elements of any one class S_i, (ii) there is an association relationship for any class S_i relative to any other class S_j, and (iii) there are set-dependent triangular parameters for any three (not necessarily distinct) sets.

The terminology here has become nonstandard, for standard usage is now "design T is partially balanced with respect to association scheme S." That is, "partial balance" is a design property, while association schemes exist independently of designs. Terminology aside, the MDPB scheme was a significant generalization of the original Bose-Shimamoto concept (which itself formalized Bose and Nair (1939)). It is what is now in the combinatorial mathematics community called a *coherent configuration*, introduced by Higman (1970) as an approach to, in the words of Brouwer (2011), "group theory without groups." Interestingly Higman, despite publishing on the latter topic for over two decades, does not seem to have ever been aware of Srivastava's work, nor vice-versa. For Srivastava, these schemes were the key to an analytic understanding of large classes of irregular fractions.

Write an N-run design for m factors as an $N \times m$ array for which column i represents factor F_i with levels in S_i. The design is partially balanced with respect to the MDPB scheme if

- there is an MDPB association scheme on $S_1, \ldots, S_m$
- each level of F_i occurs on $N/|S_i|$ runs
- if $x_{ia} \in S_i$ and $x_{jb} \in S_j$ are q^{th} associates, then the number of runs with (x_{ia}, x_{jb}) is $\lambda_q(i, j)$

For instance, a traditional PBIBD with v treatments in b blocks of size k, whose dual is also a PBIBD (and which consequently has an association scheme defined on the blocks), can be expressed as a $bk \times 2$ array that is a MDPB design. The $N = bk$ experimental units are the runs, the two columns correspond to treatments (S_1) and blocks (S_2), and a 2-class association between S_1 and S_2 is defined by a treatment being a first associate of a block if it occurs in that block, and second associate otherwise. The main idea was to capture the partial balance notion *for all cross-classifications*, regardless of the number of blocking and treatment factors. While Srivastava did publish work dealing with multiple blocking factors (e.g. Srivastava and Anderson (1971)), his *main emphasis* was on unblocked fractions.

Any ordinary association scheme generates association matrices which collectively form a linear associative algebra, called the Bose-Mesner algebra. A very interesting mathematical feature of the MDPB scheme is that it generates a *non-commutative* algebra, whereas the

Bose-Mesner algebra is commutative. Nonetheless Srivastava and Bose in paper (3) were able to construct an algorithm for inverting any matrix in the algebra, at that time a key consideration for data analysis. Parts of each of the papers (1)–(4) were devoted to analysis of various design classes, notably including irregular fractions.

There was also a nontrivial design construction problem to achieve partial balance in fractions. Srivastava was able to obtain suitable designs by exploiting an associated structural property for variances of estimated contrasts. Define a fraction to be *completely balanced* if the variance matrix of the estimates is invariant to permutation among the factors. Thus for 2^m experiments estimating main effects and two-factor interactions, with an obvious notation,

$$\textsc{Var}(\hat{F}_1) = \textsc{Var}(\hat{F}_2) = \cdots = \textsc{Var}(\hat{F}_m)$$

and

$$\textsc{Cov}(\hat{F}_{12}, \hat{F}_{13}) = \textsc{Cov}(\hat{F}_{12}, \hat{F}_{14}) = \cdots = \textsc{Cov}(\hat{F}_{m-2,m-1}, \hat{F}_{m-2,m})$$

but not necessarily

$$\textsc{Cov}(\hat{F}_{12}, \hat{F}_{13}) = \textsc{Cov}(\hat{F}_{12}, \hat{F}_{34}).$$

Complete balance is a natural statistical expression of equal interest in the factors.

How does one achieve complete balance in a fraction, short of an orthogonal array? Let $T_{N\times m}$ be a completely balanced design. Let V_T be the variance matrix of estimated effects using T. Then $V_T = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ for suitable model matrix $\mathbf{X}$. Complete balance is a permutation invariance property of $(\mathbf{X}'\mathbf{X})^{-1}$ and hence the same permutation invariance holds for $\mathbf{X}'\mathbf{X}$. Thus complete balance is directly expressible in terms of the design matrix. This led Srivastava to the idea of partially balanced arrays, or PBAs.

A *partially balanced array* of strength t on s symbols, PBA$(N, s^m; t)$, is an $N \times m$ array for which every permutation of any particular t-vector $(i_1, i_2, \ldots, i_t)$ on the s symbols occurs with the same frequency in each and every $N \times t$ subarray. For 2-level arrays ($s = 2$), taking the symbols as 0 and 1, that frequency may be denoted by μ_i for those t-vectors containing exactly i ones. An example of a 2-level, strength 4 array is displayed in Table 1. PBAs of strength t are necessarily PBAs of strength $t' < t$; the array in Table 1 is also PBA$(5, 2^4; 3)$ having $(\mu_0,\ \mu_1,\ \mu_2,\ \mu_3) = (1, 0, 1, 1)$ and PBA$(5, 2^4; 2)$ having $(\mu_0,\ \mu_1,\ \mu_2) = (1, 1, 2)$.

Table 1
PBA$(5, 2^4; 4)$ having $(\mu_0,\ \mu_1,\ \mu_2,\ \mu_3, \mu_4) = (1, 0, 0, 1, 0)$. Rows are runs of a non-orthogonal, completely balanced, resolution III fraction.

0	0	0	0
0	1	1	1
1	0	1	1
1	1	0	1
1	1	1	0

Srivastava found the nexus of complete balance and MDPB schemes in PBA's. Start with a 2-level fraction with the standard parameterization for the linear model in terms of a grand mean and main effect and interaction contrasts. The columns of the design matrix $\mathbf{X}$ then correspond to the intercept and the various contrast parameters, which may be thought of as objects on which an association scheme may be defined. It follows that $\mathbf{X}'\mathbf{X}$ for a 2-level partially balanced array is in the linear associative algebra induced by an MDPB association scheme, where the sets S_i correspond to different orders of factorial effects. Specifically, for a 2^m experiment the S_i are the effects of order i, and the association of two effects expressed as words $W_1 \in S_i$, $W_2 \in S_j$ is the number of letters in the product word W_1W_2 (and indexed by (i, j)). An example scheme appears in Figure 1. An important consequence is that $\mathbf{X}'\mathbf{X}$ for a 2-level PBA can be inverted analytically to obtain variances and so to evaluate a design's efficacy. As we will explain, Srivastava and his students exploited this in a long series of papers, but first we consider the genesis of PBA's.

Partially balanced arrays were originally defined by I. M. Chakravarti in Chakravarti (1956), in work arising from his doctoral dissertation under C. R. Rao. Chakravarti pointed out the relatively simple form for $\mathbf{X}'\mathbf{X}$ but did not explicitly justify why he chose the PBA name. It seems quite natural in light of the MDPB scheme. Srivastava, as he stated in Srivastava (1984) (page 99), independently discovered the PBA concept five years after Chakravarti by approaching from the partial balance perspective.

These arrays are known today by the simpler "balanced array" (BA), a term coined in Srivastava and Chopra (1971b). Srivastava and Chopra justified the terminological change by pointing out that the 2-level arrays of strength at least two are a generalization of BIBDs (the $N \times m$ array is also the $N \times m$ incidence matrix of a pairwise balanced block design with N blocks and m treatments). This is not a wholly satisfactory explanation; perhaps equally compelling justification lies in the original notion of complete balance in the variance matrix. In any case it is a simpler name and profitably serves to disassociate these widely

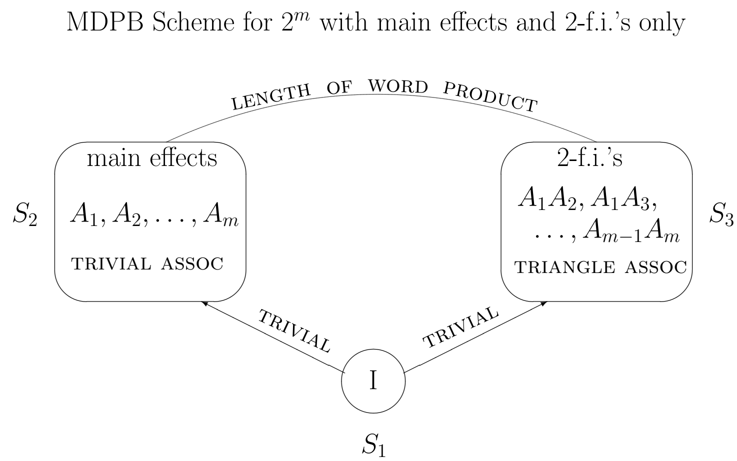

Fig. 1. A MDPB association scheme. The intercept, the main effects, and the 2-factor interactions form three groups S_1, S_2, S_3. There are two associate relationships of members of S_2 with members of S_3, depending on whether a main effect shares, or does not share, its factor with a 2-factor interaction. There are two associate relationships among members of S_3 as two 2-factor interactions share, or do not share, a common factor (this is the triangular association scheme within S_3). Aside from self-association, there are seven distinct associate relationships in this scheme.

applicable arrays (more on this below) from Bose and Nair's partial balance.

Chakravarti was not impressed by the name change. In a review (Chakravarti (1974)) of Chopra and Srivastava (1974) he wrote:

> "This reviewer introduced partially balanced arrays and gave methods of construction of such arrays for the first time [Sankhya 17 (1956), 143–164]. Subsequent papers on the subject were published by the reviewer.... It seems incongruous to call the same object by a different name fifteen years later."

If you knew Chakravarti then you know how strong a statement this was for him. However he later published two papers (Suen and Chakravarti (1985), Suen and Chakravarti (1986)) in which the BA name was used without mention of PBAs, so he, too, seems to have accepted the new name, by this time firmly ensconced. We are not aware of Chakravarti having otherwise pursued this topic.

Table 2
Quantities c_i for investigating BAs of strength 4.

$$
\begin{aligned}
c_1 =& 3\gamma_1 + (3m-5)\gamma_3 + m''\gamma_5 \\
c_2 =& 3\gamma_1^2 + 2\gamma_1\gamma_3(3m-5) + 2m''\gamma_1\gamma_5 + \frac{(m-1)(3m-8)}{2}\gamma_3^2 + m''(m-1)\gamma_3\gamma_5 - m\gamma_2^2 \\
& -\frac{m-1}{2}(2\gamma_2 + (m-2)\gamma_4)^2 \\
c_3 =& \gamma_1^3 - \binom{m}{2}(m-1)\gamma_3^2 + 2\frac{(m-1)(3m-8)}{2}\gamma_1\gamma_3^2 + (3m-5)\gamma_1^2\gamma_3 + m''(m-1)\gamma_1\gamma_3\gamma_5 \\
& + m''\gamma_1^2\gamma_5 - m\gamma_1\gamma_2^2 - mm''\gamma_2^2\gamma_5 + 2m\gamma_2^2\gamma_3 + m(m-1)(m-2)\gamma_2\gamma_3\gamma_4 \\
& -\frac{m-1}{2}\gamma_1(2\gamma_2 + (m-2)\gamma_4)^2
\end{aligned}
$$

Together with his student D. V. Chopra, Srivastava wrote a series of papers studying BAs as factorial designs. They concentrated on two major questions: (i) existence, and (ii) optimality within the class of all BAs. Both questions are somewhat delicate and required considerable ingenuity to resolve over a practical range for design size. Part of the reason for this is that unlike orthogonal arrays (OAs), which are a special case, BAs are much more widely existent. For instance (see e.g. Srivastava (1978), page 303) there is a BA(N,2^m; t) of strength $t \geq 4$ for every number of runs $N \geq 1 + \binom{m}{1} + \binom{m}{2}$. So if there are enough observations to estimate all main effects and 2-factor interactions, it is possible to estimate them with a balanced array and thus to have complete balance of the variance matrix. Unfortunately this simple existence result offers no clue as to what is the *best* strength four array for given m and N, for there can be many arrays of a given size.

Srivastava attacked the confluence of the optimality and existence questions by seeking nonnegativity conditions, as functions of the 4-vector frequencies μ_i, on the eigenvalues of the (necessarily nonnegative definite) $\mathbf{X'X}$ matrix corresponding to BA(N, 2^m; 4); see Srivastava (1972) and Srivastava and Chopra (1971b). For a 2-level strength 4 array BA($N, 2^m$) define $m' = m(m-3)/2$, $m'' = (m-2)(m-3)/2$, $\gamma_1 = \mu_0 + 4\mu_1 + 6\mu_2 + 4\mu_3 + \mu_4$, $\gamma_2 = (\mu_4 - \mu_0) + 2(\mu_3 - \mu_1)$, $\gamma_3 = \mu_4 - 2\mu_2 + \mu_0$, $\gamma_4 = (\mu_4 - \mu_0) - 2(\mu_3 - \mu_1)$, and $\gamma_5 = \mu_0 - 4\mu_1 + 6\mu_2 - 4\mu_3 + \mu_4$. Now consider the three quantities listed in Table 2. Using these, a set of necessary conditions for existence of BA(N,2^m; 4) is

$$
\begin{aligned}
(m-1)(\mu_1 + \mu_3) &\geq 2(m-5)\mu_2 \\
(m-4)\mu_2^2 &\leq \mu_2(\mu_1 + \mu_3) + (m-2)\mu_1\mu_3 \\
c_1, c_2, c_3 &\geq 0
\end{aligned}
\tag{1}
$$

Table 3
Determining the A-best fractions in the BA$(N, 2^m; 4)$ class

m	N	cite
4	$11 - 28$	Srivastava and Chopra (1971a)
4	$29 - 64$	Chopra (1975c)
5	$16 - 32$	Srivastava and Chopra (1971a)
5	$33 - 64$	Chopra (1975c)
6	$22 - 40$	Srivastava and Chopra (1971a)
6	$41 - 64$	Chopra (1975c)
7	$29 - 42$	Chopra and Srivastava (1973b)
7	$43 - 48$	Chopra and Srivastava (1975)
7	$49 - 55$	Chopra and Srivastava (1973a)
7	$56 - 68$	Srivastava and Chopra (1974)
8	$37 - 51$	Chopra and Srivastava (1974)
8	$52 - 65$	Chopra (1975a), Chopra (1975b)

Working with the inequalities (1) and directly with the eigenvalues of the information matrix, Srivastava and Chopra wrote a series of papers in the first half of the 1970s determining A-optimal BAs of strength 4 (resolution V); see Table 3. A few minor corrections are reported in Chopra, Kipngeno, and Ghosh (1986). Srivastava (1970) further proved that completely balanced main effects plans for 2-level factors are *equivalent to* BA$(N,2^m; 4)$s.

Though it was not possible at that time, we can now examine how the best BAs perform within the wider class of all competing designs. A-efficiencies for a sampling of A-best, strength 4 balanced arrays for seven and eight factors, relative to the overall best designs found by search, is displayed in Table 4. For seven factors the designs are all excellent. The small sacrifices in terms of A-efficiency seen in some cases will in most experimental settings be more than compensated for by the complete balance. On the other hand, the eight-factor designs do not fare so well. The greater dispersion in the frequencies μ_i may be interpreted as greater departure from an idealized orthogonal design, inflating variances as the cost of balance. Indeed, as pointed out by Nguyen and Miller (1997), in some cases the best BA fractions can be dominated by non-balanced designs, in the sense that *every* effect is estimated with smaller variance by the latter. In such cases BA fractions are clearly unacceptable.

The point of these observations is not to disparage the balanced array work. It is not surprising that with the passage of time and rise of computational power some better designs have been found. The point is to display and understand what was, in its time, very powerful work

Table 4
Selected A-best BA$(N, 2^m; 4)$ with A-efficiencies

m	N	μ	A-eff
7	43	$(2, 3, 3, 2, 3)$	1
	44	$(3, 3, 3, 2, 3)$	1
	45	$(4, 3, 3, 2, 3)$	0.994
	46	$(4, 3, 3, 2, 4)$	0.974
	47	$(5, 3, 3, 2, 4)$	0.954
	48	$(5, 3, 3, 2, 5)$	0.928
8	43	$(13, 4, 1, 1, 4)$	0.630
	44	$(14, 4, 1, 1, 4)$	0.602
	45	$(7, 4, 1, 2, 8)$	0.613
	46	$(8, 4, 1, 2, 8)$	0.614
	47	$(9, 4, 1, 2, 8)$	0.591
	48	$(10, 4, 1, 2, 8)$	0.578

and by doing so gain an appreciation of Srivastava's intellectual arc in design theory and its place in the history of the subject. Srivastava was pushing the research envelope for fractional designs, establishing a large body of small, useful, irregular fractions at a time when much of the world was preoccupied with orthogonal design. And we can see that this work produced a number of designs that, forty years later, still stand as either mathematically or pragmatically best in a much wider class of competitors.

Taking an even broader view, Srivastava's extensive work firmly fixed the balanced array idea in the statistical literature and the statistical consciousness. Today the search term *balanced array* produces more than 180 matches on MathSciNet, including papers with no direct connection to factorial experimentation. Recent examples include work on equidistant codes (Sinha, Wang, and Wu (2008)), on resolvable designs (Morgan and Reck (2007)), and on bootstrapping (Lin, Lu, and Sitter (2006)). One of Srivastava's enduring legacies is as the popularizer of balanced arrays.

3 Search linear models, search designs, and screening

3.1 Search linear models

As the last of the optimal BA papers was appearing, Srivastava was becoming increasingly concerned with the bias problem inherent in the use of fractions. He would later write (Srivastava (1984)):

> "All through the 1960's the author felt uneasy over the basic assumption behind the work on (fractional) factorial designs, namely, that 3-factor and higher effects are negligible. In many areas where such designs are commonly used, the assumption is not very far from correct, but certainly it is never exactly true either.... *The idea that nature usually comes in such a way that even over relatively large parts of the factor space, the above assumption about all the higher order effects being negligible holds, is unfounded and incorrect.* In view of the above, the 'optimality' of the optimal balanced designs produced in the above work is misleading, since it ignores the *almost certain* presence of *unknown* amounts of bias."

Bias considerations began to dominate Srivastava's design work from the mid-1970s. For fractional factorial experiments, he introduced the fundamental ideas of "search designs and search linear models" in Srivastava (1975a) and the attendant concepts of "optimum search designs" and "bias-free optimality criteria" in Srivastava (1977). This novel, creative work has had lasting and continuing impact, as we will show. In the pioneering 1975 paper, he wrote

> " ....in almost every experiment, there do occur a few effects which were assumed negligible, but which were not actually negligible. Although the number of such non-negligible effects is, in almost all cases, very small, it is quite difficult to pinpoint in advance which effects will turn out to be non-negligible."

This observation led to the idea of the *search linear model* (SLM), expressed as

$$\mathrm{E}(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \quad \mathrm{Var}(\mathbf{Y}) = \sigma^2\mathbf{I}, \tag{2}$$

where $\mathbf{Y}$ is a vector of n response variables, $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$ are known model matrices, $\boldsymbol{\beta}_1$ is a vector of p_1 fixed unknown parameters (effects) of primary interest, and it is believed that at most k elements of $\boldsymbol{\beta}_2$ are nonzero but the identity of these is unknown. If the nonzero elements of $\boldsymbol{\beta}_2$ can be identified, they can then be estimated along with the elements of $\boldsymbol{\beta}_1$, accounting for what would otherwise be a source of bias. In the terminology coined by Srivastava, this entails a *search* for the nonzeros of $\boldsymbol{\beta}_2$. A design that makes this search possible is called a *search design* (SD) of resolving power $\{\boldsymbol{\beta}_1; \boldsymbol{\beta}_2, k\}$; see Srivastava (1975a) and more formally following Theorem 3.1 below.

To more fully explain the search concept and the SLM in (2), consider the $\binom{p_2}{k}$ models

$$\mathrm{E}(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_{2i}\boldsymbol{\beta}_{2i}, \quad \mathrm{Var}(\mathbf{Y}) = \sigma^2\mathbf{I}, \quad i = 1, ..., \binom{p_2}{k}, \tag{3}$$

where the $\boldsymbol{\beta}_{2i}$ are the $\binom{p_2}{k}$ distinct $(k \times 1)$ sub-vectors of $\boldsymbol{\beta}_2$, and the $\mathbf{X}_{2i}$ the corresponding submatrices of $\mathbf{X}_2$. Each of these is a possible model, and if a successful search is to result, the design employed must allow every one of these models (and their submodels) to be estimated. Moreover, the design must not produce aliasing that makes competing models indistinguishable; we must be able to *discriminate* among the models so as to determine which one is correct. In the noiseless case ($\sigma^2 = 0$) a least-squares fit will be perfect but only to the correct model, so such a design should allow the correct model to be fit with certainty, that is, it should allow perfect discrimination among the $\binom{p_2}{k}$ competing models. With noisy data model discrimination cannot be perfect, but is still possible, again only if all of the competing models are estimable and distinguishable from one another. Srivastava established the condition for this in his first search design paper, shown here in a form he later preferred (as in Srivastava (1996)).

Theorem 3.1 *(Srivastava, 1975a)* **The Fundamental Theorem of the Search Linear Model**

(a) *A necessary condition for discrimination of the correct model from among the* $\binom{p_2}{k}$ *models in (3) is*

$$\mathit{rank}(\mathbf{X}_1, \mathbf{X}_{2i}, \mathbf{X}_{2i'}) = p_1 + 2k \tag{4}$$

for every $i, i' = 1, ..., \binom{p_2}{k}$ *such that* $\boldsymbol{\beta}_{2i}$ *and* $\boldsymbol{\beta}_{2i'}$ *are disjoint* $(k \times 1)$ *subvectors of* $\boldsymbol{\beta}_2$.

(b) *If* $\sigma^2 = 0$*, then the rank condition (4) is also sufficient and correct discrimination is assured. If* $\sigma^2 > 0$*, (4) implies that correct model discrimination is still possible, but with probability less than one.*

Naturally the probability of correct discrimination in the noisy case depends, in addition to the magnitude of effects relative to σ^2, on the selection procedure employed; this is discussed further in Section 3.2. In light of Theorem 3.1 we now formally define a search design of resolving power $\{\boldsymbol{\beta}_1; \boldsymbol{\beta}_2, k\}$ as a design satisfying the necessary condition ***(a)***. With such a design all the elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2i}$ are estimable for each of the models (3) regardless of σ^2, and there is no aliasing among distinct members of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_{2i}$, and $\boldsymbol{\beta}_{2i'}$ for any $i \neq i'$. In the search design literature, the estimability is sometimes phrased as "each of the models is identifiable." The non-aliasing further makes discrimination among competing, estimable models possible. Owing to the obvious necessary condition $n \geq p_1 + 2k$, a search design permits estimation of σ^2 regardless of which of the models (3) is correct.

Before the advent of fast computers, establishing the rank condition (4) for *every* pair of submatrices $\mathbf{X}_{2i}$ and $\mathbf{X}_{2i'}$ of $\mathbf{X}_2$ could be a daunting task, leading Srivastava to develop a number of methods for checking (4) in terms of smaller matrices. For example, in Srivastava (1977) the $n \times p_1$ and $n \times p_2$ matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ in (2) are partitioned as

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{X}_{1(1)} \\ \\ \mathbf{X}_{1(2)} \end{bmatrix} \quad \mathbf{X}_2 = \begin{bmatrix} \mathbf{X}_{2(1)} \\ \\ \mathbf{X}_{2(2)} \end{bmatrix}$$

with $\mathbf{X}_{1(1)}$ and $\mathbf{X}_{2(1)}$ having the same number $n_1 (< n)$ of rows and $X_{1(1)}$ having full column rank. Using this partition, define $\mathbf{Q}$ as the matrix

$$\mathbf{Q} = \mathbf{X}_{2(2)} \ - \ \mathbf{X}_{1(2)} \left(\mathbf{X}'_{1(1)} \mathbf{X}_{1(1)} \right)^{-1} \mathbf{X}'_{1(1)} \mathbf{X}_{2(1)}$$

where $\mathbf{X}'$ is the transpose of matrix $\mathbf{X}$. Srivastava showed that a necessary and sufficient condition for (4) to hold is that every set of $2k$ columns of $\mathbf{Q}$ are linearly independent (a quality that he called *property* P_{2k}). Although $\mathbf{Q}$ has the same number of columns as $\mathbf{X}_2$, it has fewer rows and deals with $\mathbf{X}_1$ separately; this makes the linear independence check easier. Later, Srivastava and Gupta (1979) took $n_1 = p_1$ with $\mathbf{X}_{1(1)}$ square and full rank, in which case $\mathbf{Q}$ simplifies to

$$\mathbf{Q} = \mathbf{X}_{2(2)} \ - \ \mathbf{X}_{1(2)} \mathbf{X}_{1(1)}^{-1} \mathbf{X}_{2(1)}.$$

Notice that if $k = 1$, as in much of the work in this area, all that is needed is to check the trivial condition that no two columns of $\mathbf{Q}$ are identical.

An alternative check was given by Srivastava and Ghosh (1977) based on the ranks of the orthogonal projections of all $n \times 2k$ submatrices $\mathbf{X}^*$ of $\mathbf{X}_2$ onto the column space of $\mathbf{X}_1$, so that a necessary and sufficient condition for (4) to hold is that, for every such $\mathbf{X}^*$,

$$\text{rank} \left[\ \mathbf{X}^{*\prime} \left(\mathbf{I} - \mathbf{X}_1 \left(\mathbf{X}'_1 \mathbf{X}_1 \right)^{-1} \mathbf{X}'_1 \right) \mathbf{X}^* \ \right] = 2k \, . \tag{5}$$

Since the matrix in (5) is a $2k \times 2k$ matrix, the rank can again be checked easily for small k. Furthermore, Srivastava and Ghosh (1977) pointed out that, for completely balanced designs (as defined in Section 2), there are many symmetries so that only a few matrices in (5) need to be checked.

3.2 *Identifying the non-zero effects*

In the search for the correct model, provided that at most k of the effects in $\boldsymbol{\beta}_2$ are non-zero, Theorem 3.1 with $\sigma^2 = 0$ guarantees selection of the non-zero effects with probability one. In the more realistic noisy case ($\sigma^2 > 0$), selection of the correct model can still be achieved but with probability less than one. A number of methods for detecting the non-zero (or *active*) effects were suggested by Srivastava; for example, using a comparison of the error sum of squares, or regression sum of squares, of competing models or, alternatively, taking a selection of the k effects that occur most frequently in the h (> 1) models with the smallest error sum of squares (Srivastava (1975a), Section 5). For $k = 1$, a method for "dichotomizing" the interactions into subsets, one of which possibly contains the non-zero effect in $\boldsymbol{\beta}_2$ and the other not, was described by Srivastava and Mallenby (1985).

Interestingly, in some of his later papers, Srivastava allowed assumptions to be made about higher order interactions being negligible. This may seem somewhat at odds with his original worries about unrealistic assumptions, but it did give scope for developing sequential strategies to detect and estimate the k non-zero interactions using as few observations as possible. Srivastava and Hveberg (1992) took this approach discounting 3-factor and higher order interactions, so that $\boldsymbol{\beta}_2$ contains only the two-factor interactions (with $\boldsymbol{\beta}_1$ containing the general mean and the main effects). Here, under the name *probing designs* (also see Srivastava (1987b)), treatment combinations are observed in an order which is specified by the values of previously observed treatment combinations (suggestive of the modern technique of sequential bifurcation; see for example Bettonvil (1995)). Their search requires a "tree structure" for the non-zero effects, similar to that of "strict weak heredity" (see, for example, Wu and Hamada (2009)) and the general approach can be extended to models which include higher order interactions. The multi-stage strategy of Srivastava and Chu (1999) is similar to the technique of resolving confounding in regular designs by adding extra effects in follow-up designs. This work, likewise, assumed negligible three-factor and higher order interactions.

In the current literature, the term "active" is sometimes taken as above to mean "different from zero," but in other cases to mean "large enough to be of practical interest." In the latter case, detection of active effects can be done with probability close to one provided that error variability is small. In fact, in Srivastava and Hveberg (1992), Srivastava stressed that

"... in practice, the words 'zero' and 'non-zero' should be interpreted as 'small' and 'relatively large', where the meaning of these words will have to depend on the experimenter."

With respect to the probability of correctly identifying the non-zero elements of $\boldsymbol{\beta}_2$, Srivastava (1977) reasoned that, in drawing inferences from data obtained from a search design,

"... if a design does not work well for the noiseless case, it can not be expected to work in the noisy case ($\sigma^2 > 0$) either."

Later, Shirakura, Takahashi, and Srivastava (1996) went on to examine the "probability of correct search" and its use in comparing search designs in the noisy case. They gave an explicit formula for the probability of correct selection of $k = 1$ non-zero effect from $\boldsymbol{\beta}_2$ when the errors follow a $N(0, \sigma^2)$ distribution, and calculated this probability for two examples for different values of the "signal to noise" ratio $\rho = \beta_{21}/\sigma$ (where β_{21} is the true value of the non-zero effect from $\boldsymbol{\beta}_2$). The examples looked at a specific MEP.1 search design (defined in Section 3.3) in the cases where (i) $\boldsymbol{\beta}_2$ contains only the two-factor interactions, and (ii) $\boldsymbol{\beta}_2$ contains the two-factor and the three-factor interactions. The probabilities are remarkably high; above 0.95 for (i) with 5–10 factors and $\rho > 1.5$, and just a little lower for (ii) where the search is more competitive. Ghosh and Teschmacher (2002) proposed two criteria based on probabilities of correct search, in addition to that of Shirakura et al. (1996), for determining which of several designs is more likely to identify the $k = 1$ non-zero interaction. That paper also generalized the three criteria for *all* values of ρ so that comparison of search designs can be made without requiring ρ to be known. A BA design was found there to be more likely than two competing OA designs to identify the non-zero interaction. This finding was no surprise to Srivastava as he mentioned many times in his writings and in conference and seminar presentations that he believed in the superiority of BAs over OAs when one is not sure about the true model.

3.3 Construction of search designs

During the 1970s, Srivastava wrote extensively on the problem of constructing search designs of resolving power $\{\boldsymbol{\beta}_1; \boldsymbol{\beta}_2, k\}$ for 2^m factorial experiments, including the issue of optimality. Many of his early papers deal with the situation that we label "Scenario 1", namely that $\boldsymbol{\beta}_1$ contains the general mean, together with the main effects of the m 2-level factors and their two-factor interactions, while $\boldsymbol{\beta}_2$ contains *all* of the higher order interactions, of which only a small number (say, $k = 1$ or $k = 2$) are expected to be non-zero. Under Scenario 1 with

$k = 1$, the number of observations needs to be at least

$$p_1 + 2 \; = \; 1 + m + m(m-1)/2 + 2 \; = \; (m^2 + m + 6)/2 \qquad (6)$$

for (4) to be satisfied. This is considerably fewer observations than would be needed for a full factorial (e.g. 18 versus 32 for $m = 5$, and 24 versus 64 for $m = 6$). Notice that effect hierarchy is not considered here, so that it would be possible, although unlikely, for a 5-factor interaction to be selected as the non-zero interaction while declaring all three-factor and four-factor interactions to be zero. Nevertheless, all hierarchical models can certainly be distinguished.

For Scenario 1, Srivastava and Ghosh (1977) constructed search designs, which they called Balanced Resolution V Plus One (BARE5.1) designs, based on BAs of strength 4. These designs, for numbers of factors in the range $4 \leq m \leq 8$, are a subset of the optimal BAs of resolution V given in the papers of Srivastava and/or Chopra (see Table 3). They enjoy the additional advantage of being BAs not only of strength four but of strength m. Tables in the paper present BARE5.1 designs in easily constructable form along with their efficiencies. A subset of these designs gives rise to an infinite series of search designs (Srivastava and Ghosh, 1976), each of which consists of all treatment combinations with weights $1, m-2, m-1$, and m, where the weight of a treatment combination is equal to the number of factors set at their "non-zero" levels. This design has $(m+1)(m+2)/2$ observations in total, which is only $m-2$ observations more than the theoretical minimum (6).

A second setting, Scenario 2, considered by Srivastava for 2^m designs, occurs when $\boldsymbol{\beta}_1$ contains the general mean and just the main effects of the m factors, while $\boldsymbol{\beta}_2$ contains *all* of the interactions. Srivastava called search designs under this scenario *main effect plus k plans* (Srivastava, 1975a), also known as MEP.k plans or designs of Resolution III.k. One possibility for constructing an MEP.k plan is to take a design (say, $\mathbf{T}_1$) that is optimal for estimating the main effects in $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2 = \mathbf{0}$ and adding to it a design $\mathbf{T}_2$ that will allow the search for the k non-negligible interactions. This strategy is attractive since optimal $\mathbf{T}_1$ are well-known, but as Srivastava pointed out, the constraint of using an optimal $\mathbf{T}_1$ may not lead to a good search design overall. Srivastava and Gupta (1979) took $\mathbf{T}_1$ to consist of the treatment combinations with weights 1 and m and obtained conditions on the weights of the treatment combinations in $\mathbf{T}_2$ to result in an MEP.1 plan. Srivastava and Arora (1991) turned attention to search of $k = 2$ interactions (MEP.2 or Resolution III.2 designs). They gave methods for checking whether a design $(\mathbf{T}_1, \mathbf{T}_2)$ is an MEP.2 design and, among other results, showed that

Table 5
Search scenarios (main effects and factor interactions denoted by m.e. and f.i.)

Scenario	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$
1	mean, m.e. and 2-f.i.	all q-f.i., $q \geq 3$
2	mean, m.e.	all interactions
3	mean, m.e.	2-f.i., 3-f.i.
4	mean, m.e.	2-f.i.
5	mean	all other effects
6	mean	m.e., 2-f.i.
7	mean	m.e.

$\mathbf{T}_1$ as above coupled with $\mathbf{T}_2$ containing treatment combinations with weights 0 and 2 does satisfy these conditions. Scenario 2 was later examined by Gupta (1992) who considered existence conditions for MEP.k designs for general k and presented search designs for 2^7 experiments when $k = 1$ and $k = 2$. From the 1980's onward, many other authors began working on search design problems, expanding beyond the two scenarios described above. MEP.1 and MEP.2 designs were constructed by Ohnishi and Shirakura (1985) for Scenario 3 (see Table 5). Details about these and other plans are available in Ghosh (1996).

Much of the work on construction of search designs has been done in the context of 2-level factors, but a number of authors have looked at more general cases; for example, Ghosh and Zhang (1987), Anderson and Thomas (1980) and Chatterjee and Mukerjee (1993). For a more expansive treatment of the major developments in the construction of search designs, along with discussion and insights, see Ghosh, Shirakura, and Srivastava (2007).

An interesting issue in search design selection is how to assess design efficiency. Srivastava and Ghosh (1977) measured the efficiency of estimation of $\boldsymbol{\beta}_1$ in a search design relative to a design in which these are the only parameters of interest. Srivastava (1977), on the other hand, looked at the sum and product of determinants of the information matrices for estimation of the parameters in the possible models (3) to define the six average optimality criteria: AD-, GD-, AT-, GT-, $AMCR$-, and $GMCR$, where A and G indicate arithmetic and geometric mean, respectively, and D, T and MCR represent determinant, trace, and maximum characteristic root, respectively. The AD criterion was used by Shirakura and Ohnishi (1985) in designs for which $\boldsymbol{\beta}_1$ contains up to ℓ-factor interactions and $\boldsymbol{\beta}_2$ contains $(\ell + 1)$-factor interactions only. Ghosh and Tian (2006) presented tables of optimal search designs with respect to all six criteria for Scenario 4, for the maximum value of k in 2^4 and 2^5 factorial experiments.

3.4 Screening designs

The idea of *screening* was formulated as an application of the search linear model in the pioneering paper Srivastava (1975a). The defining attribute of this setting is that $\boldsymbol{\beta}_1$ contains only the mean, reflecting the experimenter's belief that of the factors being examined, there is no *a priori* sense of which are to be included. Thus $\boldsymbol{\beta}_2$ contains *all* of the potentially relevant effects, for which there are several standard options, here labeled Scenarios 5–7 (see Table 5). Recently, the topic of screening has found renewed interest due to its importance for experiments in industry when budgets allow few observations in comparison with the number of effects to be estimated. For simplicity, as well as greatest relevance to screening applications, we consider only 2-level factors in this section (though Theorem 3.2 is not so restricted).

For the screening context, define an *active factor* to be a factor with a non-zero main effect and/or involved in at least one non-zero interaction. A modification of Theorem 3.1 for screening, that explicitly takes hierarchy into account and can be used in any of Scenarios 5-7, is (see Cheng (2006)):

Theorem 3.2 *Suppose that there are at most k^* active factors among m potentially active factors, where k^* is known, but it is unknown which of the factors are active. Let each of S_1 and S_2 be a set of k^* potentially active factors and let $\mathbf{X}(S_1, S_2)$ denote the matrix whose columns are contrasts that measure the main effects of the factors in $S_1 \cup S_2$, and the interactions between factors within each of S_1 and S_2. A design under model (2) can identify all of the active factors when $\sigma^2 = 0$ if and only if*

$$[\underline{1},\ \mathbf{X}(S_1, S_2)] \text{ is of full column rank for all possible } S_1 \text{ and } S_2\,.$$

Notice that, while Srivastava's rank condition in Theorem 3.1 deals with a search for k non-zero factorial effects in $\boldsymbol{\beta}_2$, here Theorem 3.2 searches for a model containing k^* active *factors*, with "active" as defined above the theorem. Nonetheless there are strong connections between the two results. Take $p_1 = 1$ in Theorem 3.1, and take any k^* that is no more than half the number of factors. It follows immediately that in Scenario 7, the numbers k in Theorem 3.1 and k^* in Theorem 3.2 are equal and the two theorems are equivalent. In fact, Srivastava presented essentially this result as a special case of Theorem 3.1 in the 1975 paper (his Theorem 2 restricted to Scenario 7).

For Scenario 6, if we do not cap how many two-factor interactions are allowed among the k^* active factors, then setting $k = k^*(k^* + 1)/2$ in Theorem 3.1 implies Theorem 3.2 for any $k^* \leq m/2$. The two theorems are now distinguished by the sets of models within which they discriminate. Cheng's theorem considers each possible model with k^* main effects and their $k^*(k^* - 1)/2$ two-factor interactions. Srivastava's does much more, discriminating among models containing *any* k parameters from among the m main effects and $m(m - 1)/2$ two-factor interactions. Similar statements can be made when the number of 2-factor interactions allowed in the model is capped at less than $k^*(k^* - 1)/2$, and analogously for Scenario 5.

Though the Theorem 3.1 condition with suitable k implies the Theorem 3.2 condition, the implication, as expected, is not generally reversible: the two theorems are not equivalent outside of Scenario 7. Again for Scenario 6 consider any design found by selecting six columns from a 12-run Plackett-Burman design (Hadamard matrix). The goal is to screen for two factors and their possible interaction. Taking $k^* = 2$ and so $k = 3$, any one of these designs satisfies the rank condition of Theorem 3.2 and so can distinguish among the 15 competing hierarchical models (and their submodels). However the rank condition of Theorem 3.1 is not satisfied; no such design with just 12 runs can discriminate among all 1330 models containing any 3 parameters from among the main effects and two-factor interactions.

Search designs for screening factor main effects (Scenario 7) were studied by Ghosh and Avila (1985), among others. For main effects screening, the number of observations is only required to be at least $1+2k^*$, and this leads to the consideration of *supersaturated designs* where the number of observations is less than the number of factors. There is now a huge literature on the construction and analysis of supersaturated designs, not just for main effects screening, but also for screening interactions (see Georgiou, Draguljic, and Dean (2009) for references and further details).

The Theorem 3.2 rank requirement for screening designs has echoes in a design property that has seen interest in the last 20 years; namely that of *projectivity* of non-regular fractions (orthogonal arrays, in particular). Box and Tyssedal (1996) defined a design to be of *projectivity* p if all subsets of p columns of the design contain a full factorial design among their runs. Cheng (1998), dedicating his paper to Srivastava on his 65th birthday, gives a very clear accounting of how projectivity and resolving power of two-level designs relate to search design conditions. Cheng (1998) also discusses another related property, called *hidden projectivity* by Wang and Wu (1995). This refers to a design which "allows some (or all) interactions to be estimated even when the design does not have the right resolution or other combinatorial design property for the same interactions to be estimated". Consider $m = 8$ factors specified by a choice of eight columns from a 12-run Plackett-Burman design. One of the resulting, hidden projection properties is that any four main effects and three two-factor interactions among them can be estimated. This is an instructive example of the limitation of designing for projection properties alone, which though certainly necessary, are not sufficient for the screening problem. While this design can estimate any of the models just mentioned, *it cannot discriminate amongst all such models.* This is Scenario 6 with $k^* = 4$ and number of 2-factor interactions capped at three, for which either Theorem 3.1 (put $k = 7$) or Theorem 3.2 requires at least 15 runs to distinguish any two of these models. For a recent review of work on design projection and related properties with many pertinent references, see Xu, Phoa, and Wong (2009).

In this section, we have tried to demonstrate that Srivastava's application of the search linear model for factor screening was well ahead of its time. While most of Srivastava's direct contributions to screening were earlier, the ideas are still being explored, added to, and rediscovered, at a rate that has accelerated from the 1990's on.

4 Bias and Row-Column Designs

We now turn to another aspect of the bias problem - its implications for row-column designs such as Latin squares. Looking back, Srivastava wrote in Srivastava and Wang (1998):

> "The possibility of the presence of nonadditivity [due to row-column interaction], particularly in large Latin Squares, was apprehended by Fisher and Yates (1948), who pointed out that this may invalidate the analysis of the experimental data, which is based on the additive linear model. This remark was picked up by one of the authors (Srivastava), who (in a conversation with R.C. Bose in October 1959) deemphasized large sized Latin Squares from the statistical angle. The same point of view was taken for general row-column designs when the papers of Kiefer (1975a), Kiefer (1975b) were being processed....
>
> In 1981, Srivastava and Kiefer planned to systematically attack the non-additivity problem, but the project did not start because of Kiefer's sudden death. Only in 1990, Srivastava began the investigation, which were published in Srivastava (1993). Some alarming conclusions were reached in this study...."

Table 6
Analysis of published data for $p \times p$ Latin squares. See Srivastava and Wang (1998) for full details, including individual cites and precise deletions.

Source	p	c	s^2_{full}	s^2_{del}	$\frac{s^2_{full}}{s^2_{del}}$	p-value
Bliss	4	3	35.67	0.93	38.22	3.73×10^{-7}
Montgomery	4	3	1.75	0.17	10.29	5.14×10^{-3}
Senn	4	3	0.073	0.001	73.00	1.02×10^{-11}
Montgomery	5	3	10.67	2.83	3.77	0.1065
Montgomery	5	4	3.13	0.13	24.08	9.39×10^{-5}
Cox	6	3	1.62	0.71	2.28	0.4677
Evans and Alldredge	6	3	0.66	0.29	2.28	0.4677
Montgomery	6	3	9.90	5.06	1.96	0.6422
Cochran and Cox	6	4	3.33	0.95	3.51	0.1252

The Srivastava and Wang paper contains ample theory concerning identification of row-column interaction and its impact on treatment comparisons. Some of this employs a search linear model approach. Here we give information from that paper that helps to illustrate the reality of the problem. Table 6 displays the impact on the variance estimate s^2 due to deletion of c cells. The ratio for the two variance values is displayed, as well as the simulated p-value for such an extreme ratio under the null hypothesis of no row-column interaction. Of nine cases, four show clear indication of nonadditivity, and two others are marginal. This is not just an academic problem.

It can be argued that row-column interaction makes no sense from a randomization viewpoint. In the randomization theory, the individual cells all have their own effects which are averaged over rows and columns in the randomization process (producing row and column effects in the randomization model). Srivastava would have none of this: though a believer in randomization for practical reasons, he did not accept it as a valid basis for model justification and analysis. He wrote about this on several occasions, perhaps most succinctly in Srivastava (1996):

> "To avoid biases ... the concept of *randomization* was introduced. This ... should not be confused with *randomization analysis* of designs. It is shown that, approximately, it gives the same answers ... as the ones based on linear models and normality. Some people regard this as an added justification of, or even a more desirable basis of, the usual analysis. However, the latter is controversial. Consider two *different* acts of randomization A_i ($i = 1, 2$) with two different systems of probability P_i .... Since P_1 and P_2 are different, A_1 and

> A_2 will generate two different universes (say, U_1 and U_2). However A_1 and A_2 could both result in the same placement Π of treatments on the units. Now, the experiment depends only on Π.... However, the analysis of the experiment using A_i ($i = 1, 2$) would be done by averaging out U_i, which would generally lead to two *different* conclusions. This is a bit discomforting ...."

Even if you disagree with this argument, there is another reason to dismiss a randomization justification in the current situation. What Srivastava does not seem to have mentioned in print, but which is easily demonstrated, is that if a few units have substantially different effects than others (a plausible manifestation of serious non-additivity), then the normal theory test can be an unacceptably poor approximation to the randomization F-test. Unless one wants to use randomization analysis entirely, this seems adequate reason to take the interaction problem seriously.

Though search design ideas could be implemented to detect the problem, Srivastava recommended *bypassing search* in this instance. In Srivastava (1977) he wrote:

> "However, in large designs of this type, like, for example, large row-column designs, the author is of the opinion that instead of using such designs and searching for interactions, we could almost bypass the latter by simply not using such designs. The reason is that a new class of designs, whose concept we now introduce, seems to be by far more efficient in an overall sense."

This new class Srivastava called "block-treatment designs with two-dimensional blocks." Here is an example given in the 1977 paper for 5 treatments in 5 blocks of size 2×2:

1	2		2	3		3	4		4	0		0	1
3	4		4	0		0	1		1	2		2	3

Lattice squares and other resolvable row-column designs were already available; Srivastava's notion was that the nesting block be incomplete rather than complete or multiply complete. This would allow for shorter rows and columns, dampening the likelihood of departure from the additive model due to fitting constants for larger collections of units.

The idea was picked up very quickly, notably by Singh and Dey (1979), who chose the alternative terminology "block design with nested rows and columns." This terminology would become the norm, sometimes shortened to "nested row-column design." For an extensive introduction to the topic, with many references, see Morgan (1996). MathSciNet now finds well over 50 publications dealing with this class of designs.

So Srivastava's idea definitely caught on, though he did almost no work on nested row-column designs after the 1977 paper. He put forth the concept, but little else; others developed his idea, changing the name in the process. This placed Srivastava on the other side of a situation paralleling the introduction and development of balanced arrays, a small irony that, had he taken notice, he surely would have appreciated. Just as Chakravarti later accepted and used Srivastava's BA terminology, Srivastava later accepted and used "nested row-column" (e.g. Srivastava and Beaver (1986)).

5 Other Design Work

A review of Srivastava's design work would not be complete without mention of at least two other areas in which he made significant contributions. Not surprisingly given its close association with balanced arrays, one of these is orthogonal arrays.

An orthogonal array of strength t, OA$(N, s^m; t)$, is just a BA$(N, s^m; t)$ for which the permutation-invariant frequency for row vectors in any $N \times t$ subarray is constant across all s^t vectors (for $s = 2$ this is equality of the μ_i defined in Section 2). One standard construction for an OA uses the finite field GF$_s$ as follows. For any vector $\boldsymbol{v}$ whose elements are drawn from GF$_s$, define the weight of $\boldsymbol{v}$ to be the number of nonzero elements in $\boldsymbol{v}$. Let $\boldsymbol{A}_{r\times m}$ and $\boldsymbol{c}_{r\times 1}$ be a known matrix and vector respectively with elements in GF$_s$, and let $\boldsymbol{A}$ have rank r. Then the s^{m-r} solutions $\boldsymbol{x}$ to the equations $\boldsymbol{Ax} = \boldsymbol{c}$ are the rows of OA$(s^{m-r}, s^m; t)$ if, and only if, every vector in the row space of $\boldsymbol{A}$ has weight at least $t+1$ (Rao (1947),Rao (1950)). Evidently the solutions $\boldsymbol{x}$ form an affine subspace, also called an $(m-r)$-*flat*, in the m-dimensional Euclidean geometry EG$(m$,$s)$.

Srivastava and Throop (1990) sought to understand the conditions that would produce an OA when solving several such sets of equations. For $i = 1, \ldots, f$ let $\boldsymbol{T}_i$ be the $s^{m-r} \times m$ array whose rows are all solutions $\boldsymbol{x}$ of $\boldsymbol{Ax} = \boldsymbol{c_i}$ where $\boldsymbol{A}$ is as above and the $\boldsymbol{c_i}$ are known, distinct vectors in EG$(r$,$s)$. Let $\boldsymbol{T}_{(fs^{m-r}\times m)} = [\boldsymbol{T_1'}|\boldsymbol{T_2'}|\ldots|\boldsymbol{T_f'}]'$ and let $\boldsymbol{C} = [\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_f}]$. They (also see Srivastava and Chopra (1973), where the result was first stated) proved:

Theorem 5.1 $\boldsymbol{T}$ *is an orthogonal array of strength t if and only if, for any vector* $\boldsymbol{u}_{(r\times 1)}$ *in EG*(r, s) *with* $\boldsymbol{u}'\mathbf{A} \neq \mathbf{0}$ *and weight* $(\boldsymbol{u}'\boldsymbol{A}) \leq t$*, the* $(f \times 1)$ *vector* $\boldsymbol{C}'\boldsymbol{u}$ *is an orthogonal array of strength 1.*

The $\boldsymbol{T}_i$ share no common rows and so are parallel flats in EG(m,s). Consequently $\boldsymbol{T}$ is termed a *parallel flats fraction*. Parallel flats fractions were introduced in Connor (1960) and Connor and Young (1961) for $s = 2, 3$. Srivastava, Anderson, and Mardekian (1984) built a general theory for these designs. Srivastava (1987a), Srivastava and Li (1996), and Li and Srivastava (1997) further advanced the theory and provided optimal fractions. Many other references for work in this area may be found in these papers.

Turning back to his earlier years, while a doctoral student at UNC, Srivastava began a collaboration with S. N. Roy on design for experiments with multivariate response. This work became more vigorous during his post-doctoral appointment (see Section 1) and continued until the sudden demise of Roy in 1964. Significant products of their collaboration included Roy and Srivastava (1965) and the monograph Roy, Gnanadesikan, and Srivastava (1971) (a project begun in the summer of 1962). These works established theory for *hierarchical designs* for situations where fewer measurements were to be made on response variables of lesser interest. They also introduced *p-block designs* for settings where variability in different responses is impacted by different blocking variables (the p refers to the number of response variables). Srivastava was also independently authoring design papers for multivariate response (Srivastava (1966), Srivastava (1968)) in addition to other work in multivariate inference. Those papers were shortly followed by a series of articles with his doctoral student L. L. McDonald (Srivastava and McDonald (1969), Srivastava and McDonald (1970), Srivastava and McDonald (1971)) incorporating cost considerations into multi-response design. Srivastava was a leader in this area through the 1960's and into the early 1970's, concurrent with other authors (notably Farrell, Kiefer, and Walbran (1967) and Fedorov (1972)), opening research paths for others to follow and generalize.

6 Editor and Conference Organizer

Srivastava was tireless in the promotion of new statistical and combinatorial thinking in experimental design. He organized four international conferences at Fort Collins, in 1971, 1973, 1978, and 1995. He edited volumes containing papers from the first three of these: Srivastava, Harary, Rao, Rota, and Shrikhande (1973), Srivastava (1975b), and Srivastava (1980). He published papers from the 1995 Bose Memorial Conference as special issues of *Journal of Statistical Planning and Inference* (JSPI), volumes 72 and 73.

The 1973 conference "Statistical Design and Analysis of Experiments and Linear Models" was particularly significant in terms of both participation and impact. Srivastava (1975b) contains the invited papers from this conference, including Srivastava (1975a) (discussed at length earlier in this article) and the influential papers Hedayat and Afsarinejad (1975), Huber (1975), and Kiefer (1975b). With emphasis in the research world about to shift from purely combinatorial to optimal design, the conference brought together a wide swath of statisticians at the cusp of a new era for planning experiments. It was also during this conference that Srivastava began building widespread support for the creation of a new journal that would be particularly encouraging of work in design and sampling. He and others sensed that the *Annals of Mathematical Statistics* in its last years had become less receptive to papers in design theory, a trend they saw continuing with the newly created *Annals of Statistics*. Sam Hedayat, then associate editor for *Annals of Statistics*, was experiencing a degree of this perceived design antipathy first hand, and was sympathetic to Srivastava's view. He was a participant in the 1973 conference as well as in much of the correspondence on the "new journal issue" even while doing his best to push quality design papers in the *Annals*. According to Hedayat,

> "... there was no society which was willing to sponsor a new journal targeted to design and sampling. I should say this much, if Srivastava was not behind the idea then there would have been no new journal. He was passionate about it and frankly upset about the direction of *Annals of Statistics*."

Enlisting support from an impressive list of distinguished statisticians (interested readers may wish to peruse the initial editorial board), Srivastava partnered with North Holland to create JSPI with a driving spirit of cooperation, encouraging the interplay of design with all areas of statistical activity. The first sentence of JSPI **1**, issue 1 (1977), highlights this perspective: "*Journal of Statistical Planning and Inference* has been established to provide a common medium for the dissemination of significant information in all branches of statistical planning and related inference problems." By JSPI **7** (1982), this had evolved to "This is a broad based journal covering *all* branches of statistics, with special encouragement to workers in the field of statistical planning and related combinatorial mathematics." Under Srivastava's direction a few prominent mathematicians were always present on the editorial board of JSPI to encourage cross-fertilization, and new researchers were brought in to keep the process dynamic and vigorous.

JSPI was officially "born" with the first call for papers in late 1975, with Srivastava as Editor-in-Chief, a position he held with clarity of vision and a commitment to knowledge for 35 years. Preparing to step down, he wrote in Srivastava (2010):

> "The Journal truly belongs to the world scientific community. I hope they continue to nourish it and be nourished by it, without regard to geographical boundaries, race, religion, ethnicity, and other such factors."

But Srivastava would pass before his term ended. He wrote in 2010:

> "It is said that all things that have a beginning also have an end. The same is true of me as the Editor-in-Chief of JSPI. I will soon be moving onwards from this position. There is a long list of people whom I must thank for the success of the journal. In spite of my many inefficiencies, they have helped bring it into the category of the largest prestigious Statistics journals in the world."

7 Srivastava and design: an exceptional career

Not all of Srivastava's ideas have stood the test of time, but no one's do; this is part and parcel of scientific progress. The MDPB association scheme underlying much of his earliest design work, picked up by a few other authors, now seems inactive in the statistical literature (though it has enjoyed an independent life among group theorists). Yet like many non-trivial ideas, it put Srivastava on a path that has had many ramifications. His extensive explorations on balanced arrays popularized the idea and has led to work that continues today, even in areas removed from their originally intended application. He was an early proponent of run-size economy in fractions, an early pursuant of optimal fractions, and a consistent voice warning of the dangers of bias. Much of his work in fractions, especially that arising out of the search design concept, is foundational to currently active research. He introduced the notion of nested row-column designs. He was founding editor of JSPI. Ignoring all of his work in other areas, his design contributions alone comprise an exceptional career.

Srivastava made significant contributions to areas of statistics apart from design, including multivariate analysis and sampling; the interested reader is directed to Ghosh (1999) as a good starting point for these topics. A deeply spiritual man, some of his writings in this regard are available at *higherreality.googlepages.com*. A tribute to Srivastava and a list of his publications are available in Ghosh (1999) and his obituary in Ghosh (2010).

We accept the reality: "... that all things that have a beginning also have an end." Srivastava wrote in *Sparks of Consciousness* (Srivastava (2009)):

No one knows where we all came from,
We do not know if we shall ever meet again.

Yes, we do not know whether we shall see Jagdish again but we will long feel his absence at professional meetings and conferences with his insightful, thought-provoking, and penetrating questions. He is present and will remain present in our thoughts.

Acknowledgments

This paper arose from talks we individually delivered in the Srivastava Memorial Session that led off the Memphis conference. Our thanks to R. A. Bailey who, during that session, spotted the previously unknown connection to Higman's work on coherent configurations. We also thank S. Hedayat for providing information on Srivastava's role in the founding of JSPI, D. Majumdar for directing us to Professor Hedayat and also for speaking in the Memorial Session, and D. V. Chopra for correspondence regarding balanced arrays. This paper was completed at the Isaac Newton Institute for the Mathematical Sciences during the 2011 Programme on the Design and Analysis of Experiments. This work was sponsored, in part, by the National Science Foundation under Agreement DMS-0806134 (The Ohio State University).

References

Anderson, D. A. and Thomas, A. M. (1980), "Weakly Resolvable IV.3 Search Designs for the p^n Factorial Experiment," *Journal of Statistical Planning and Inference*, 4, 299–312.

Bettonvil, B. (1995), "Factor Screening by Sequential Bifurcation," *Communications in Statistics: Simulation and Computation*, 24, 165–185.

Bose, R. C. and Kishen, K. (1940), "On the problem of confounding in the general symmetrical factorial design," *Sankhyā*, 5, 21–36.

Bose, R. C. and Nair, K. R. (1939), "Partially balanced incomplete block designs," *Sankhya*, 4, 337–372.

Bose, R. C. and Shimamoto, T. (1952), "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Statist. Assoc.*, 47, 151–184.

Bose, R. C. and Srivastava, J. N. (1963), "Mathematical theory of factorial designs," *Bull. Inst. Internat. Statist*, 40, 780–794.

— (1964a), "Analysis of irregular factorial fractions," *Sankhyā Ser. A*, 26, 117–144.

— (1964b), "Multidimensional partially balanced designs and their analysis, with applications to partially balanced factorial fractions," *Sankhyā Ser. A*, 26, 145–168.

— (1964c), "On a Bound Useful in the Theory of Factorial Designs and Error Correcting Codes," *The Annals of Mathematical Statistics*, 35, 408–414.

Box, G. and Tyssedal, J. (1996), "Projective Properties of Certain Orthogonal Arrays," *Biometrika*, 83, 950–955.

Brouwer, A. E. (2011), "Coherent configurations," Unpublished notes available at *www.abc*.

Chakravarti, I. M. (1956), "Fractional replication in asymmetrical factorial designs and partially balanced arrays," *Sankhyā*, 17, 143–164.

— (1974), "Review of 'Optimal balanced 2^8 fractional factorial designs of resolution V, $37 \leq N \leq 51$' by D. V. Chopra and J. N. Srivastava," *Math. Rev.*, 52.

Chatterjee, K. and Mukerjee, R. (1993), "Search Designs for Estimating Main Effects and Searching Several Two-factor Interactions in General Factorials," *Journal of Statistical Planning and Inference*, 35, 131–138.

Cheng, C. S. (1998), "Projectivity and Resolving Power," *Journal of Combinatorics, Information and System Sciences*, 23, 47–58.

— (2006), "Projection Properties of Factorial Designs for Factor Screening," in Screening: Methods for Experimentation in Industry, Drug Discovery and Genetics., eds. Dean, A. M. and Lewis, S. M., New York: Springer Verlag., chap. 7, pp. 156–168.

Chopra, D. V. (1975a), "Balanced optimal 2^8 fractional factorial designs of resolution V, $52 \leq N \leq 59$," in *A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973)*, Amsterdam: North-Holland, pp. 91–100.

— (1975b), "Optimal balanced 2^8 fractional factorial designs of resolution V, with 60 to 65 runs," in *Proceedings of the 40th Session of the International Statistical Institute (Warsaw, 1975), Vol. 3. Contributed papers*, vol. 46, pp. 161–166 (1976).

— (1975c), "Some investigations on trace optimal balanced designs of resolution V and balanced arrays," in *Wichita State University Bulletin LI-4, Univ. Stud.*, vol. 105.

Chopra, D. V., Kipngeno, W. A. K., and Ghosh, S. (1986), "More precise tables of optimal balanced 2^m fractional factorial designs of Srivastava and Chopra, $7 \leq m \leq 10$," *J. Statist. Plann. Inference*, 15, 115–121.

Chopra, D. V. and Srivastava, J. N. (1973a), "Optimal balanced 2^7 fractional factorial designs of resolution V, $49 \leq N \leq 55$," *Comm. Statist.*, 2, 59–84.

— (1973b), "Optimal balanced 2^7 fractional factorial designs of resolution V, with $N \leq 42$," *Ann. Inst. Statist. Math.*, 25, 587–604.

— (1974), "Optimal Balanced 2^8 Fractional Factorial Designs of Resolution V, $37 \leq N \leq 51$," *Sankhyā, Series A*, 36, 41–52.

— (1975), "Optimal Balanced 2^7 Fractional Factorial Designs of Resolution V, $43 \leq N \leq 48$," *Sankhyā, Series B*, 37, 429–447.

Connor, W. S. (1960), "Construction of fractional factorial designs of the mixed $2^m \times 3^n$ series," in *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling*, Stanford: Stanford University Press, pp. 168–181.

Connor, W. S. and Young, S. (1961), "Fractional factorial designs for experiments with factors at two and three levels," *Nat. Bur. Standards Appl. Math. Ser. No.*, v+65.

Farrell, R. H., Kiefer, J., and Walbran, A. (1967), "Optimum multivariate designs," in *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, Berkeley, Calif.: Univ. California Press, pp. Vol. I: Statistics, pp. 113–138.

Fedorov, V. V. (1972), *Theory of optimal experiments*, New York: Academic Press, translated from the Russian and edited by W. J. Studden and E. M. Klimko, Probability and Mathematical Statistics, No. 12.

Fisher, R. A. and Yates, F. (1948), *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 3d ed.

Georgiou, S. D., Draguljic, D., and Dean, A. M. (2009), "An overview of two-level supersaturated designs with cyclic structure," *Journal of Statistical Theory and Practice*, 3, 489–504, submitted.

Ghosh, S. (1996), "Sequential Assembly of Fractions in Factorial Experiments," in *Design and Analysis of Experiments [Handbook of Statistics 13]*, eds. Ghosh, S. a. and Rao, C. R. a., Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 407–435.

Ghosh, S. (ed.) (1999), *Multivariate analysis, design of experiments, and survey sampling*, vol. 159 of *Statistics: Textbooks and Monographs*, New York: Marcel Dekker Inc., a tribute to Jagdish N. Srivastava.

Ghosh, S. (2010), "Obituary: Jagdish N. Srivastava, June 20, 1933 − November 18, 2010," *Journal of Statistical Planning and Inference*, 141, 1633–1634.

Ghosh, S. and Avila, D. (1985), "Some new factor screening designs using the search linear model," *Journal of Statistical Planning and Inference*, 11, 259–266.

Ghosh, S., Shirakura, T., and Srivastava, J. N. (2007), "Model identification using search linear models and search designs," in *Entropy, search, complexity*, Springer, vol. 16 of

Bolyai Society Mathematical Studies, pp. 85–112.

Ghosh, S. and Teschmacher, L. (2002), "Comparisons of Search Designs Using Search Probabilities," *Journal of Statistical Planning and Inference*, 104, 439–458.

Ghosh, S. and Tian, Y. (2006), *Journal of Multivariate Analysis*, 97, 1437 – 1450.

Ghosh, S. and Zhang, X. (1987), "Two new series of search designs for 3^m factorial experiments," *Utilitas Mathematica*, 32, 245–254.

Gupta, B. (1992), "On the existence of main effect plus k plans for 2^m factorials and tables for main effect plus 1 and 2 plans for 2^7 factorials," *Communications in Statistics, Theory and Methods*, 21, 1137–1143.

Harter, H. L. (1963), "Review of 'Mathematical theory of factorial design' by R. C. Bose and J. N. Srivastava," *Math. Rev.*, 31.

Hedayat, A. and Afsarinejad, K. (1975), "Repeated measurements designs. I," in *A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973)*, Amsterdam: North-Holland, pp. 229–242.

Higman, D. G. (1970), "Coherent configurations. I," *Rend. Sem. Mat. Univ. Padova*, 44, 1–25.

Huber, P. J. (1975), "Robustness and designs," in *A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins Colo., 1973)*, Amsterdam: North-Holland, pp. 287–301.

Kiefer, J. (1975a), "Balanced block designs and generalized Youden designs. I. Construction (patchwork)," *Ann. Statist.*, 3, 109–118.

— (1975b), "Construction and optimality of generalized Youden designs," in *A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973)*, Amsterdam: North-Holland, pp. 333–353.

Kishen, K. and Srivastava, J. N. (1959a), "Confounding in asymmetrical factorial designs in relation to finite geometries," *Current Science*, 28, 98–100.

— (1959b), "Mathematical theory of confounding in asymmetrical and symmetrical factorial designs," *J. Indian Soc. Agric. Statist*, 11, 73–110.

Li, J. and Srivastava, J. N. (1997), "Optimal 2^n factorial designs of parallel flats type," *Comm. Statist. Theory Methods*, 26, 2473–2488.

Lin, C. D., Lu, W. W., and Sitter, R. R. (2006), "Fast approximately balanced bootstrap without construction," *Statist. Probab. Lett.*, 76, 1861–1872.

Morgan, J. P. (1996), "Nested designs," in *Design and analysis of experiments*, Amsterdam: North-Holland, vol. 13 of *Handbook of Statist.*, pp. 939–976.

Morgan, J. P. and Reck, B. H. (2007), "Resolvable designs with large blocks," *Ann. Statist.*, 35, 747–771.

Nguyen, N.-K. and Miller, A. J. (1997), "2^m fractional factorial designs of resolution V with high A-efficiency, $7 \leq m \leq 10$," *J. Statist. Plann. Inference*, 59, 379–384.

Ohnishi, T. and Shirakura, T. (1985), "Search Designs for 2^m Factorial Experiments," *Journal of Statistical Planning and Inference*, 11, 241–245.

Rao, C. R. (1947), "Factorial experiments derivable from combinatorial arrangements of arrays," *Supp. J. Royal Stat. Soc.*, 9, 128–139.

— (1950), "The theory of fractional replication in factorial experiments," *Sankhyā*, 10, 81–86.

Roy, S. N., Gnanadesikan, R., and Srivastava, J. N. (1971), *Analysis and design of certain quantitative multiresponse experiments*, New York: Pergamon Press.

Roy, S. N. and Srivastava, J. N. (1965), "Hierarchical and p-block multiresponse designs and their analysis," in *Contributions to Statistics*, Calcutta: Statist. Publ. Soc., pp. 419–428.

Shirakura, T. and Ohnishi, T. (1985), "Search Designs for 2^m Factorials Derived from Balanced Arrays of Strength $2(l+1)$ and Ad-optimal Search Designs," *Journal of Statistical Planning and Inference*, 11, 247–258.

Shirakura, T., Takahashi, T., and Srivastava, J. N. (1996), "Searching Probabilities for Nonzero Effects in Search Designs for the Noisy Case," *The Annals of Statistics*, 24, 2560–2568.

Singh, M. and Dey, A. (1979), "Block designs with nested rows and columns," *Biometrika*, 66, 321–326.

Sinha, K., Wang, Z., and Wu, D. (2008), "Good equidistant codes constructed from certain combinatorial designs," *Discrete Math.*, 308, 4205–4211.

Srivastava, J. (ed.) (1980), *Combinatorial mathematics, optimal designs and their applications*, Ann. Discrete Math.

Srivastava, J. (1984), "Sensitivity and revealing power: two fundamental statistical criteria other than optimality arising in discrete experimentation," in *Experimental design, statistical models, and genetic statistics*, New York: Dekker, vol. 50 of *Statist. Textbooks Monogr.*, pp. 95–117.

— (1987a), "Advances in the general theory of factorial designs based on partial pencils in Euclidean n-space," *Utilitas Math.*, 32, 75–94.

— (2009), *Sparks of Consciousness: Spiritual Poems, Beyond Religion*, Atma Ram & Sons.

Srivastava, J., Anderson, D., and Mardekian, J. (1984), "Theory of factorial designs of the parallel flats type. I. The coefficient matrix," *J. Statist. Plann. Inference*, 9, 229–252.

Srivastava, J. and Hveberg, R. (1992), "Sequential Factorial Probing Designs for Identifying and Estimating Non-negligible Factorial Effects for the 2^m Experiment under the Tree

Structure," *Journal of Statistical Planning and Inference*, 30, 141–162.

Srivastava, J. N. (1966), "Incomplete multiresponse designs," *Sankhyā Ser. A*, 28, 377–388.

— (1968), "On a general class of designs for multiresponse experiments," *Ann. Math. Statist.*, 39, 1825–1843.

— (1970), "Optimal balanced 2^m fractional factorial designs," in *Essays in Probability and Statistics*, Univ. of North Carolina Press, Chapel Hill, N.C., pp. 689–706.

— (1972), "Some general existence conditions for balanced arrays of strength t and 2 symbols," *J. Combinatorial Theory Ser. A*, 13, 198–206.

— (1975a), "Designs for Searching Non-negligible Effects," in *Statistical Design and Linear Models*, Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 507–520.

Srivastava, J. N. (ed.) (1975b), *A survey of statistical design and linear models*, North-Holand Publishing Co., Amsterdam-Oxford.

Srivastava, J. N. (1977), "Optimal Search Designs, or Designs Optimal under Bias Free Optimality Criteria," in *Statistical Decision Theory and Related Topics, Proceedings of a Symposium at Purdue University, May, 1976, in two volumes, Volume 2*, Academic Press, pp. 375–410.

— (1978), "A Review of Some Recent Work on Discrete Optimal Factorial Designs for Statisticians and Experimenters," in *Developments in Statistics – Volume 1*, ed. Krishnaiah, P. R., Academic Press, pp. 267–329.

— (1987b), "On the inadequacy of the customary orthogonal arrays in quality control and general scentific experimentation and the need of probing designs of higher revealing power," *Communications in Statistics, Theory and Methods*, 16, 2901–2941.

— (1993), "Nonadditivity in row-column designs," *J. Combin. Inform. System Sci.*, 18, 85–96.

— (1996), "A Critique of Some Aspects of Experimental Design," in *Design and Analysis of Experiments [Handbook of Statistics 13]*, eds. Ghosh, S. a. and Rao, C. R. a., Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 309–341.

— (2001), "Logic, Goedel's theorem, rationality, and spirituality," in *Thoughts on Synthesis of Science and Religion*, Kolkata: Bhaktivedanta Institute, pp. 157–174.

— (2010), "Bon voyage," *Journal of Statistical Planning and Inference*, 140, 1619.

Srivastava, J. N. and Anderson, D. A. (1971), "Factorial Association Schemes with Applications to the Construction of Multidimensional Partially Balanced Designs," *The Annals of Mathematical Statistics*, 42, 1167–1181.

Srivastava, J. N. and Arora, S. (1991), "An infinite sesries of resoultion III.2 designs for the 2^m foactorial experiment," *Discrete Mathematics*, 98, 35–56.

Srivastava, J. N. and Beaver, R. J. (1986), "On the Superiority of the Nested Multidimensional Block Designs, Relative to the Classical Incomplete Block Designs," *Journal of Statistical Planning and Inference*, 13, 133–150.

Srivastava, J. N. and Bose, R. C. (1966), "Some economic partially balanced 2^m factorial fractions," *Ann. Inst. Statist. Math.*, 18, 57–73.

Srivastava, J. N. and Chopra, D. V. (1971a), "Balanced Optimal 2^m Fractional Factorial Designs of Resolution V, $m \leq 6$," *Technometrics*, 13, 257–269.

— (1971b), "On the Characteristic Roots of the Information Matrix of 2^m Balanced Factorial Designs of Resolution V, with Applications," *The Annals of Mathematical Statistics*, 42, 722–734.

— (1973), "Balanced arrays and orthogonal arrays," in *A survey of combinatorial theory (Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1971)*, Amsterdam: North-Holland, pp. 411–428.

— (1974), "Balanced trace-optimal 2^7 fractional factorial designs of resolution V, with 56 to 68 runs," *Utilitas Math.*, 5, 263–279.

Srivastava, J. N. and Chu, J. Y. (1999), "Multistage Design Procedures for Identifying Two-factor Interactions, When Higher Effects Are Negligible," *Journal of Statistical Planning and Inference*, 78, 149–189.

Srivastava, J. N. and Ghosh, S. (1976), "A Series of Balanced 2^m Factorial Designs of Resolution V Which Allow Search and Estimation of One Extra Unknown Effect," *Sankhyā, Series B*, 38, 280–289.

— (1977), "Balanced 2^m factorial designs of resolution V which allow search and estimation of one extra unknown effect, $4 \leq m \leq 8$," *Communications in Statistics, Theory and Methods*, 6, 141–166.

Srivastava, J. N. and Gupta, B. C. (1979), "Main Effect Plan of 2^m Factorials Which Allow Search and Estimation on One Unknown Effect," *Journal of Statistical Planning and Inference*, 3, 259–266.

Srivastava, J. N., Harary, F., Rao, C. R., Rota, G.-C., and Shrikhande, S. S. (eds.) (1973), *A survey of combinatorial theory*, North-Holland Publishing Col., Amsterdam-London.

Srivastava, J. N. and Li, J. (1996), "Orthogonal designs of parallel flats type," *J. Statist. Plann. Inference*, 53, 261–283.

Srivastava, J. N. and Mallenby, D. W. (1985), "On a Decision Rule Using Dichotomies for Identifying the Nonnegligible Parameter in Certain Linear Models," *Journal of Multivari-*

ate Analysis, 16, 318–334.

Srivastava, J. N. and McDonald, L. (1971), "On the costwise optimality of certain hierarchical and standard multiresponse models under the determinant criterion," *J. Multivariate Anal.*, 1, 118–128.

Srivastava, J. N. and McDonald, L. L. (1969), "On the costwise optimality of hierarchaical multiresponse randomized block designs under the trace criterion," *Ann. Inst. Statist. Math.*, 21, 507–514.

— (1970), "On the hierarchical two-response (cyclic PBIB) designs, costwise optimal under the trace criterion," *Ann. Inst. Statist. Math.*, 22, 507–518.

Srivastava, J. N. and Throop, D. (1990), "Orthogonal arrays obtainable as solutions to linear equations over finite fields," *Linear Algebra Appl.*, 127, 283–300.

Srivastava, J. N. and Wang, Y. C. (1998), "Row-column Designs: Non-additivity Makes Them Hazardous to Use," *Journal of Statistical Planning and Inference*, 73, 277–315.

Suen, C.-Y. and Chakravarti, I. M. (1985), "Balanced factorial designs with two-way elimination of heterogeneity," *Biometrika*, 72, 391–402.

— (1986), "Efficient two-factor balanced designs," *J. Roy. Statist. Soc. Ser. B*, 48, 107–114.

Wang, J. C. and Wu, C. F. J. (1995), "A Hidden Projection Property of Plackett-Burman and Related Designs," *Statistica Sinica*, 5, 235–250.

Wu, C. F. J. and Hamada, M. (2009), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley, 2nd ed.

Xu, H., Phoa, F. K. H., and Wong, W. K. (2009), "Recent developments in nonregular fractional factorial designs," *Statistics Surveys*, 3, 18–46.