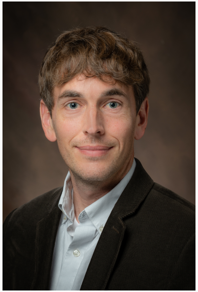
STOP TREATING SUPERSATURATED DESIGNS LIKE OTHER SCREENING DESIGNS

Maria Weese
Information Systems & Analytics
Miami University
JSM 2020

weeseml@miamioh.edu
Twitter: @MFWeese

CO-AUTHORS

David Edwards
VCU
Richmond, Virginia



Byran Smucker
Miami University
Oxford, Ohio

Jon Stallrich
NC State
Raleigh, North Carolina

“I think it is perfectly natural and wise to do some supersaturated experiments.”

–John Tukey

(from a discussion of Satterthwaite 1959)

WHAT IS SCREENING?

- Screening is the first stage of a sequential experimental procedure.
- Screening involves the choice of a design and analysis combination.
- Screening aims to classify factors into those that should be further studied (i.e. "potentially active") and those that can be ignored (i.e. "inactive").

Two-level supersaturated designs (SSDs) use $n < k+1$ runs to examine k factors. This design uses $n = 6$ runs to examine $k = 9$ factors.

$$D = \begin{pmatrix} -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

BASIC SCREENING MODEL

The most basic screening model includes just the linear main effects:

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where n is the number of runs, $\epsilon_i \sim N(0, \sigma^2)$, x_{ij} is the j-th factor's setting for run i, and β_j is an unknown parameter.

The model is equivalently written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is the $n \times (k+1)$ model matrix, $\boldsymbol{\beta}$ is a $(k+1)$-vector of model parameters, and $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ are n-vectors, with $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

The j-th factor is considered active if $|\beta_j| > t$ for some threshold $t \geq 0$.

The screening classification rule depends on the experimenter's willingness to risk classifying an inactive factor as potentially active (type 1 error) and classifying a truly active factor as inactive (type 2 error)

	$\lvert\beta_j\rvert > t$	$\lvert\beta_j\rvert \leq t$
Label: Potentially Active	Correct	Type 1 Error
Label: Inactive	Type 2 Error	Correct

THE TRADE OFF

A trade-off must be made that depends on the budget for future experimentation and the overall goals:

1. Is it important that few, if any, active factors are omitted, even at the expense of more type 1 errors?
2. Or is the goal to to identify as many active factors as possible, while controlling the type 1 error more stringently?

THE TRADE OFF

A trade-off must be made that depends on the budget for future experimentation and the overall goals:

1. Is it important that few, if any, active factors are omitted, even at the expense of more type 1 errors?
2. Or is the goal to to identify as many active factors as possible, while controlling the type 1 error more stringently?

The best choice of supersaturated design (SSD) construction and analysis depends on this practitioner-specified trade-off.

The fundamental principle of experimental design is that the data collection procedure strongly influences an estimator's statistical properties, and hence factor classification.

In particular, $\hat{\beta}_{LS} = (X^T X)^{-1} X^T Y$ has good screening properties for a given design if $\sigma^2 (X^T X)^{-1}$ is "small."

The ideal matrix, $X^T X = nI_n$, comes from regular and nonregular fractional factorial designs, having an X with settings ± 1 and mutually orthogonal columns.

TRADITIONAL SSD CONSTRUCTION

SSDs have been constructed via heuristic measures of orthogonality based on the off-diagonals of $\mathbf{X}^T\mathbf{X} = (s_{ij})$.

For example, the E(s^2)-criterion forces $s_{0j} = 0$ and minimizes $E(s^2) = \binom{k}{2}^{-1} \sum_{1 \le i < j \le k} s_{ij}^2$.

The unconditional E(s^2)-criterion, or UE(s^2)-criterion (Jones and Majumdar 2014; Weese et al. 2015) is similarly defined, but does not require $s_{0j} = 0$; that is, $UE(s^2) = \frac{2}{k(k+1)} \sum_{0 \le i < j \le k} s_{ij}^2$.

Such criteria intend to approximate the ideal structure $\mathbf{X}^T\mathbf{X} = n\mathbf{I}_n$ as closely as possible.

For SSDs, such an X cannot exist and, even worse, the main-effect model is not least-squares estimable.

For SSDs, such an X cannot exist and, even worse, the main-effect model is not least-squares estimable.

There has been no clear consensus about the optimal pairing of SSD construction criteria and analysis strategy.

This may partially explain why practitioners are hesitant to adopt SSDs for screening.

SSDs as a Design and Analysis Combination

GO-SSDS VS. var(s+)

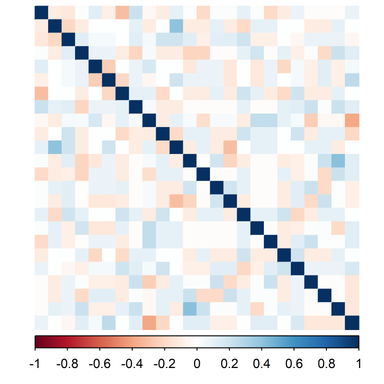
GO-SSDs (Jones et al. (2019))

- GO-SSDs include “fake” factor columns to estimate σ^2.
- The true factor columns are partitioned into mutually-orthogonal groups.
- GO-SSDs are generated using a Kronecker product of a Hadamard matrix, $\mathbf{H}_m$ and a small generating SSD

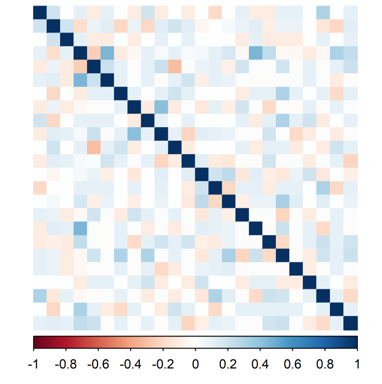
Var(s+), (Weese et al. (2017))

- The Var(s+) criterion minimizes the variance of the off-diagonal s_{ij}'s subject to some constraints.
- This criterion allows the s_{ij}'s to be, on average, more positive than those in the approximately UE(s^2)-optimal design, but with less variability.

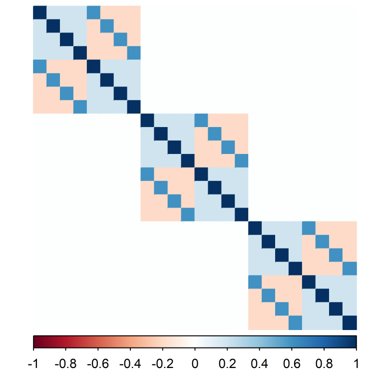
SSDS



$n = 20$, $k = 24$ UE(s^2) SSD



$n = 20$, $k = 24$ Var($s+$) SSD



$n = 20$, $k = 24$ GO-SSD

GO-SSDs: MaxPower

- Two stage analysis using the estimate of σ^2 from the fake factor columns.
- First stage is group testing.
- Second stage is factor testing.

Var(s+): Dantzig Selector

- Pre-specify effect directions.
- Visually analyze the experiment with a profile plot of the Dantzig selector solutions.
- Even if effect directions are misspecified, power is equivalent to traditional SSDs

SSD Comparison

For several SSD (n, k) sizes, each SSD we generated 5000 responses according to model (1) by:

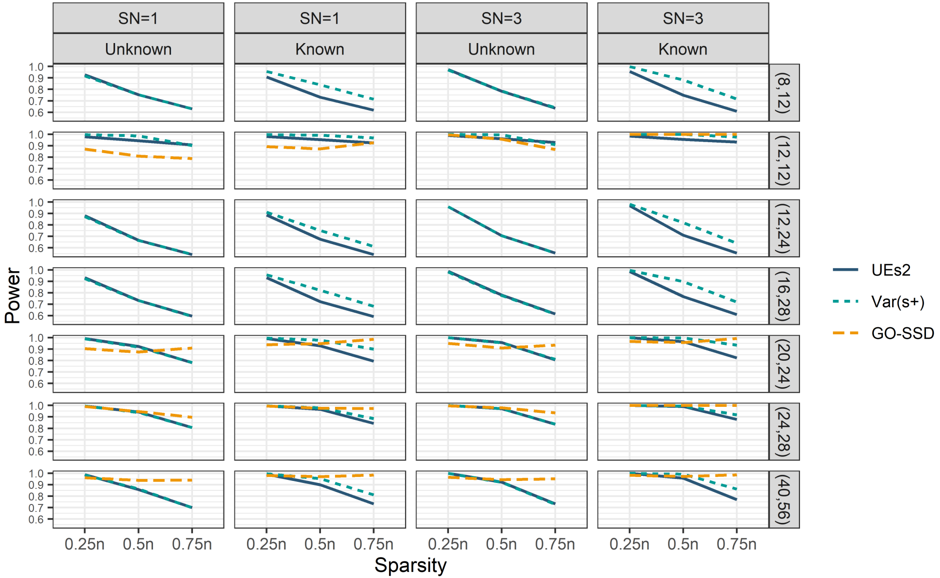
1. Randomly assigning columns to be active based on effect sparsity levels: 0.25n, 0.5n, and 0.75n.
2. Assigning the magnitude of the active effects generated randomly from $\text{Exp}(1) + \text{SN}$, where SN, meaning signal-to-noise ratio, was set to either 1 or 3.
3. Assigning the remaining inactive columns a coefficient from $N(0, 6^{-2})$.
4. Adjusting the signs of all coefficients to either be all positive (directions known) or mixed positive and negative (directions unknown).

SIMULATION PROTOCOL, CONT.

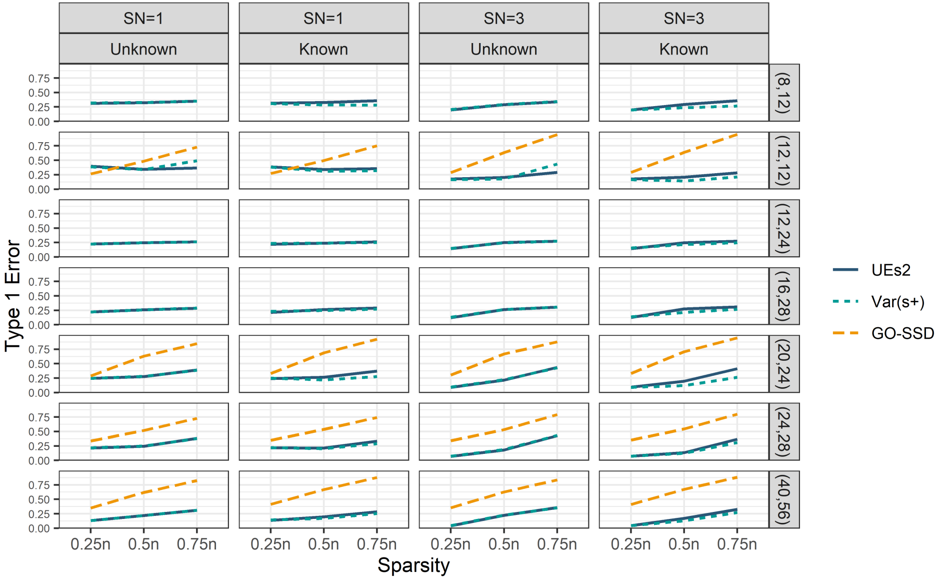
For each SSD and analysis combination,

1. GO-SSD and MaxPower,
2. Var($s+$) and the Gauss-Dantzig selector,
3. UE(s^2) and the Gauss-Dantzig selector,

we report the average the power (proportion of effects correctly classified as potentially active) and type 1 error (proportion of inactive effects classified as potentially active).



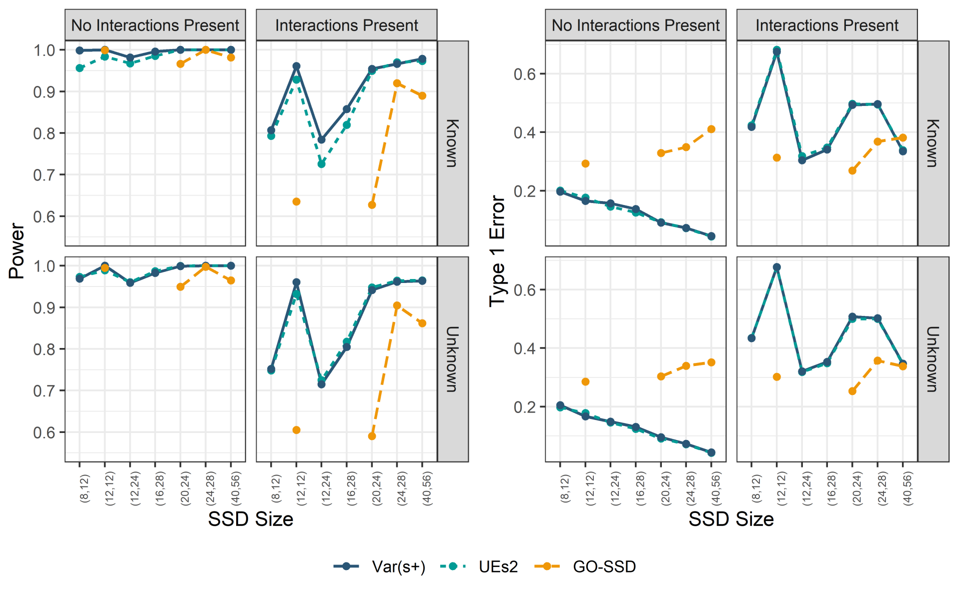
SN=1
Unknown
SN=1
Known
SN=3
Unknown
SN=3
Known
(8,12)
(12,12)
(12,24)
(16,28)
(20,24)
(24,28)
(40,56)
1.0
0.9
0.8
0.7
0.6
Power
0.25n
0.5n
0.75n
Sparsity
UEs2
Var(s+)
GO-SSD


SN=1
Unknown
SN=1
Known
SN=3
Unknown
SN=3
Known
(8,12)
(12,12)
(12,24)
(16,28)
(20,24)
(24,28)
(40,56)
0.75
0.50
0.25
0.00
0.25n
0.5n
0.75n
Type 1 Error
Sparsity
UEs2
Var(s+)
GO-SSD

SIMULATION WITH INTERACTIONS PRESENT

We consider the case where interactions are present in the true model, but ignored in the analysis.

1. We generated 5000 responses containing two interaction effects exhibiting weak heredity with the active main-effect columns
2. We fixed the factor sparsity to 0.25n for the main effects.
3. Both main and interaction effects were assigned $SN = 3$.
4. We considered both known and unknown effect directions.


No Interactions Present
Interactions Present
Known
Unknown
Power
Type 1 Error
1.0
0.9
0.8
0.7
0.6
0.2
0.4
0.6
(8,12)
(12,12)
(12,24)
(16,28)
(20,24)
(24,28)
(40,56)
SSD Size
Var(s+)
UEs2
GO-SSD

GO-SSD+MaxPower

1. Conservative screening
2. Measure of sparsity
3. Easy construction
4. Limited design sizes
5. Less robust to interactions

Var(s+)+Dantzig Selector

1. More aggressive screening
2. Algorithmically constructed
3. Flexible design sizes
4. More robust to interactions

SSDs should be considered on their own terms, with experimental goals and analysis methods specified and effectively exploited.

CONCLUSION

SSDs should be considered on their own terms, with experimental goals and analysis methods specified and effectively exploited.

Maria Weese
weeseml@miamioh.edu
Twitter: @MFWeese

Jones, B., Majumdar, D. (2014). Optimal supersaturated designs. Journal of the American Statistical Association, 109(508), 1592-1600.

Jones, B., Lekivetz, R., Majumdar, D., Nachtsheim, C. J., Stallrich, J. W. (2019). Construction, Properties, and Analysis of Group-Orthogonal Supersaturated Designs. In press: Technometrics.

Weese, M. L., Smucker, B. J., Edwards, D. J. (2015). Searching for powerful supersaturated designs. Journal of Quality Technology, 47(1), 66-84.

Weese, M. L., Edwards, D. J., Smucker, B. J. (2017). A criterion for constructing powerful supersaturated designs when effect directions are known. Journal of Quality Technology, 49(3), 265-277.

Appendix

Note that both $n, m = 0 \pmod 4$ and values of k are restricted since $w > p/2$ to prevent complete confounding within a group.

To construct the $n = 20$, $k = 24$ design GO-SSD:

1. Create a Hadarmard matrix with $m = 4$.
2. Construct $\mathbf{T}_{5\times 8}$ using the first 5 columns of a Hadarmard matrix with $m = 8$.
3. Take the Kronecker product of $\mathbf{H}_4$ and $\mathbf{T}_{5\times 8}$ giving $n = 4 \times 5 = 20$ and $k^* = 4 \times 8 = 32$.
4. This GO-SSD has a group size $p = 8$ and 4 groups.
5. The first group contains the intercept column and the 7 “fake” factor columns for estimating σ^2. This group is not shown in the correlation plot on slide 12.

Stage 1: Group Testing

1. Using the fake factor columns obtain an estimate for σ^2.
2. Test the significance of each group (recall the groups are orthogonal) using the estimate obtained in step 1.
3. Pool the degrees of freedom for any non-significant groups.

GO-SSD: MAXPOWER ANALYSIS

Stage 1: Group Testing

1. Using the fake factor columns obtain an estimate for σ^2.
2. Test the significance of each group (recall the groups are orthogonal) using the estimate obtained in step 1.
3. Pool the degrees of freedom for any non-significant groups.

Stage 2: Factor Testing

1. For the factors in the groups identified in stage 1, test all 1 factors models. Choose the model with the lowest residual sum of squares.
2. Repeat for models of up to size $\lfloor r/2 \rfloor$.
3. If all models of rank $\lfloor r/2 \rfloor$ are significant, designate all factors in the group as ”potentially active”.