Screening Strategies in the Presence of Interactions – comparison of strategies –

Angela Dean
The Ohio State University & University of Southampton

DEMA 2011
Isaac Newton Institute, Cambridge

Joint work with
Danel Draguljić, David Woods, Susan Lewis, Anna Vine

Outline

- Screening Strategies in the Presence of Interactions
- Screening designs
  - Two-stage group screening
  - Single stage supersaturated design
- Analysis methods
  - Dantzig Selector, LASSO, SCAD,
  - Bayesian Model Selection, Bayesian MAP estimation
- Simulation study comparing methods
- Issues for simulations

Screening

Screening is the process of

- sifting through a large number of potentially important factors
- in order to discover the few important (“active”) factors

Screening

Screening is the process of

- sifting through a large number of potentially important factors
- in order to discover the few important ("active") factors

- **Important** or **Active** means

  has a *substantial effect* on the
  response average or response variability
  through main effects or interactions

Screening

Screening is the process of

- sifting through a large number of potentially important factors
- in order to discover the few important ("active") factors

- **Important** or **Active** means

  has a *substantial effect* on the
  response average or response variability
  through main effects or interactions

- It does not mean "has a non-zero effect"

- There needs to be a cost-effective benefit to changing current product design

Screening

- Assumes factor sparsity

Screening

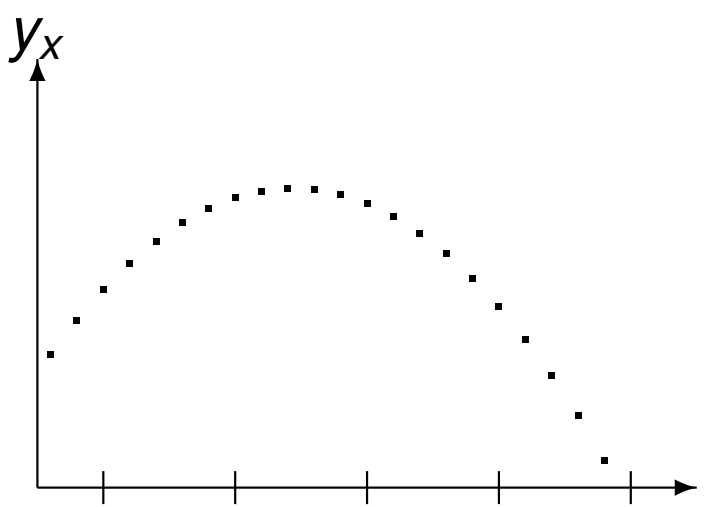
- Assumes factor sparsity
- Relatively few observations are taken
  - active factors are followed up later
  - for building detailed models for prediction and optimization

Screening

- Assumes factor sparsity
- Relatively few observations are taken
  - active factors are followed up later
  - for building detailed models for prediction and optimization
- Experiments with 2-level factors are common
- For factors with expected non-linear response, attempt to take observations at the steepest part of the response

Screening

- Assumes factor sparsity
- Relatively few observations are taken
  - active factors are followed up later
  - for building detailed models for prediction and optimization
- Experiments with 2-level factors are common
- For factors with expected non-linear response, attempt to take observations at the steepest part of the response



Screening Experiment: Cold Start Optimization

- Vine, Lewis, Dean, and Brunson, (2008), *Technometrics*.

Aim of experiment:

- To find which factors most affect the cold start performance of the new generation Jaguar engine

Screening Experiment: Cold Start Optimization

- 40 factors identified and assessed
- 10 factors in a screening experiment
- Investigated 10 main effects, 45 2-factor interactions

Traditional approach to screening

Two-stage approach

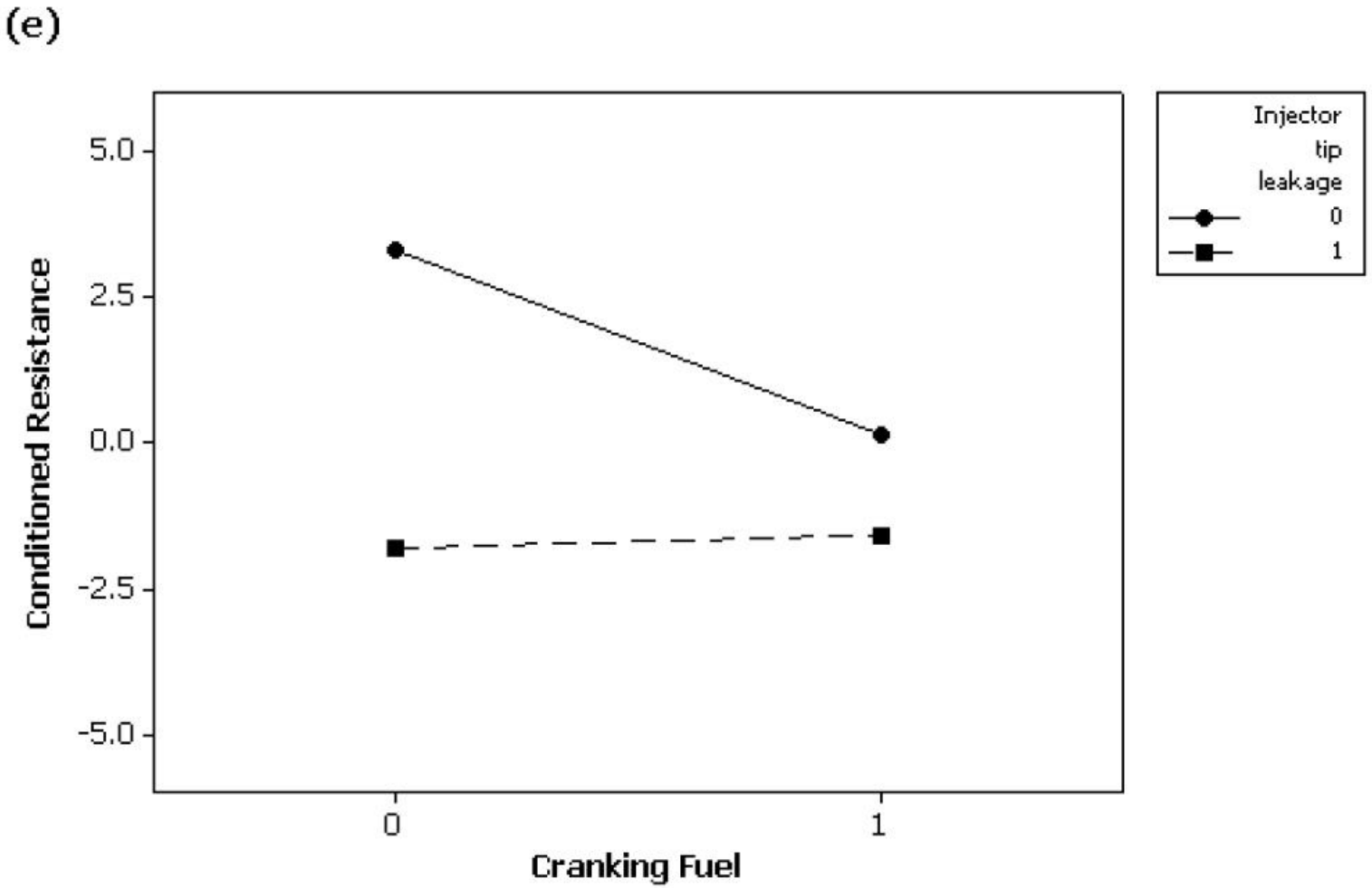
- Run a small experiment to estimate only main effects
- Select factors with large (active) main effects
- Use follow-up experimentation to investigate interactions
- Strategy works well if interactions are small or occur only between factors with large main effects
  - (*strong effect heredity*);
  - Hamada & Wu (1992); Chipman (1996)

Traditional approach to screening

Two-stage approach

- Run a small experiment to estimate only main effects
- Select factors with large (active) main effects
- Use follow-up experimentation to investigate interactions
- Strategy works well if interactions are small or occur only between factors with large main effects
  - (*strong effect heredity*);
  - Hamada & Wu (1992); Chipman (1996)
- In the previous example, this approach would not have worked

Interaction Illustration: Jaguar experiment



- Main effect of cranking fuel is small
- Would be screened out at stage 1

Incorporating interactions

- $\Rightarrow$ Important to measure interactions at stage 1
- Other studies stressing importance of screening interactions
  - Phoa, Wong, and Xu, 2009
  - Scinto, Wilkinson, and Lin, 2011

Screening Strategies

for investigating interactions

- Two-stage group screening
- Single-stage supersaturated design

I. Two-stage Group Screening

	$\{A, B\}$	$\{C, D\}$	$\{E, F\}$	$\{G, H\}$	$\{J, K\}$
	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
Grouped Factors	G_1	G_2	G_3	G_4	G_5

Stage 1: perform an experiment on the **grouped factors**

- Factors within a group always have matching levels

Stage 2: dismantle active groups and experiment on their **individual factors**

I. Two-stage Group Screening

Stage 1: **grouped factors**

- Estimate main effects and interactions
- Grouped factors with “large” effects are declared active

I. Two-stage Group Screening

Stage 1: **grouped factors**

- Estimate main effects and interactions
- Grouped factors with “large” effects are declared active

Stage 2: **individual factors** witihin a group

If group factor

- was inactive, hold individual factors at nominal levels
- was active, measure main effects and two-factor interactions among factors in the group

If two grouped factors

- had active interaction, measure interactions between individual factors in these groups

I. Two-stage Group Screening

- Aliasing among main effects and two-factor interactions is induced by the grouping
- e.g. 10 factors in groups of 2: {A , B}; {C , D}; {E , F}; {G , H}; {J , K};

I. Two-stage Group Screening

- Aliasing among main effects and two-factor interactions is induced by the grouping
- e.g. 10 factors in groups of 2: {A , B}; {C , D}; {E , F}; {G , H}; {J , K};

$$
\begin{array}{lccccccc}
mean = AB = CD = EF = GH = JK & & & & & & & \\
G_i \Rightarrow & A = B & & C = D & & E = F & & G = H \quad J = K \\
G_1 G_2 \Rightarrow & AC & = & BD & = & AD & = & BC \\
G_1 G_3 \Rightarrow & AE & = & BF & = & AF & = & BE \\
 & \vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

I. Two-stage Group Screening

- Aliasing among main effects and two-factor interactions is induced by the grouping
- e.g. 10 factors in groups of 2: {A , B}; {C , D}; {E , F}; {G , H}; {J , K};

$$
\begin{array}{lccccccc}
mean = AB = CD = EF = GH = JK & & & & & & & \\
G_i \Rightarrow & A = B & & C = D & & E = F & & G = H \quad J = K \\
G_1 G_2 \Rightarrow & AC & = & BD & = & AD & = & BC \\
G_1 G_3 \Rightarrow & AE & = & BF & = & AF & = & BE \\
 & \vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

- each set of aliased two-factor interactions involves factors from within only two groups

I. Two-stage Group Screening

- Aliasing among main effects and two-factor interactions is induced by the grouping
- e.g. 10 factors in groups of 2: {A , B}; {C , D}; {E , F}; {G , H}; {J , K};

$$
\begin{array}{llllllll}
mean = AB = CD = EF = GH = JK & & & & & & & \\
G_i \Rightarrow & A = B & & C = D & & E = F & & G = H \quad J = K \\
G_1 G_2 \Rightarrow & AC & = & BD & = & AD & = & BC \\
G_1 G_3 \Rightarrow & AE & = & BF & = & AF & = & BE \\
 & \vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

- each set of aliased two-factor interactions involves factors from within only two groups
- This remains true even if hundreds of factors
- and no matter the group size

II. Supersaturated designs

In a supersaturated design:

- there are fewer runs than effects to be estimated
- no main effects or interactions are completely aliased, but estimates of some main effects and interactions are correlated

II. Supersaturated designs

In a supersaturated design:

- there are fewer runs than effects to be estimated
- no main effects or interactions are completely aliased, but estimates of some main effects and interactions are correlated

- For good results, need design column correlations ≤ 0.33

II. Supersaturated designs

In a supersaturated design:

- there are fewer runs than effects to be estimated
- no main effects or interactions are completely aliased, but estimates of some main effects and interactions are correlated

- For good results, need design column correlations ≤ 0.33
- There are many different construction methods in the literature
- Or can use algorithm of Jones, Lin, & Nachtsheim (2008) to find best (*D*-optimal) SSD for estimating a given set of factorial effects

Analysis methods

Simulation: designs and analyses

- Single-stage Supersaturated Design (SSD)
  - SCAD: Smoothly Clipped Absolute Deviation
  - Lasso: Least Absolute Shrinkage and Selection Operator
  - DS: Gauss-Dantzig Selector
  - Bayes: Bayesian Model Selection
  - Bayes-MAP: Bayesian Model Averaging with maximum a posteriori (MAP) estimation

Simulation: designs and analyses

- Single-stage Supersaturated Design (SSD)
  - SCAD: Smoothly Clipped Absolute Deviation
  - Lasso: Least Absolute Shrinkage and Selection Operator
  - DS: Gauss-Dantzig Selector
  - Bayes: Bayesian Model Selection
  - Bayes-MAP: Bayesian Model Averaging with maximum a posteriori (MAP) estimation
- Two-stage Group Screening
  - GS5: Group Screening with R^2
  - GS5-DS: Group Screening with Dantzig Selector
  - GS5-B: Group Screening with Bayesian Model Selection
  - GS5-B-MAP: Grp Screening with Bayesian MAP estimation

Tuning Parameters

SCAD, Lasso and DS are all penalized regression methods

- Choice of tuning parameters in the penalty functions affects performance of each procedure

Tuning Parameters

SCAD, Lasso and DS are all penalized regression methods

- Choice of tuning parameters in the penalty functions affects performance of each procedure

- For Lasso and Dantzig Selector, tuning parameters chosen using a modified AIC
- For SCAD, used generalized cross-validation

Tuning Parameters

SCAD, Lasso and DS are all penalized regression methods

- Choice of tuning parameters in the penalty functions affects performance of each procedure

- For Lasso and Dantzig Selector, tuning parameters chosen using a modified AIC
- For SCAD, used generalized cross-validation

Small effects less than a threshold t screened out

Two Bayesian Analysis Methods

1. Select union of effects in models with high posterior prob. (e.g. $>$ .3 of the probability of the posterior modal model.)

Two Bayesian Analysis Methods

1. Select union of effects in models with high posterior prob. (e.g. $>$ .3 of the probability of the posterior modal model.)

2. Select effects whose regression coefficients have *maximum a posteriori* (MAP) estimates $> t$

Two Bayesian Analysis Methods

1. Select union of effects in models with high posterior prob. (e.g. $>$ .3 of the probability of the posterior modal model.)

2. Select effects whose regression coefficients have *maximum a posteriori* (MAP) estimates $> t$

– tuning parameters in the prior distributions chosen via graphical methods

Two Bayesian Analysis Methods

1. Select union of effects in models with high posterior prob. (e.g. $>$ .3 of the probability of the posterior modal model.)

2. Select effects whose regression coefficients have *maximum a posteriori* (MAP) estimates $> t$

– tuning parameters in the prior distributions chosen via graphical methods

–for group screening, we developed prior distributions at stage 2 based on the posterior distributions at stage 1

Simulation Study Details

to compare designs and analysis methods

Simulation: designs

Number of factors f equal to 10 or 15 or 20

Simulation: designs

Number of factors f equal to 10 or 15 or 20

I. Group screening used

- stage 1: 2_V^{5-1} fraction with 5 groups
- stage 2: adapted alg. of Jones, Lin, and Nachtsheim (2008) to generate design with $n = \#$ effects to be estimated $+5$

Simulation: designs

Number of factors f equal to 10 or 15 or 20

I. Group screening used

- stage 1: 2_V^{5-1} fraction with 5 groups
- stage 2: adapted alg. of Jones, Lin, and Nachtsheim (2008) to generate design with $n = \#$ effects to be estimated $+5$

II. Single stage supersaturated design

- generated from alg. of Jones, Lin, and Nachtsheim (2008)
- size = median number of runs required by Group Screening

Simulation: designs

Number of factors f equal to 10 or 15 or 20

I. Group screening used

- stage 1: 2_V^{5-1} fraction with 5 groups
- stage 2: adapted alg. of Jones, Lin, and Nachtsheim (2008) to generate design with $n = \#$ effects to be estimated $+5$

II. Single stage supersaturated design

- generated from alg. of Jones, Lin, and Nachtsheim (2008)
- size = median number of runs required by Group Screening
  - For $f = 10, p = 56$ effects, $n = 32$.
  - For $f = 15, p = 121$ effects, $n = 58$.
  - For $f = 20, p = 211$ effects, $n = 94$.

Simulation: effect size generation

- Every main effect is active with probability q_{me}
  - ($q_{me} \in \{.05, .1, .15, .2\}$)

Simulation: effect size generation

- Every main effect is active with probability q_{me}
  - ($q_{me} \in \{.05, .1, .15, .2\}$)
- A two-factor interaction is active with conditional probability
  - $w_{00} = 0.005$; $w_{10} = w_{01} = 0.125$; $w_{11} = 0.25$

Simulation: effect size generation

- Every main effect is active with probability q_{me}
  - ($q_{me} \in \{.05, .1, .15, .2\}$)
- A two-factor interaction is active with conditional probability
  - $w_{00} = 0.005$; $w_{10} = w_{01} = 0.125$; $w_{11} = 0.25$
- Generated active effect sizes from $N(6, 1), N(12, 4), N(24, 4)$
- Generated inactive effect sizes from $N(0, 1)$ or $N(0, 16)$

Simulation: data generation

- Given the generated effect sizes β
- and the model $y = X\beta + \epsilon$

Simulation: data generation

- Given the generated effect sizes β
- and the model $y = X\beta + \epsilon$
- generated data y for given design X with error ϵ from $N(0, 1)$

Simulation: data generation

- Given the generated effect sizes β
- and the model $y = X\beta + \epsilon$
- generated data y for given design X with error ϵ from $N(0, 1)$
- Analyzed the data using each of the specified methods
- Repeated a total of 1,000 times for each setting

Simulation: performance measures

To compare methods,

obtained boxplots of following measures of goodness:

Simulation: performance measures

To compare methods,

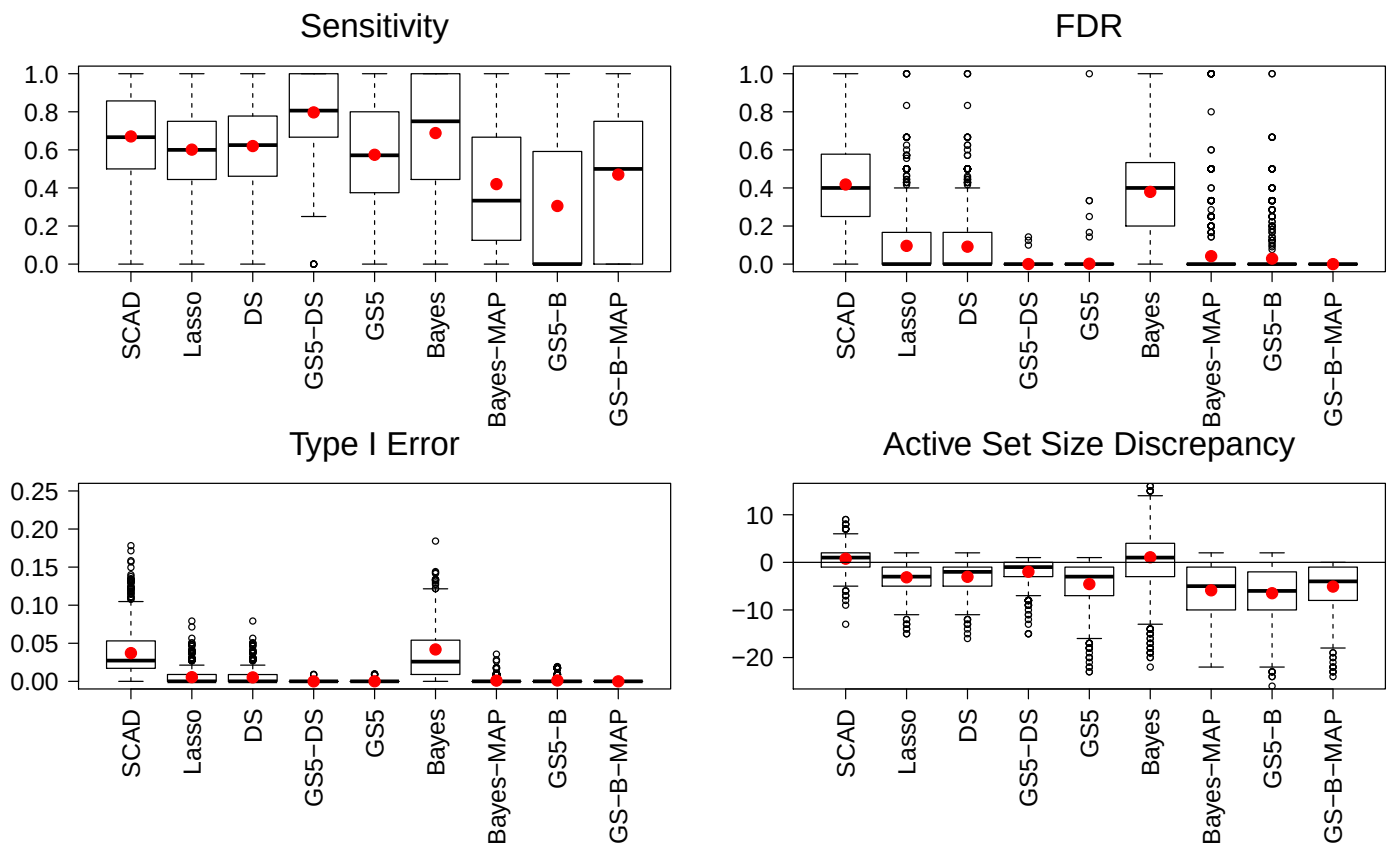
obtained boxplots of following measures of goodness:

- Sensitivity: proportion of active effects identified as active
- Type I error: proportion of inactive effects identified as active
- False discovery rate: proportion of effects identified as active that are truely inactive
- Active Set Size Discrepancy:

  true # active effects $-$ declared # active effects

Comparisons of designs and analyses; f=15 factors

$[q_{me} = 0.20$, Active $N(24, 4)$, Inactive $N(0, 16)$]



3b.pdf

Comparisons of designs and analyses—summary

In general:

- GS5-DS seems to perform the best overall
- followed by Lasso and DS in the supersaturated design

Comparisons of designs and analyses—summary

In general:

- GS5-DS seems to perform the best overall
- followed by Lasso and DS in the supersaturated design
- SCAD in the ssd tended to have high FDR
- GS5 (using R^2) tended to have low sensitivity

Comparisons of designs and analyses—summary

In general:

- GS5-DS seems to perform the best overall
- followed by Lasso and DS in the supersaturated design
- SCAD in the ssd tended to have high FDR
- GS5 (using R^2) tended to have low sensitivity
- Bayesian model selection in ssd sometimes had very high Type I error rate
- Bayesian methods had a greater drop in performance for the difficult cases than the frequentist methods

Issues about simulation studies

- Columns of a SSD assigned to active effects have a large bearing on how easily the active effects can be detected.

Issues about simulation studies

- Columns of a SSD assigned to active effects have a large bearing on how easily the active effects can be detected.
- In a simulation, assign active effects at random to columns, and compute summary measures over a large number of such assignments.

Issues about simulation studies

- For comparison, the sizes of group screening and supersaturated designs need to match
  - We used n for SSD size = median number of runs for group screening procedure
  - Possible to match more closely and compare performance on each individual run of the simulation.

Issues about simulation studies

- For comparison, the sizes of group screening and supersaturated designs need to match
  - We used n for SSD size = median number of runs for group screening procedure
  - Possible to match more closely and compare performance on each individual run of the simulation.
- In a simulation study, automatic decisions have to be made at stage 1 of group screening about active groups.
  - More control can be exercised in a practical experiment.

Issues about simulation studies

- There is a small chance in group screening that aliased active effects may cancel each other and not be taken through to stage 2. This possibility can be minimized for the main effects by eliciting the “direction” of the effects and setting the levels accordingly

Issues about simulation studies

- There is a small chance in group screening that aliased active effects may cancel each other and not be taken through to stage 2. This possibility can be minimized for the main effects by eliciting the “direction” of the effects and setting the levels accordingly
- In group screening, aliased small effects may amalgamate, resulting in a non-active group appearing to be active at stage 1. However, a high FDR can be corrected at stage 2
- Such amalgamation can also happen in any fractional factorial experiment

Future extensions

- Active main effects and interaction effects could be drawn from different distributions.

Future extensions

- Active main effects and interaction effects could be drawn from different distributions.
- To mimic a situation where effects can be classified in terms of their likelihood of being active, the active effect distributions could be set differently

Future extensions

- Active main effects and interaction effects could be drawn from different distributions.
- To mimic a situation where effects can be classified in terms of their likelihood of being active, the active effect distributions could be set differently
- We used the same threshold t for screening out small effects regardless of whether the effect was a main effect or interaction. Could be set differently.

Future extensions

- Active main effects and interaction effects could be drawn from different distributions.
- To mimic a situation where effects can be classified in terms of their likelihood of being active, the active effect distributions could be set differently
- We used the same threshold t for screening out small effects regardless of whether the effect was a main effect or interaction. Could be set differently.
- The Bayesian approaches are more computationally intensive than the frequentist methods. Need to refine the procedures so that larger numbers of factors can be handled
- Need to tune Bayes procedures to be more successful overall

Thank you!