Large Row-Constrained Supersaturated Designs for High-throughput Screening

Byran Smucker

Department of Public Health Sciencies, Division of Biostatistics
Henry Ford Health, Detroit, MI

June 17, 2025

Nonclinical Biostatistics Conference, Rutgers University

HENRY FORD HEALTH

This work is a team effort

Large Row-Constrained Supersaturated Designs for High-throughput Screening

Byran J. Smucker* [1], Stephen E. Wright[1], Isaac Williams[1], Richard C. Page[2], Andor J. Kiss[3], Surendra Bikram Silwal[2], Maria Weese[4], and David J. Edwards[5]

[1]Department of Statistics, Miami University, Oxford, OH, 45056
[2]Department of Chemistry and Biochemistry, Miami University, Oxford, OH, 45056
[3]Center for Bioinformatics and Functional Genomics, Miami University, Oxford, OH, 45056
[4]Department of Information Systems & Analytics, Miami University, Oxford, OH, 45056
[5]Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA, 23284

This work is a team effort

Brian Ahmer, PhD, The Ohio State University, Department of Microbiology
Meng Wu, PhD, The Ohio State University, Medicinal Chemistry Shared Resource - High-Throughput Screening Core
Emily Rego, The Ohio State University, Department of Microbiology
Zhihong Lin, PhD, The Ohio State University, Medicinal Chemistry Shared Resource - High-Throughput Screening Core

Outline

Outline

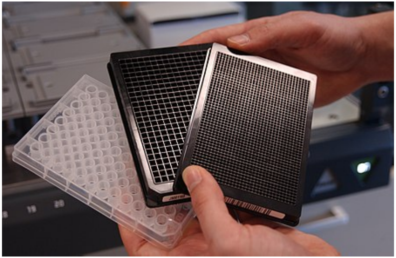
HENRY FORD HEALTH

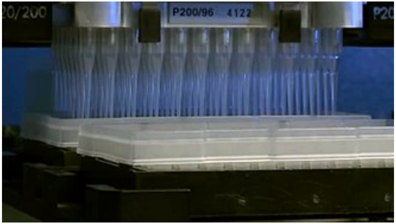
High Throughput Screening

Used in drug discovery and chemical biology.

High Throughput Screening

Used in drug discovery and chemical biology.

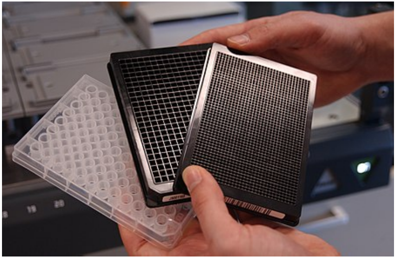


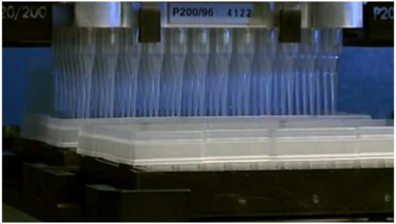


https://en.wikipedia.org/wiki/Microplate

High Throughput Screening

Used in drug discovery and chemical biology.





https://en.wikipedia.org/wiki/Microplate

Later, we'll discuss two examples related to the problem of antibiotic-resistant bacteria.

Take-Home Message

In many high-throughput screening campaigns, **pooling** multiple compounds per well will improve hit rates and reduce false positives.

HTS Design and Analysis Approaches

Typical: One-compound-one-well (OCOW). Relatively simple in terms of statistical design and analysis.

HTS Design and Analysis Approaches

Typical: One-compound-one-well (OCOW). Relatively simple in terms of statistical design and analysis.

Proposed: Constrained Row Screening (CRowS), which treats HTS as a multi-factor experimental design problem.

HTS Design and Analysis Approaches

Typical: One-compound-one-well (OCOW). Relatively simple in terms of statistical design and analysis.

Proposed: Constrained Row Screening (CRowS), which treats HTS as a multi-factor experimental design problem.

- Each compound is treated as a factor with two levels (absent, present)
- Each well is treated as an independent experimental unit
- Each well receives *multiple* compounds, *but no more than c compounds per well*
- The number of compounds k can be more than the number of wells n, so use machinery of *supersaturated designs*

HENRY FORD HEALTH

Supersaturated Designs

Two-level supersaturated designs (SSDs) use $n < k$ runs to study k factors.

Supersaturated Designs

Two-level supersaturated designs (SSDs) use $n < k$ runs to study k factors. Here is a toy $n = 6$, $k = 9$ example:

$$X = \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

SSDs generate data to estimate $\mathbf{y} = \beta_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where X is $n \times k$.

Supersaturated Designs

Two-level supersaturated designs (SSDs) use $n < k$ runs to study k factors. Here is a toy $n = 6$, $k = 9$ example:

$$X = \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

SSDs generate data to estimate $\mathbf{y} = \beta_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where X is $n \times k$. We use the Lasso to estimate $\boldsymbol{\beta}$.

Standard Supersaturated Design Criterion, UE(s^2)

Let $L = [\mathbf{1}, X]$ and $S = L^{'}L$ with elements s_{ij}.

Unconditional $E(s^2)$-criterion:

$$\text{UE}(s^2) = \binom{k+1}{2}^{-1} Q(X)$$

where $Q(X) = \sum_{0 \le i < j \le k} s_{ij}^2$.

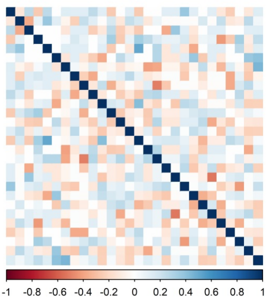
Standard Supersaturated Design Criterion, UE(s^2)

Let $L = [\mathbf{1}, X]$ and $S = L'L$ with elements s_{ij}.

Unconditional $E(s^2)$-criterion:

$$\text{UE}(s^2) = \binom{k+1}{2}^{-1} Q(X)$$

where $Q(X) = \sum_{0 \le i < j \le k} s_{ij}^2$.



S matrix for a 16×28 $UE(s^2)$-optimal design

Pooling in the Literature

Pooling multiple compounds in a single well has been previously proposed and studied (e.g., see reviews by Kainkaryam and Woolf 2009).

Elkin et al. (2015) calls pooling "controversial ... with a long history of limited success" but shows a successful example of a two compounds/well campaign.

Some recent works have proposed new pooling procedures: Ji et al. (2023) and Liu et al. (2024). Both use statistical methods to screen; neither study their approaches as supersaturated designs, nor do they compare with competing methods.

Outline

HENRY FORD HEALTH

CRowS Designs

We say an $n \times k$ design X is a Constrained Row Screening design if $n \leq k$, $[X]_{ij} = x_{ij}$, and the design solves the following optimization problem:

$$\min \quad UE(s^2) \tag{1}$$

$$\text{s.t.} \quad x_{ij} \in \{-1, 1\} \ \forall i, j \tag{2}$$

$$\sum_{j=1}^{k} x_{ij} \leq 2c - k \ \forall i \tag{3}$$

where c is the maximum number of compounds/well.

HENRY FORD HEALTH

CRowS Designs

We say an $n \times k$ design X is a Constrained Row Screening design if $n \leq k$, $[X]_{ij} = x_{ij}$, and the design solves the following optimization problem:

$$\min \quad UE(s^2) \tag{1}$$

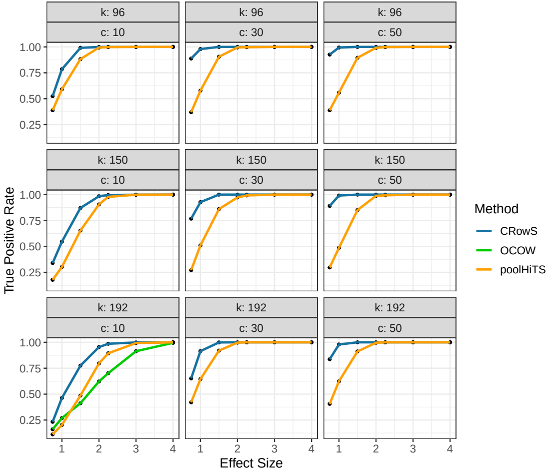
$$\text{s.t.} \quad x_{ij} \in \{-1, 1\} \; \forall i, j \tag{2}$$

$$\sum_{j=1}^{k} x_{ij} \leq 2c - k \; \forall i \tag{3}$$

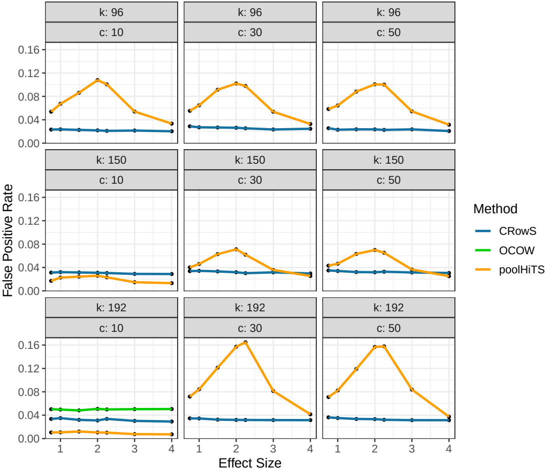
where c is the maximum number of compounds/well. We construct designs using a coordinate-exchange type algorithm.

HENRY FORD HEALTH

Comparisons: True Positive Rate



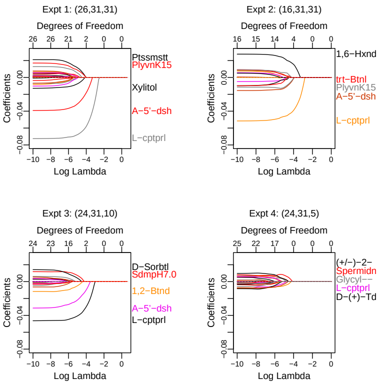
Comparisons: False Positive Rate



Outline

HENRY FORD HEALTH

Detecting Inhibitors of Metallo-β-lactamases



- Working with Rick Page's lab at Miami University
- Goal to find inhibitors of enzyme that causes bacterial resistance to antibiotics
- Four mini-experiments, spiked with a known inhibitor, `L-captopril`

HENRY FORD HEALTH

Searching for New Antimicrobials

Brian Ahmer's lab at The Ohio State University.

Searching for New Antimicrobials

Brian Ahmer's lab at The Ohio State University.

MtlD is an enzyme some bacteria use to process mannitol; without it, bacteria weaken or die when mannitol is present.

Searching for New Antimicrobials

Brian Ahmer's lab at The Ohio State University.

MtlD is an enzyme some bacteria use to process mannitol; without it, bacteria weaken or die when mannitol is present.

The lab developed a knockout strain without MtlD for several different bacterial species.

Searching for New Antimicrobials

Brian Ahmer's lab at The Ohio State University.

MtlD is an enzyme some bacteria use to process mannitol; without it, bacteria weaken or die when mannitol is present.

The lab developed a knockout strain without MtlD for several different bacterial species.

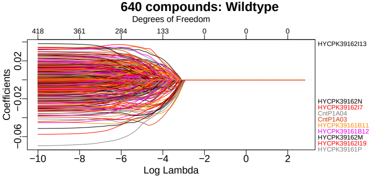
Goal: identify compounds that hit on Wildtype but not on Knockout

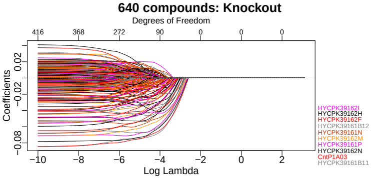
Pilot Studies with 320 Wells

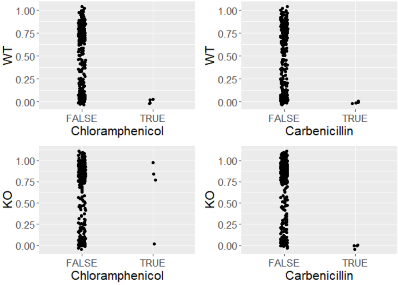
Check to make sure we can correctly detect a known hit, and check how aggressively we can reasonably screen.

Pilot Studies with 320 Wells

Check to make sure we can correctly detect a known hit, and check how aggressively we can reasonably screen.







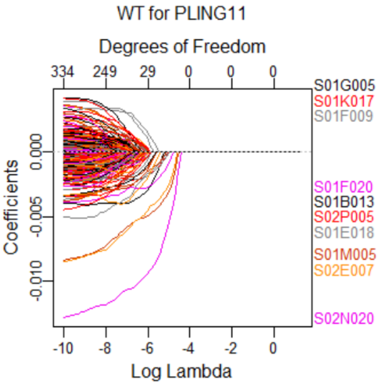
CntP1A04==Chloramphenicol and CntP1A03==Carbenicillen

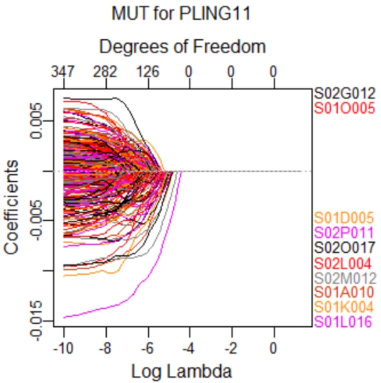
Initial 10,000+ Compound Screen

METHOD

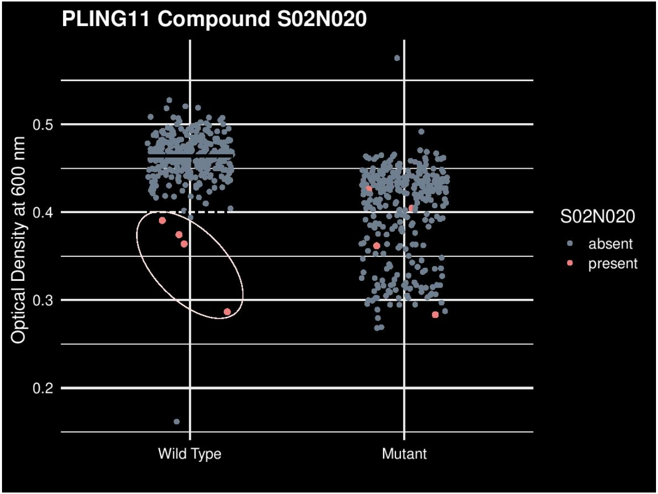
1. **Plan** the pools using a CRowS design
2. **Construct** the pools using a liquid handling system
3. **Obtain and clean** the assay data
4. **Generate** an initial hit list using the Lasso
5. **Trim** hit list as desired by isolating compounds that show consistent inhibition across replications

Out of 16 “PLINGs”, PLING 11 is notable





S02N020 from PLING 11 is promising



Outline

HENRY FORD HEALTH

Summary of Work

Our contributions:

1. CRowS, a statistically principled and high-quality solution to the pooling problem in HTS
2. a new class of *row-constrained* supersaturated designs motivated by these applications

Summary of Work

Our contributions:

1. CRowS, a statistically principled and high-quality solution to the pooling problem in HTS
2. a new class of *row-constrained* supersaturated designs motivated by these applications

Challenges:

1. Application-specific success
2. Interactions?

Summary of Work

Our contributions:

1. CRowS, a statistically principled and high-quality solution to the pooling problem in HTS
2. a new class of *row-constrained* supersaturated designs motivated by these applications

Challenges:

1. Application-specific success
2. Interactions?

Ongoing and future work: more applications; comparisons with other methods; handling interactions; theoretical development to better understand and construct the designs

HENRY FORD HEALTH

Thank you!

Byran Smucker: bsmucke1@hfhs.org

Find me on Twitter/X (@ByranSmucker) and LinkedIn

arXiv preprint: “Large Row-Constrained Supersaturated Designs for High-throughput Screening” (Smucker et al. 2024)

Outline

Features of Row-Constrained SSDs: $(n = 96, k = 144)$