



Optimal design of High Dimensional Experiments

Dr Joff Jones

Quantitative Biology & Data Sciences, Discovery
Sciences, Biopharma R&D, AstraZeneca

The Discovery Centre, Cambridge, UK

23rd September 2026



Outline

1. What are High Dimensional Experiments?
2. How high dimensional should they be?
3. Spatially optimal assay plate designs
4. Increasing accessibility to DoE with agentic AI



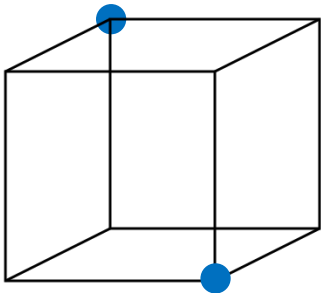
1

What are High Dimensional Experiments?



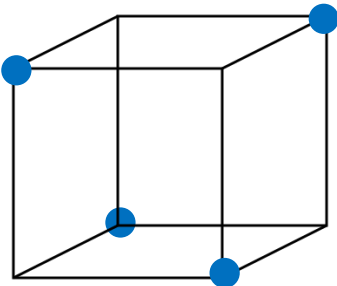
DoE is typically performed in multiple sequential experiments, minimising the total number of runs, but taking time

Preliminary
Exploration



Known settings
Maximal settings

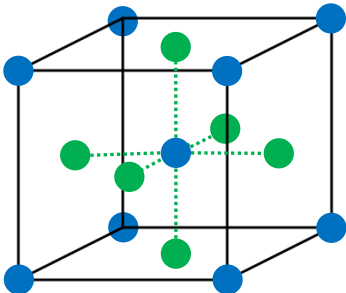
Factor
Screening



D-Optimal

$$D = \det[\mathbf{X}'\mathbf{X}]$$

Optimisation

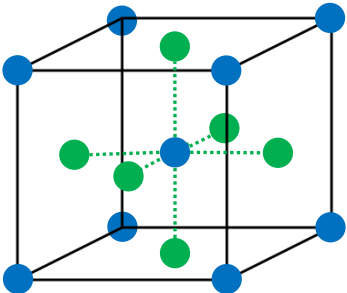


Response-surface
I-Optimal

$$I = \int_R f(\mathbf{x})'(\mathbf{X}'\mathbf{X})^{-1}f(\mathbf{x})d\mathbf{x} = \text{trace}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}]$$

$$\mathbf{M} = \int_R f(\mathbf{x})f(\mathbf{x})'d\mathbf{x}$$

Further
Optimisation



Response-surface
I-Optimal



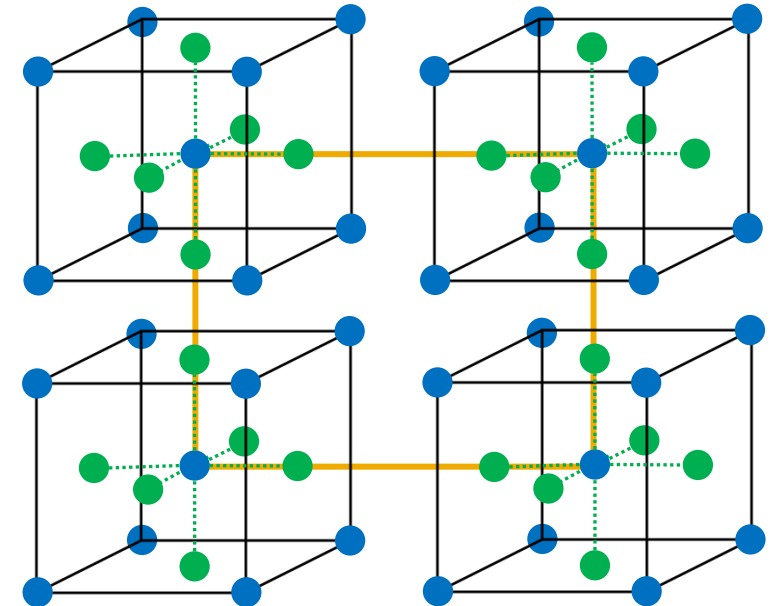
High Dimensional Experimentation (HDE) aims to maximise organisational efficiency

Sequential experiments takes time

- Cellular assays can be long (up to 2-3 weeks) & the biology can be complex

HDE can remove experimental cycles

- Minaturised assays (e.g. 384-well plates) minimise cost and enable high-capacity experiments
- Test more factors/levels and higher order effects
- Faster progress
- Broader scope for Optimisation



2

2nd or 3rd order HDEs? A
simulation study



When to design for 3rd order effects?



Emiel Ceunen



Dr Saif Hameed



Prof Peter Goos

Rationale

- 2nd order response surface models are the convention for optimization because 3rd order effects are assumed to be rare and small
- However, this is not necessarily true for preclinical assays with complex biology and highly interdependent factors - e.g. Reagent * (Reagent Concentration)²

Study design

We designed 180 run experiments with 8 continuous factors at 3 levels based on an empirical example. 100 simulations was performed per combination of the following:

- DoE model: 2nd or 3rd order
- Fitted model: 2nd or 3rd order
- Proportion of active 3rd order effects: 0, 5, 10, 15, 20, 30%
- Size of third order effects: 0.05, 0.125, 0.2

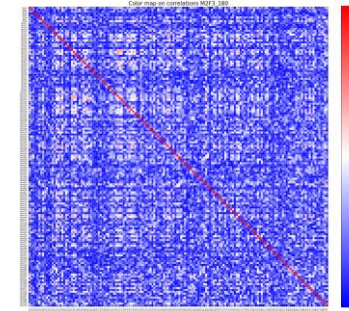
We assessed the **relative distance between the predicted and true maximum response**



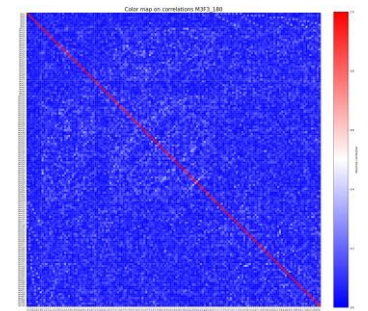
There is a tipping point when 3rd order DoEs perform best

- With 2nd order DoEs, parameter correlations involving 3rd order terms could be very high
 - This inflates variance and decreases power (63% vs 85%)
- 3rd order DoEs perform best when 3rd order effects are 0.125 or larger and 10% or more prevalent

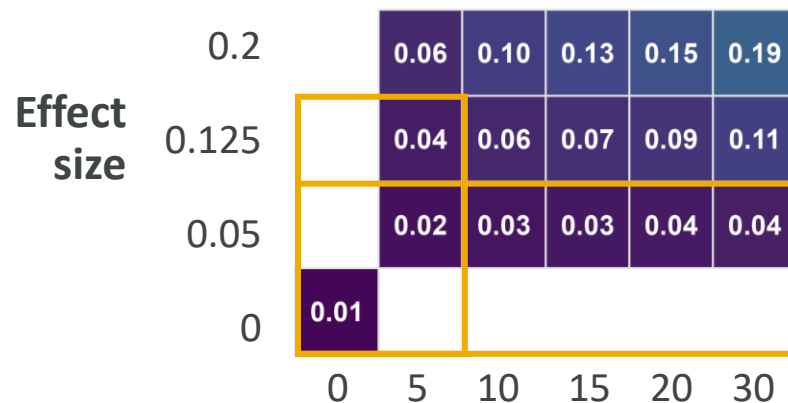
2nd order DoE, 3rd order fit



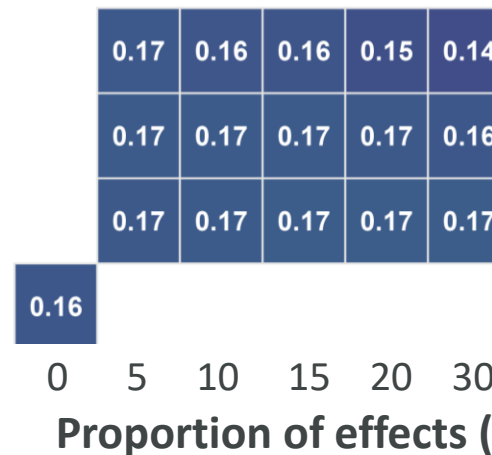
2nd order DoE, 3rd order fit



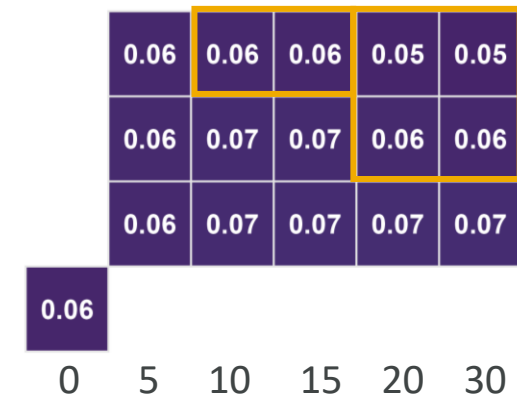
2nd order DoE, 2nd order fit



2nd order DoE, 3rd order fit



3rd order DoE, 3rd order fit



3

Spatially optimal assay plate designs



Designing for spatial blocks & trends



Emiel Ceunen



Dr Saif Hameed



Prof Peter Goos

Rationale

- Spatial effects are highly prevalent in preclinical assays and can confound or increase variance in effects of interest
- Covariate designs find the optimal design to estimate experimental factors and uncontrolled covariates
- 2-way split-plot designs introduce random blocks for columns and rows
- However, these methods in combination are not available in commercial software (e.g. JMP)

Method

- We developed a novel coordinate exchange algorithm
- It minimizes the average prediction variance (I-Optimal) for a design with two crossed random blocks (column & row) and a column*row response surface
- We tested this for a 3-factor response surface design on a 6x4 grid
- We then calculated the average prediction variance:

$$\begin{aligned}\text{Average prediction variance} &= \frac{\int_{\mathcal{X}} \mathbf{f}(\mathbf{x})^{\top} (\mathbf{X}^{\top} \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{X}} d\mathbf{x}} \\ &= 2^{-p} \text{tr} \left[(\mathbf{X}^{\top} \mathbf{V}^{-1} \mathbf{X})^{-1} \int_{\mathcal{X}} \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^{\top} d\mathbf{x} \right]\end{aligned}$$

Example model formula

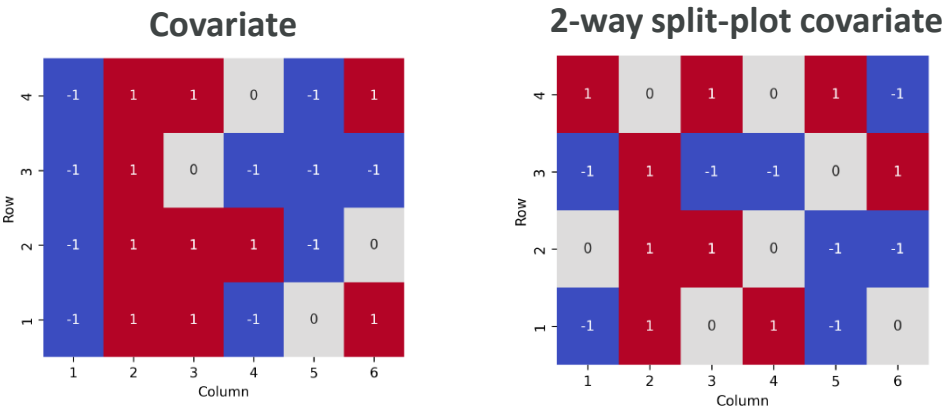
$$\begin{aligned}Y_{ij} = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ & + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\ & + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \\ & + \theta_1 z_1 + \theta_2 z_2 + \theta_{12} z_1 z_2 + \theta_{11} z_1^2 + \theta_{22} z_2^2 \\ & + \gamma_i + \delta_j + \varepsilon_{ij},\end{aligned}$$



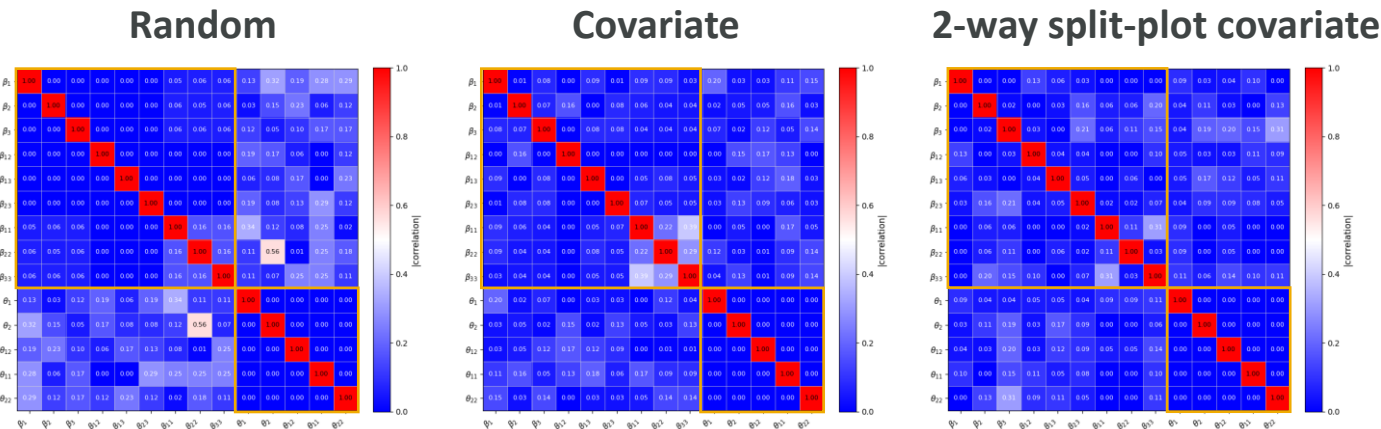
The 2-way split-plot covariate DoE performed best

- The covariate design did not prevent correlation between parameters and specific columns or rows
- Relative to the covariate design, it resulted in comparable correlations between parameter and spatial trends, and mean prediction variance

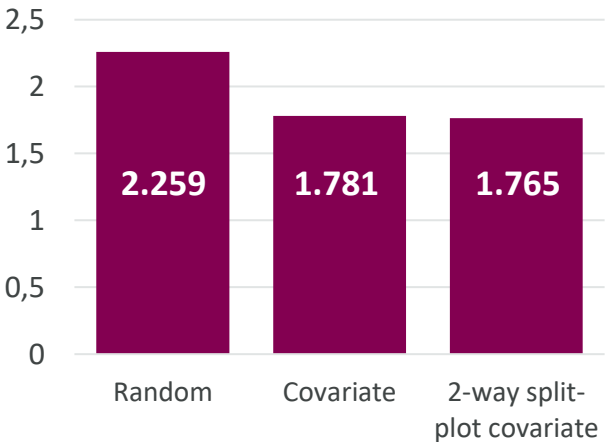
Level assignment of X2



Parameter & covariate correlation matrices



Comparison of mean prediction variance



4

Agentic DoE



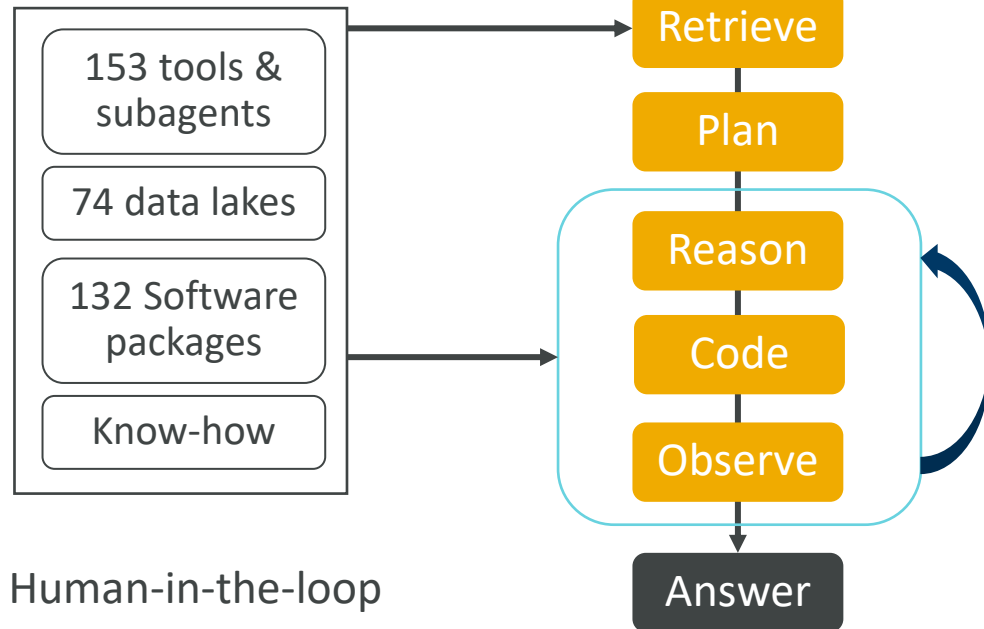
AI agent paired with know-how

Chen Qian



Biomedical coding agent

- Multi-agent system
- LLM (Claude Sonnet 4.6)
- Sandbox



- Human-in-the-loop

In vitro DoE know-how

- Step-by-step instructions for design generation & evaluation
- Identify response type
- Rules to infer the DoE model from an experimental goal
- Identify hard-to-change factors & blocking
- Code examples in skpr (R) for full-factorial, optimal, & split-plot designs

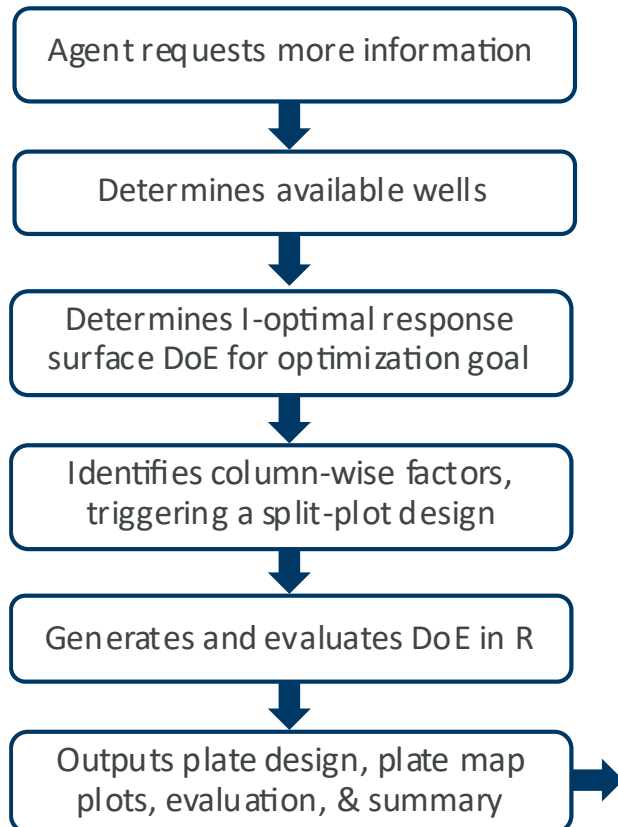


A split-plot DoE example

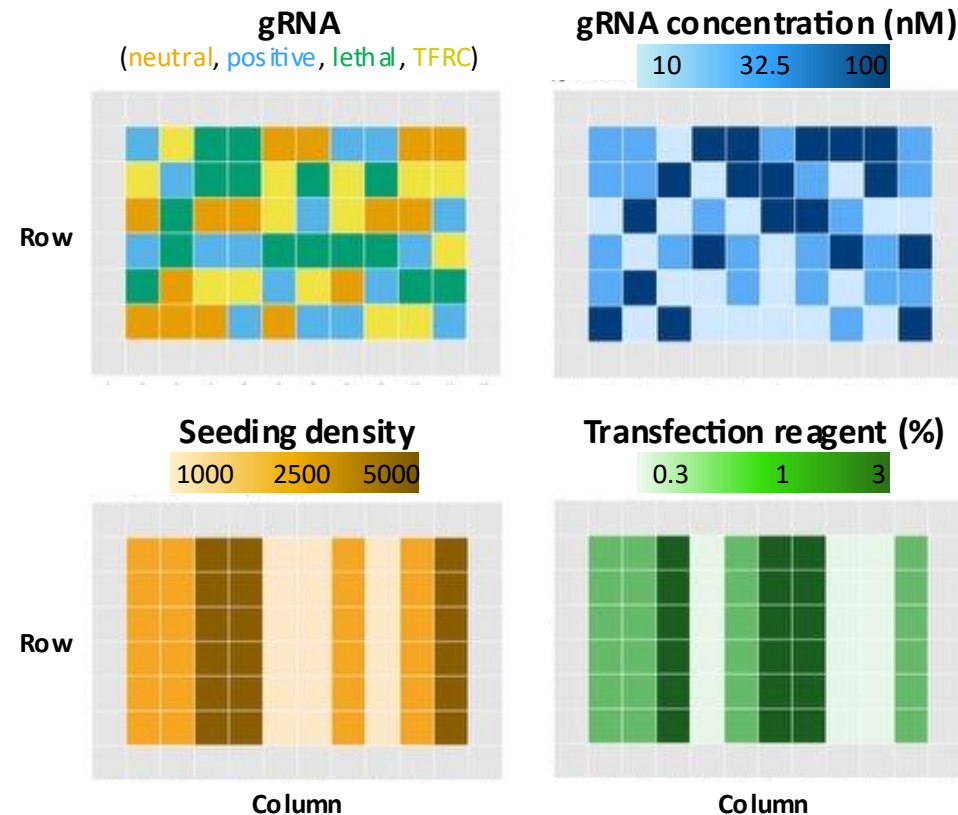


Please help me to design a lipid-based CRISPR gene-editing optimisation experiment.

Agentic workflow



Generated plate maps



Summary

1. High Dimensional Experiments acquire high-quality data and achieve high operational efficiency
2. 3rd order DoEs are beneficial with prevalence $\geq 10\%$ and effect sizes ≥ 0.125
3. The 2-way crossed-block covariate design minimises parameter correlation with spatial trends and blocks
4. Agentic AI with DoE know-how can make expert-level DoE widely accessible to scientists

Acknowledgements

KU Leuven – Bioscience Engineering

- Emiel Ceunen
- Dr Saif Hameed
- Prof Peter Goos

AstraZeneca

- Chen Qian
- Yin Hai Wang
- Amrutha Saseendran
- Chen Jin
- Valentina Quarantotti
- Helena Peilot Sjögren



Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful.

AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK
+44(0)203 749 5000
www.astrazeneca.com



I-optimal split-plot design



High Dimensional Experimentation (HDE)

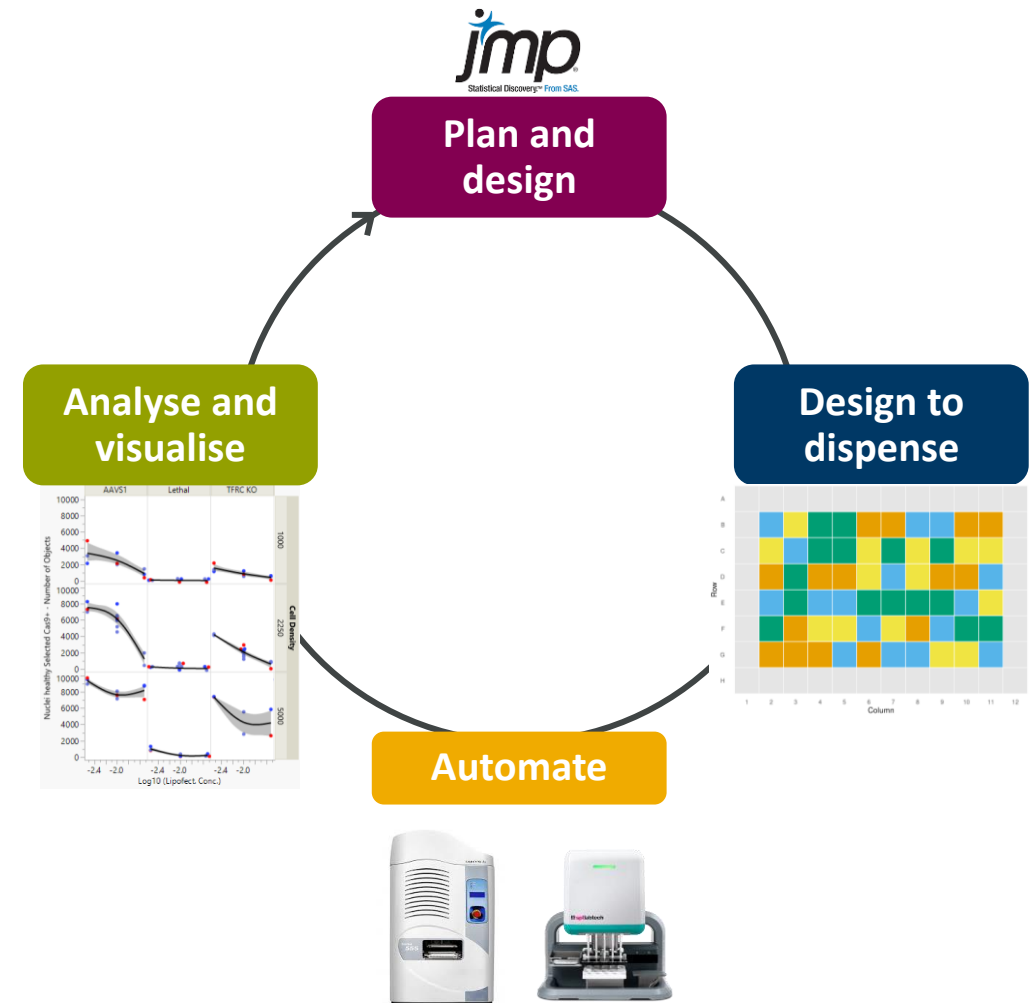
HDE is an integrated strategy for *in vitro* experimental **plate design**, **lab automation**, and **data analysis**

HDE generates high-quality data with fewer experiments

- Minimises handling errors with **automation**
- Controls for noise variability using **randomisation**
- Uses **robust statistical modelling** to predict outcomes
- **Reduces** wells, cells, reagents and overall cost 
- Tests **multiple factors and their interactions** in one experiment
- Generates **FAIR data**

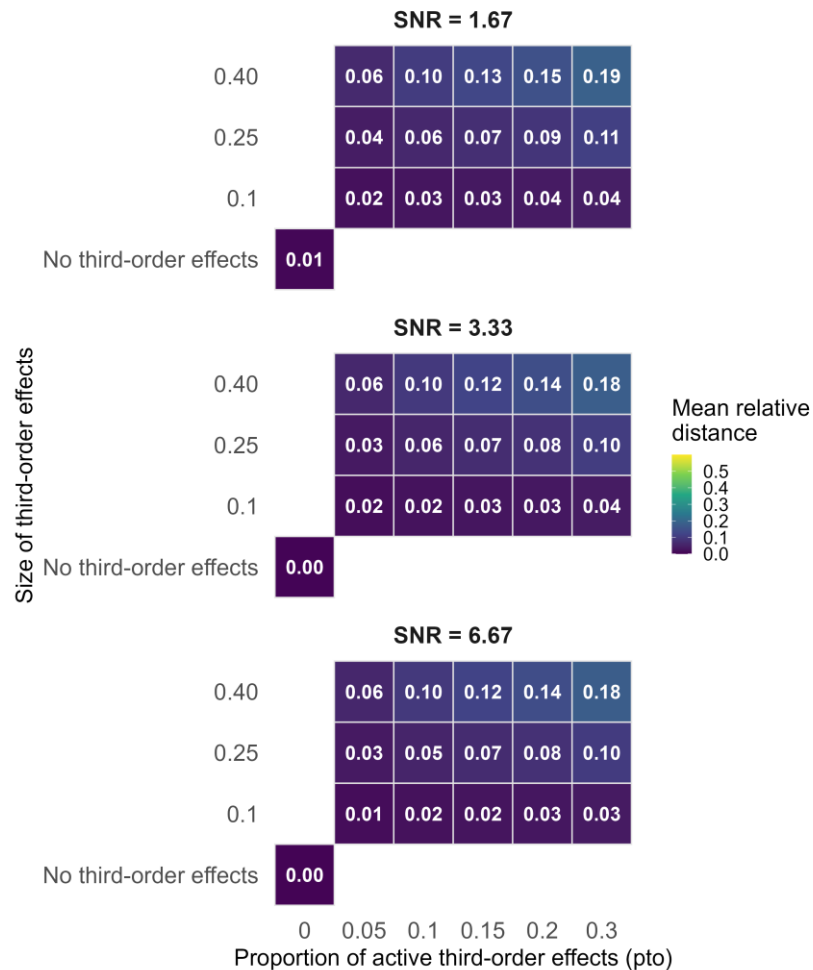


- ✓ **Applicable for most experiments with two or more factors!!**
 - ✓ Enables otherwise “impossible” experiments

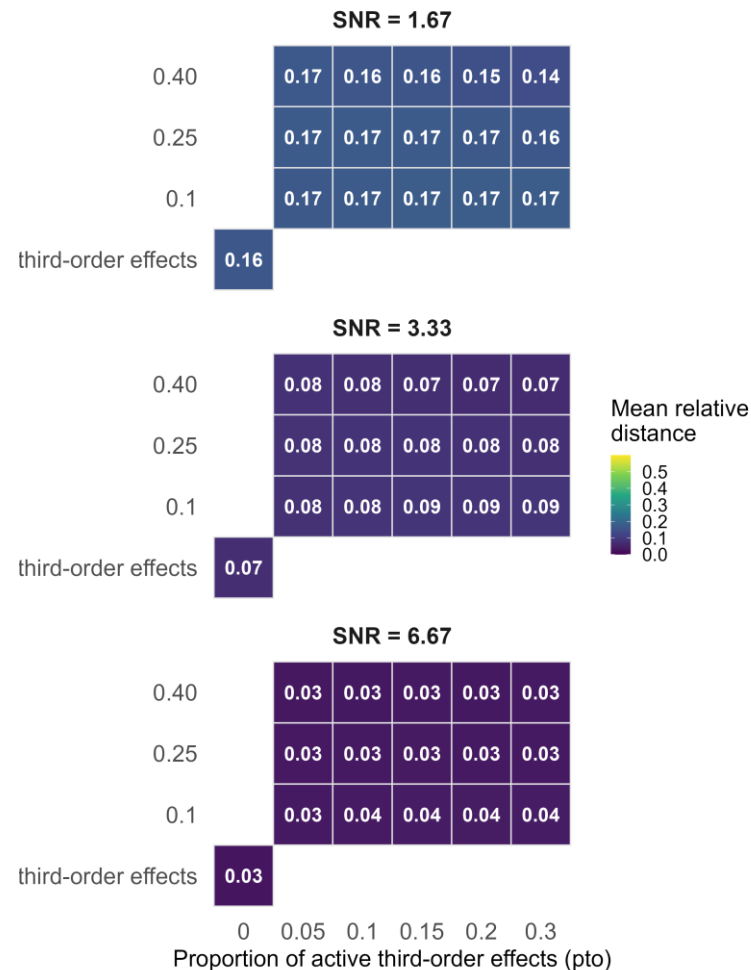


Relative distances at different SNR

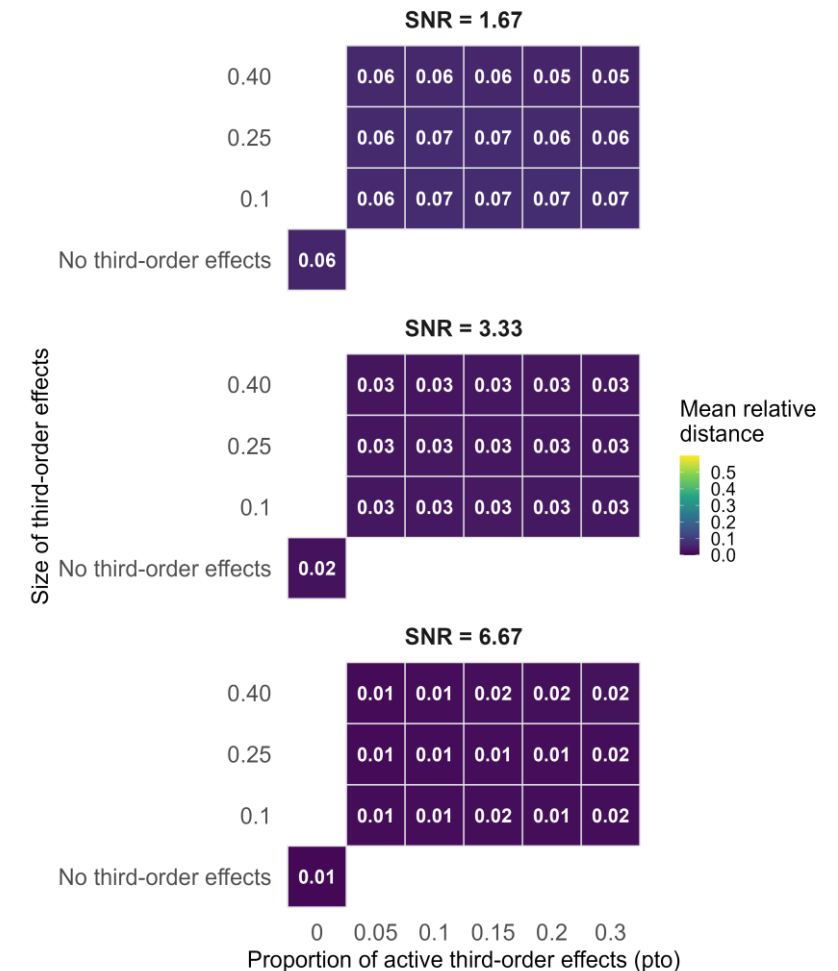
2nd order DoE, 2nd order fit



2nd order DoE, 3rd order fit



3rd order DoE, 3rd order fit



$$\beta_{\text{third-order}} = \text{size}_{\text{3rd}} \cdot (\text{SNR} \times \sigma + \omega), \quad \omega \sim \text{Exponential}(\lambda = 5).$$

Sigma = 0.3

