Statistical Detection of Synergy: New Methods and a Comparative Study

Olivier Thas

UHASSELT
I-BioStat
KU LEUVEN
Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Introduction

This is joint work with Annelies Tourny, Bie Verbist, Stijn Hawinkel, Maxim Nazarov, Kathy Mutambanengwe and Luc Bijnens.

This talk is based on

- Van der Borght, K., Tourny, A., Bagdziunas, R., Thas, O., Nazarov, M., Turner, H., ... and Ceulemans, H. (2017). BIGL: Biochemically Intuitive Generalized Loewe null model for prediction of the expected combined effect compatible with partial agonism and antagonism. *Scientific reports*, 7(1), 1-9.
- Thas, O., Tourny, A., Verbist, B., Hawinkel, S., Nazarov, M., Mutambanengwe, K. and Bijnens, L. (2022). Statistical detection of synergy: New methods and a comparative study. *Pharmaceutical Statistics*, 21(2), 345-360.
- The BIGL R package.

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Introduction

Combination therapies are therapies in which two or more drugs are combined.

It is increasingly adopted as standard of care for various diseases, e.g. tuberculosis, malaria, HIV, and many advanced cancers.

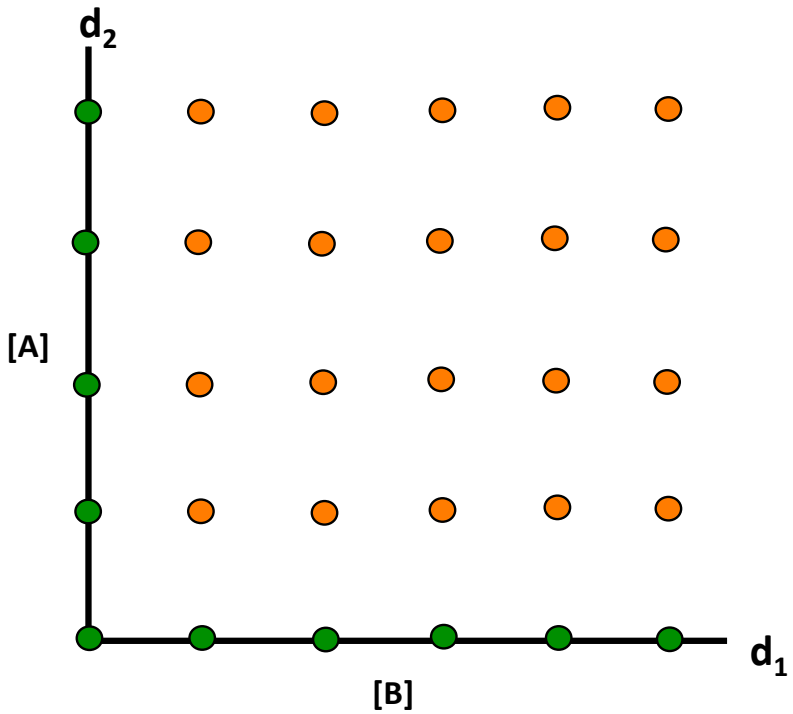
Advantages:

- improve treatment response
- minimize development of monotherapy resistance
- sometimes lower doses are possible (avoiding intolerable dose ranges), hence reducing side-effects

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Introduction

Drug combinations are screened early in the drug development process for synergistic effects, for example with a checkerboard design.

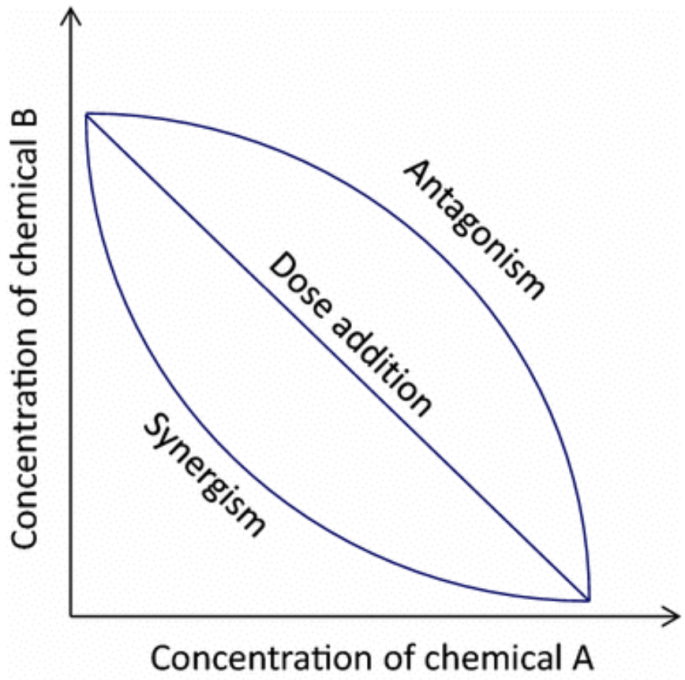


On-axis points (monotherapy)

Off-axis points (comb. exp.)

But what is synergism?

We start with defining aditivity of two drugs.
Dose additivity: for resulting in the same response,
$d_{\mathsf{AB}} = d_{\mathsf{A}} + d_{\mathsf{B}}$.



Hernández, Gil and Lacasaña. Archives of Toxicology (2017)

Introduction

Many models for additivity have been proposed: null models.

These models are often

- mechanistically inspired
- biochemically interpretable
- based on the monotherapeutic dose-response curves

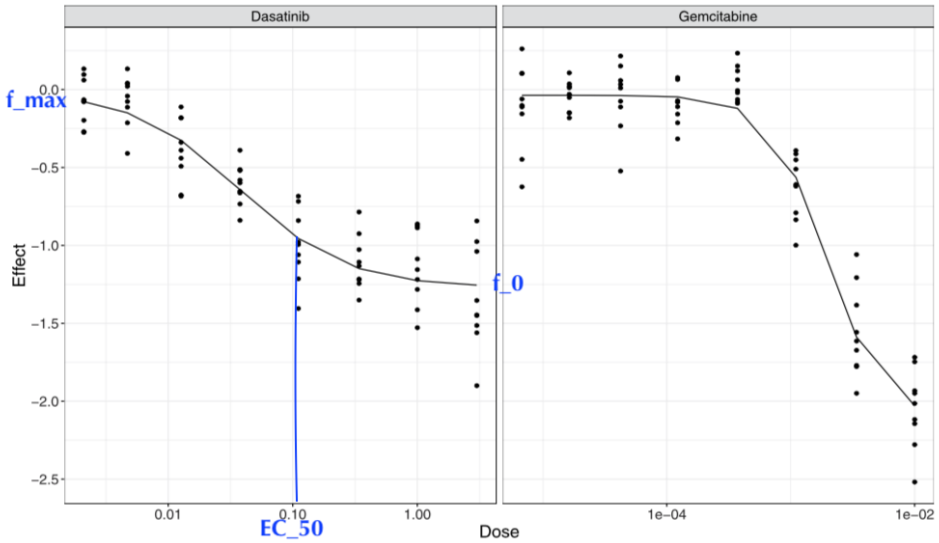
These null models

- give the expected outcome $f(d_1, d_2)$ under additivity when drugs 1 and 2 have doses d_1 and d_2
- depend on the monotherapeutic dose response curves, $f(d_1, 0)$ and $f(0, d_2)$.

Later in this talk, we discuss several null models.

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Monotherapeutic dose-response curves



Van der Borght et al. Scientific Reports (2017)

The expected outcome f at dose d is modelled by the Hill eq.,

$$f(d;\boldsymbol{\beta}) = f_0 + \frac{(f_{\text{max}} - f_0)\, d^h}{\text{EC}_{50}^h + d^h}$$

with $\boldsymbol{\beta}^t = (f_{\text{max}}, f_0, h, \text{EC}_{50})$.

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Parameters of the two dose-response models are estimated by means of least squares.

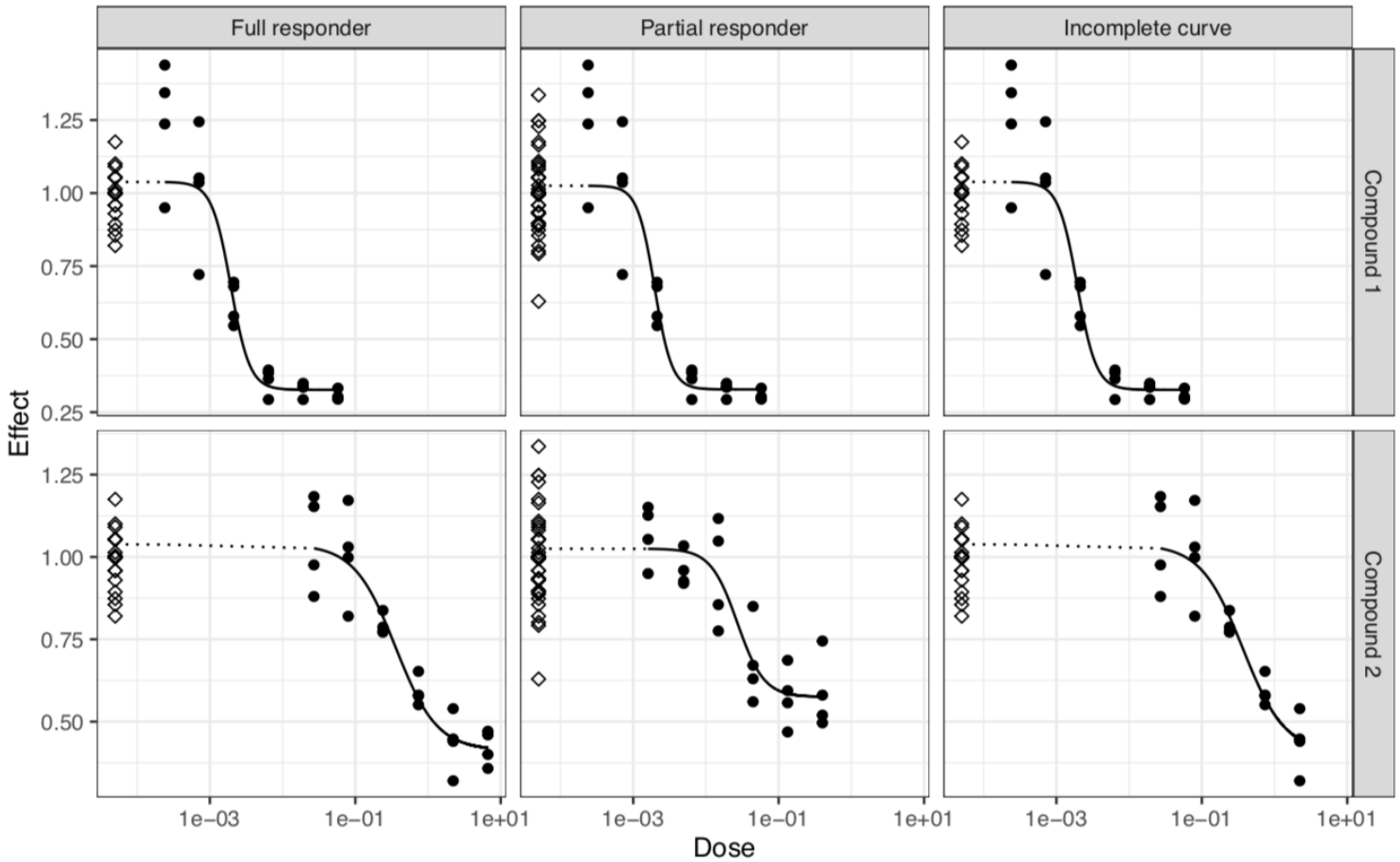
With Y_{ij} the observed outcome of drug $j = 1, 2$ with dose d_i, $i = 1, \ldots, n_j$,

$$Y_{ij} = f(d_i; \boldsymbol{\beta}_j) + \varepsilon_{ij}$$

with E $\{\varepsilon_{ij}\} = 0$ and Var $\{\varepsilon_{ij}\} = \sigma^2_{0j}$.

The least squares parameter estimates are denoted by $\hat{\boldsymbol{\beta}}_j$.

Some more dose-response terminology.



Thas et al. Pharmaceutical Statistics (2022)

Introduction

Later in this presentation:

1. some null models
2. statistical methods for detecting synergism
3. a comprehensive comparative simulation study

Bliss independence

$$\begin{aligned} f(d_1, d_2) &= f_1(d_1) + (1 - f_1(d_1))\, f_2(d_2) \\ &= f_1(d_1) + f_2(d_2) - f_1(d_1) f_2(d_2) \end{aligned}$$

- adopts probabilistic perspective
- equals sum of independent drug responses minus their joint effect under additivity (cfr. stochastic independence)
- assumes independent sites of action of the two compounds
- model can be adapted to allow for partial responders

Classical Loewe independence

$$\frac{d_1}{f_1^{-1}(f_{12}(d_1,d_2))} + \frac{d_2}{f_2^{-1}(f_{12}(d_1,d_2))} = 1$$

- assumes a constant potency ratio of the two compounds
- model can be adapted to allow for partial responders (based on the concept of occupancy)

Null Models

Highest Single Agent (HSA

$$f(d_1, d_2) = \max(f_1(d_1), f_2(d_2))$$

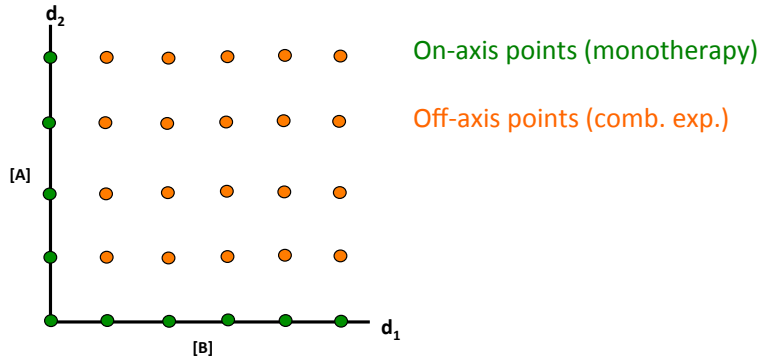
Loewe2: an alternative generalisation of Loewe

- combination of classical Loewe and HSA
- classical Loewe only works if $f_0 \leq f_{12} \leq \min(f_{\text{max},1}, f_{\text{max},2})$
- If e.g. $f_{12} > f_{\text{max},1}$ we now set $f_1^{-1}(f_{12}) = +\infty$.
  This results in $f_{12} = f_2(d_2)$, mimicking the HSA approach.

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Testing for Synergy

Recall that we have observations from several combination therapies (e.g. checkerboard design).



At each off-axis we have possibly replicated outcomes:

$$Y_{ik} \quad \text{at doses} \quad d_{1ik}, d_{2ik}$$

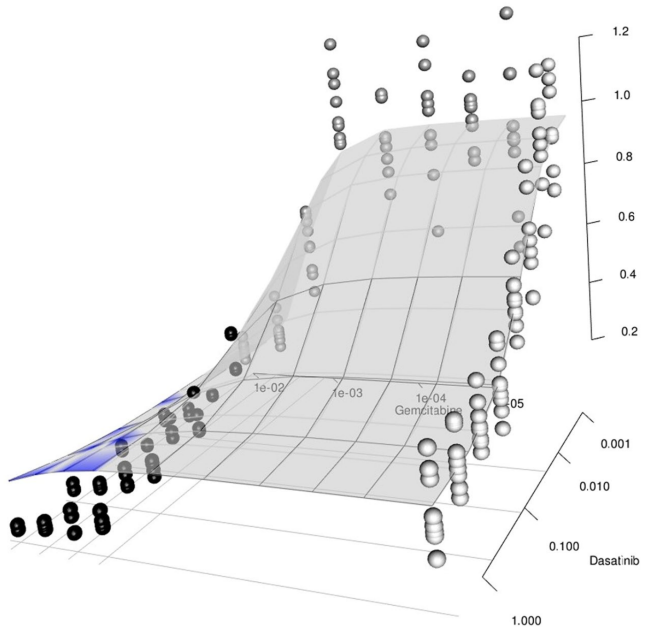
with

- $i = 1, \ldots, n_1$: off-axis point
- $k = 1, \ldots, m_i$: replicate

Testing for Synergy

The overall approach that we take is by looking at the mean residuals at the off-axis points,

$$E_i = \bar{Y}_i - f(d_{1i}, d_{2i}; \hat{\boldsymbol{\beta}}) = \bar{Y}_i - \hat{f}(d_{1i}, d_{2i})$$



Van der Borght et al. Scientific Reports (2017)

Testing for Synergy

We will present several versions of the MeanR and MaxR tests.

The variance-covariance matrix of the vector of the mean residuals, $\boldsymbol{E}^t = (E_1, \ldots, E_{n_1})$:

$$\boldsymbol{\Sigma}_E = \text{Var}\{\boldsymbol{E}\} = \sigma_0^2 \boldsymbol{C} + \sigma_1^2 \boldsymbol{D},$$

with

- σ_0^2 the residual variance on the on-axis points
- σ_1^2 the residual variance on the off-axis points
- $\boldsymbol{D}$ a diagonal matrix with on the diagonal $(1/m_1, \ldots, 1/m_{n_1}$
- $\boldsymbol{C}$ the variance-covariance matrix of the vector of predictions $(\hat{f}(d_{11}, d_{21}), \ldots, \hat{f}(d_{1n_1}, d_{2n_1}))$

Testing for Synergy

Suppose we have an estimator of $\boldsymbol{\Sigma}_E$, say $\hat{\boldsymbol{\Sigma}}_E$, then we construct two test statistics for testing the null hypothesis

$$H_0 : \mathsf{E}\left\{\bar{Y}_i\right\} = f(d_{1i}, d_{2i}) \quad \text{for all } i.$$

- MeanR:

$$T_{\text{MeanR}} = \boldsymbol{E}^t \hat{\boldsymbol{\Sigma}}_E^{-1} \boldsymbol{E}$$

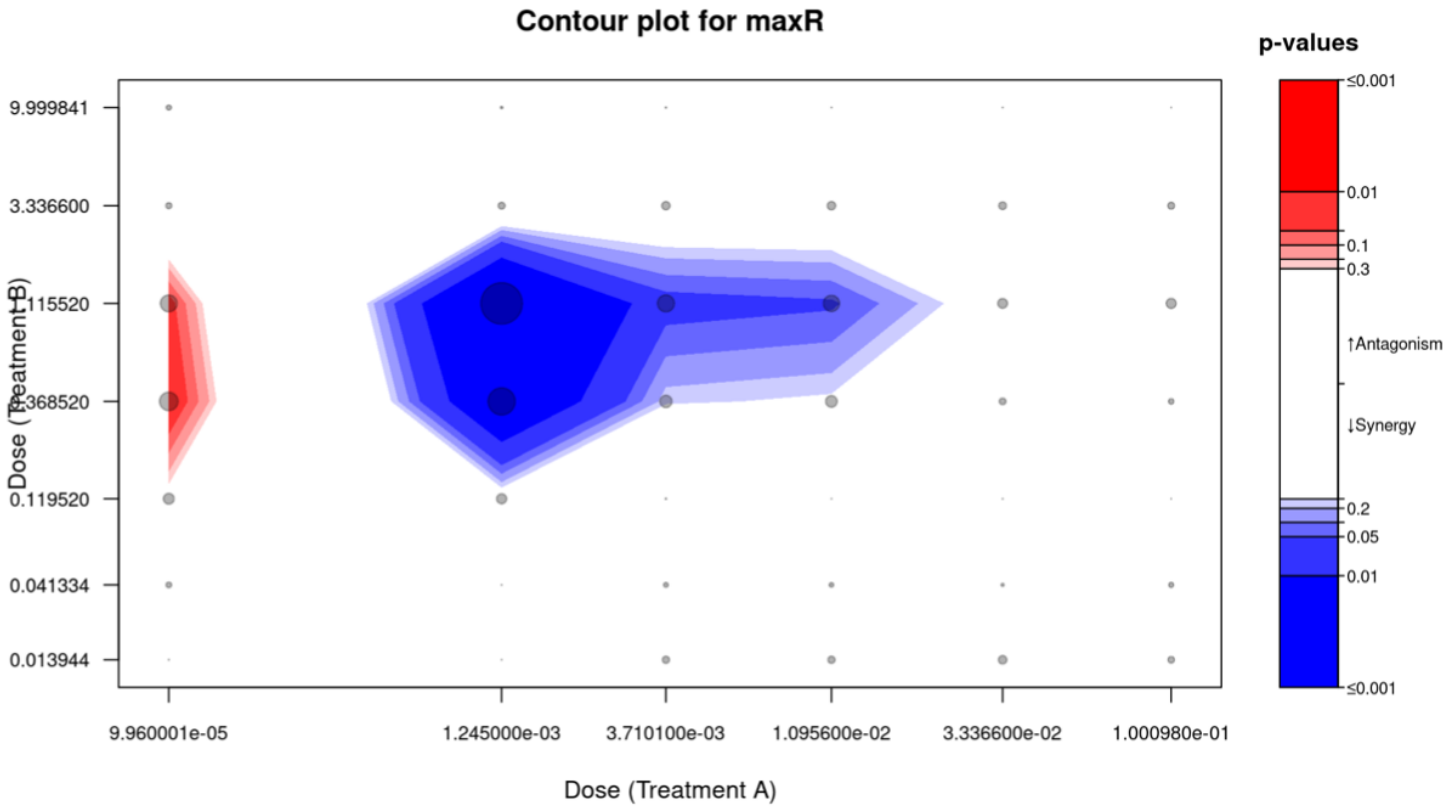
- MaxR:

$$T_{\text{MaxR}} = \max_i \left| \left(\boldsymbol{E}^t \hat{\boldsymbol{\Sigma}}_E^{-1/2} \right)_i \right|$$

The advantage of MaxR, is that at rejection of H_0, the individual $\left(\boldsymbol{E}^t \hat{\boldsymbol{\Sigma}}_E^{-1/2} \right)_i$ can used for identifying the off-axis points i at which there is a significant deviation from additivity.

Testing for Synergy

The advantage of MaxR, is that at rejection of H_0, the individual $\left(\boldsymbol{E}^t \hat{\boldsymbol{\Sigma}}_E^{-1/2}\right)_i$ can used for identifying the off-axis points i at which there is a significant deviation from additivity. (union-intersection test)



Testing for Synergy

The estimation of $\mathbf{\Sigma}_E$ depends on the assumptions we are willing to make:

- equal variance on all on-axis and off-axis points:
  $\sigma_0^2 = \sigma_1^2$
- equal variance on all on-axis and off-axis points, separately:
  $\sigma_0^2 \neq \sigma_1^2$
- mean-variance relation for the off-axis points:
  $\sigma_1^2(d_1, d_2) = \beta_0 + \beta_1 f(d_1, d_2)$

Equal variance on all on-axis and off-axis points:

$$\hat{\sigma}_0^2 = \hat{\sigma}_1^2 = \text{MSE}_0 \text{ from fits of Hill equations}$$

This may not be optimal, as the off-axis point observation are not used for the estimation.

The initial motivation was that the MeanR test statistic has a conventient F null distribution under the normality assumption,

$$T_{\text{MeanR}} \overset{H_0}{\sim} F_{n_1,q},$$

with q the degrees of freedom of MSE_0.

However, we recommend a bootstrap procedure for enumerating the null distributions.

Equal variance on all on-axis and off-axis points, separately:

$$\hat{\sigma}_0^2 = \text{MSE}_0 \text{ from fits of Hill equations}$$

and

$$\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} S_i^2$$

with S_i^2 the sample variance of the n_i replicates Y_{ik} at off-axis point i.

The null distributions are obtained by a bootstrap procedure, consiting of seperately resampling residuals for the on and off axis points.

Mean-variance relation for the off-axis points:

The parameters of the variance model

$$\sigma_1^2(d_{1i}, d_{2i}) = \text{Var}\{E_i\} = \beta_0 + \beta_1 f(d_{1i}, d_{2i})$$

are obtained by regressing S_i^2 on $\hat{f}(d_{1i}, d_{2i})$ (least squares).

With $\hat{\sigma}_1^2(d_{1i}, d_{2i})$ the predictions from the model fit, we construct

$$\boldsymbol{V} = \text{diag}(\hat{\sigma}_1^2(d_{11}, d_{21}), \ldots, \hat{\sigma}_1^2(d_{1n_1}, d_{2n_1}))$$

and use

$$\hat{\boldsymbol{\Sigma}}_E = \hat{\sigma}_0^2 \boldsymbol{C} + \boldsymbol{V},$$

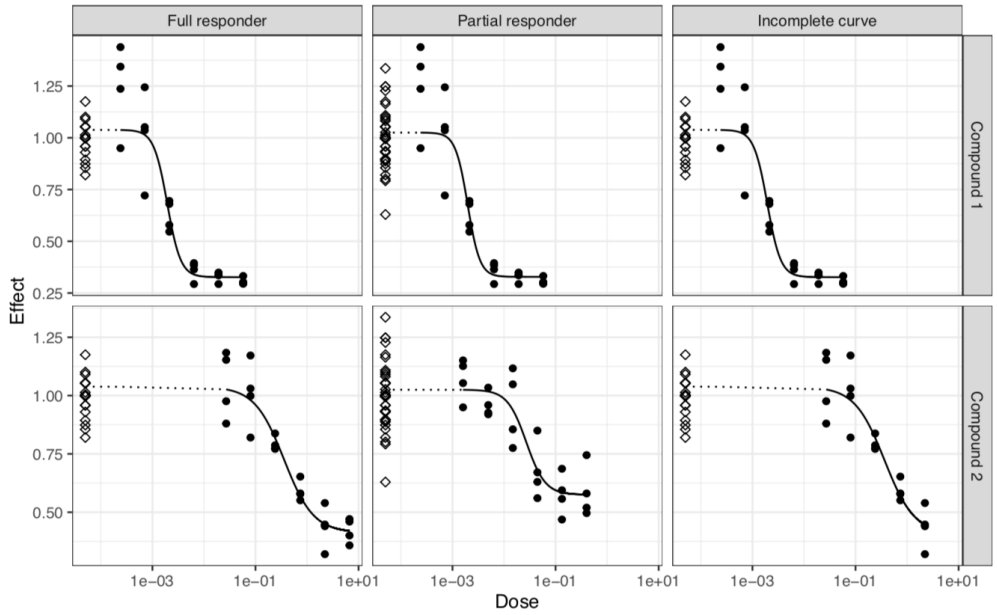
The null distributions are obtained by a bootstrap procedure, consiting of separately resampling residuals for the on and off axis points.

Objective

- evaluate MeanR and MaxR tests in terms of
  - type I error rate and power (for MeanR)
  - FWER and sensitivity (for MaxR)
- under various scenarios: combinations of
  - null model for predicting additive effects
  - synergistic effects
- we hope to find a null model that gives good results under all scenarios (even when null model is misspecified)

Data simulation framework:

- Start from real datasets (checkerboard designs) for estimating the parameters of the two monotherapeutic dose-response curves:
  - template 1: two full responders
  - template 2: a full and a partial responder



Thas et al. Pharmaceutical Statistics (2022)

Simulation Study

- Simulate data from a Data Generating Model (DGM):
  - combination of two fitted monotherapeutic dose-response curves
  - estimate expected outcomes $f(d_{1i}, d_{2i})$ for on and off axis points i (checkerboard design), based on a null model
  - add synergistic effects at some of the off-axis points i (see later):

    $$Y_{ik}^{*} = \hat{f}(d_{1i}, d_{2i}) + \delta_{ik}$$

    with

    $$\delta_{ik} \sim N(\Delta_i, \sigma_1^2(d_{1i}, d_{2i}))$$

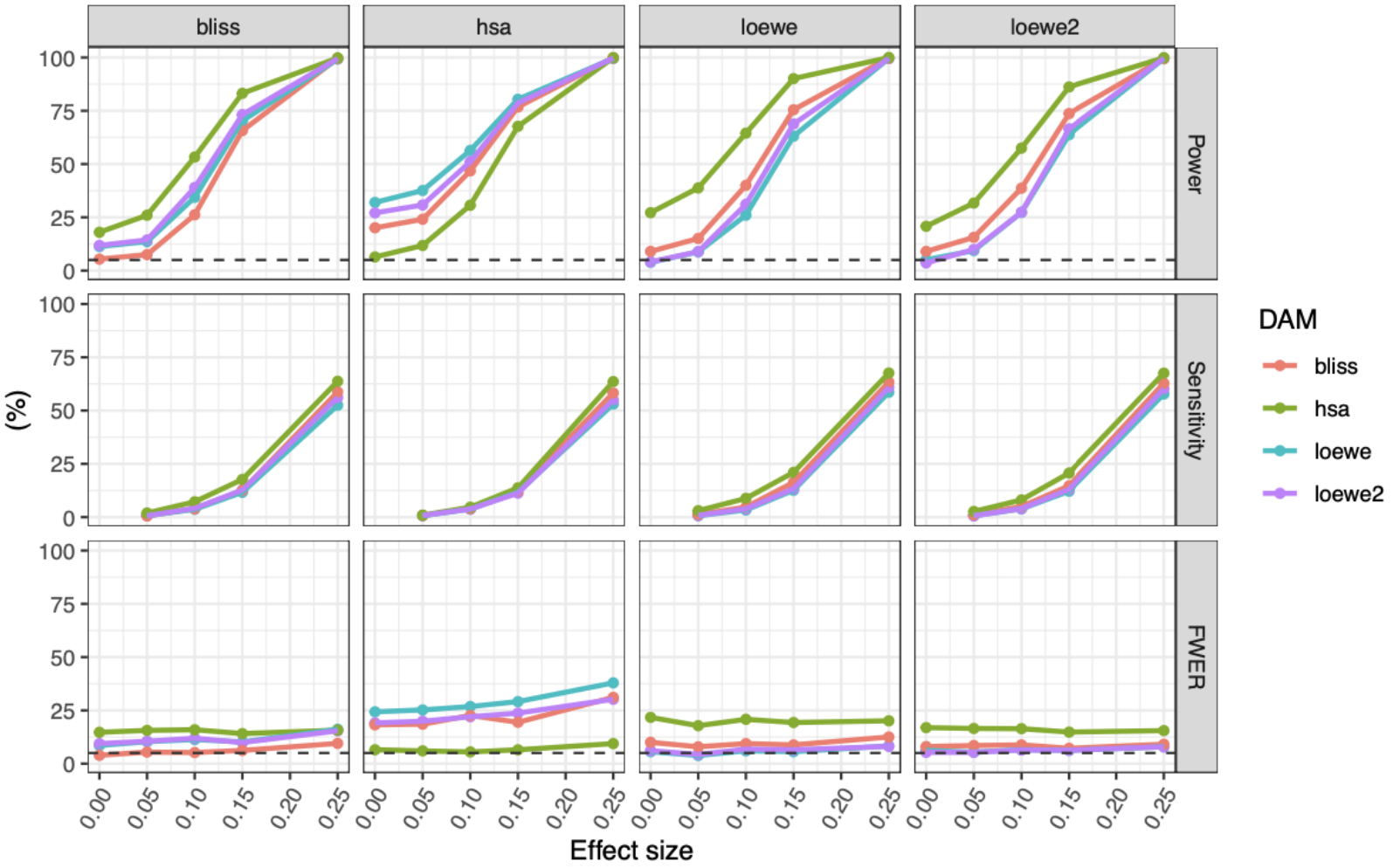
    and Δ_i the synergistic effect size at point i.
- Perform MeanR and MaxR at the 5% significance level according to different null models (Data Analysis Models (DAM)).
- compute FWER, power (of MeanR) and sensitivity (of MaxR)

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Overview of the 7 scenarios

Scenario	Monotherapies	Variances	Outliers	Synergy
1	Complete	$\sigma_0^2 = \sigma_1^2$	None	Random
2	Complete	$\sigma_0^2 \neq \sigma_1^2$	None	Area
3	Complete	$\sigma_0^2 = \sigma_1^2$	2 off axis points	Area
4	Complete	$\sigma_0^2 = \sigma_1^2$	1 batch	Area
5	Partial responder	$\sigma_0^2 = \sigma_1^2$	None	Area
6	Incomplete	$\sigma_0^2 = \sigma_1^2$	None	Area
7	Complete	$\sigma_0^2 \neq \sigma_1^2 = v^{-1}(\beta_0 + \beta_1 \mu)$	None	Area

Results for Scenario 1



Simulation Study

All results: see online suppl. mat. of Thas et al. (2022)

Statistical Detection of Synergy: Simulation Study Annelies Tourny and Stijn Hawinkel

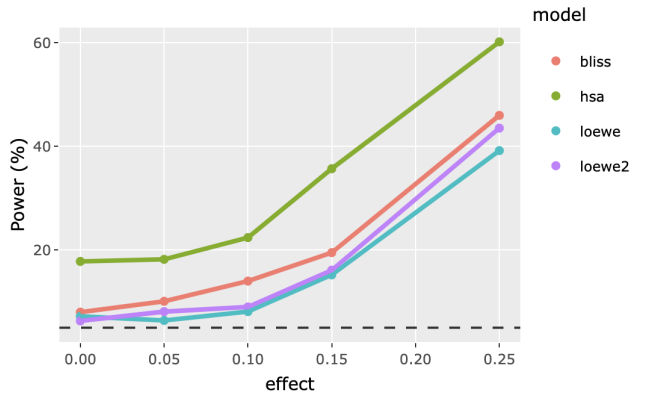
SIM 1 | SIM 2 | SIM 3 | SIM 4 | SIM 5 | SIM 6 | SIM 7 | EFFECT SIZES

Simulation specification:

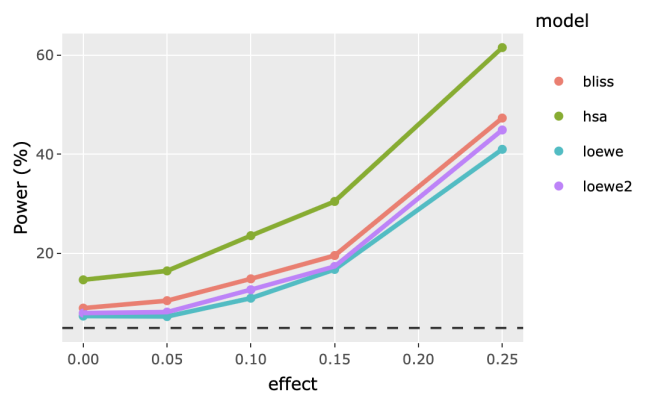
- More or less equal responders
- effects are added in an area resulting in 3-5 synergistic points.
- 2 clear outliers are added randomly
- Range of effect sizes: 0, 0.05, 0.1, 0.15, 0.25
- Constant variability

Power of the meanR test

Data generating model: Loewe



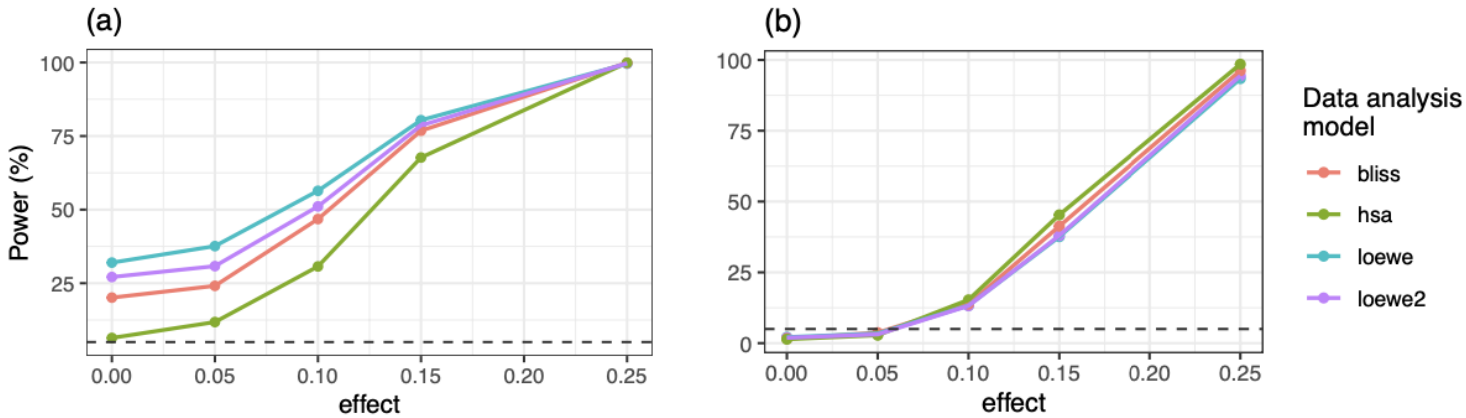
Data generating model: Loewe2



Simulation Study

Some conlcusions:

- meanR and maxR control type I error rate and FWER when DAM=DGM
- when DAM≠DGM, tests are often slightly liberal
  - DAM=HSA: tests are more liberal under model misspecification
  - DGM=HSA: too many antogonism calls when DAM≠HSA



Left: all calls - Right: only synergy calls

Simulation Study

- sensitivity of maxR is not much affected by choice of DAM, but for DAM=HSA, the test becomes liberal
- outliers are detrimental (too many false calls)
- Partial or incomplete dose-response curves:
  - no serious effect on type I error rate , FWER, power or sensitivity
  - incomplete curves can result in imprecise parameter estimates of dose-response curves
- Deviation from constant variance:
  - New methods perform better than the original methods
  - New bootstrap method helps, but still no good FWER control for MaxR

Simulation Study

Some recommendations:

- Report outliers
- Perform visual assessment of fit of monotherapeutic dose-response curves
- Model variance heterogeneity (only if necessary), and use the new bootstrap procedure

Effect Size

Hypothesis tests may have some drawbacks:

- they are often interpreted as a binary decision
- they do not provide information on the relevance of the synergistic effect
- their results may strongly depend on the sample size

It is good statistical practice to also report effect size estimates.

We propose two effect sizes:

- pointwise: $\delta_i = \mathsf{E}\{Y_{ik}\} - f(d_{1i}, d_{2i})$
- average: $\Delta = \frac{1}{n_1} \sum_{i=1}^{n_1} \delta_i$

Effect Size

The effect sizes can be estimated:

- $\hat{\delta}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} \left(Y_{ik} - \hat{f}(d_{1i}, d_{2i}) \right)$
- $\hat{\Delta} = \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{\delta}_i$

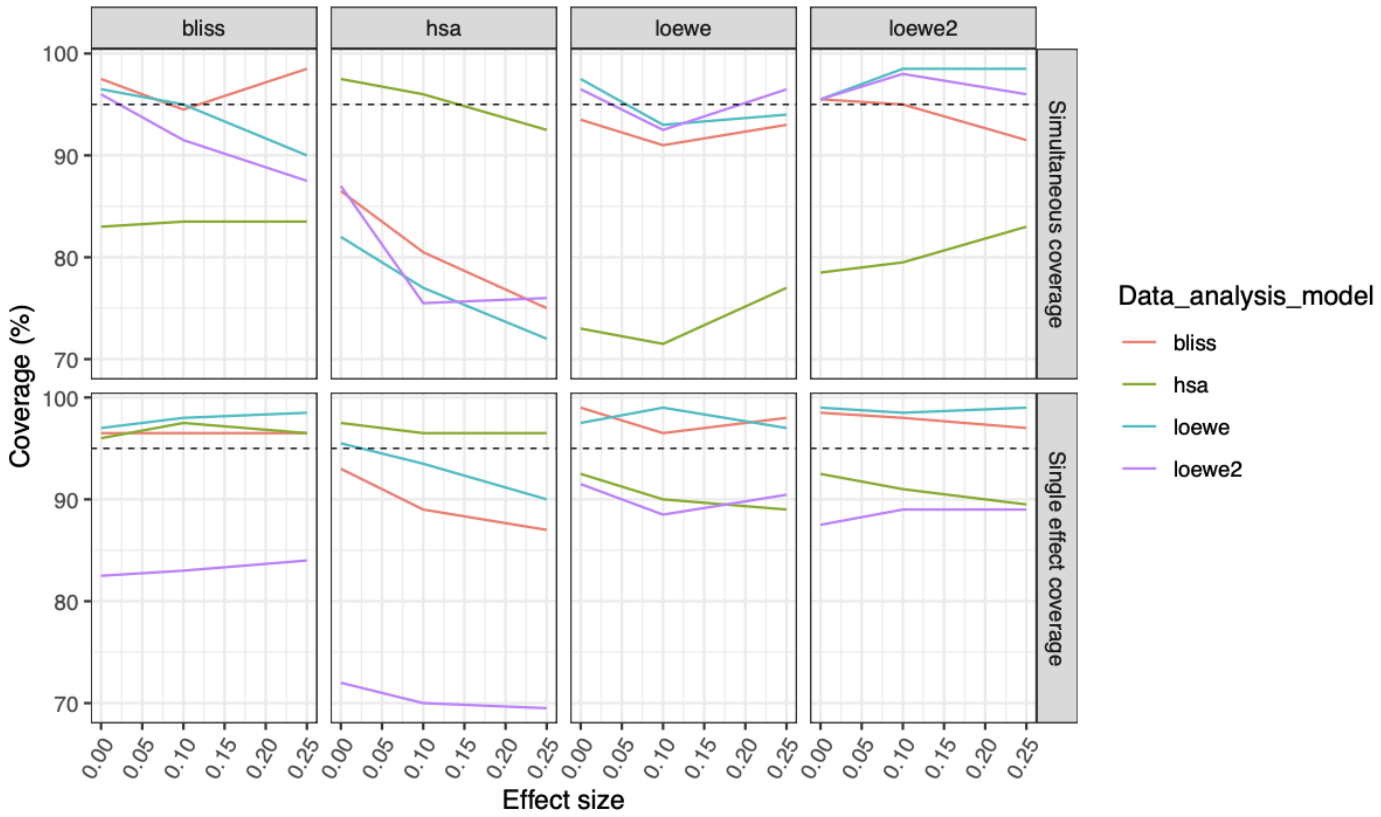
We have developed bootstrap procedures for the construction of confidence intervals.

This research is still ongoing.

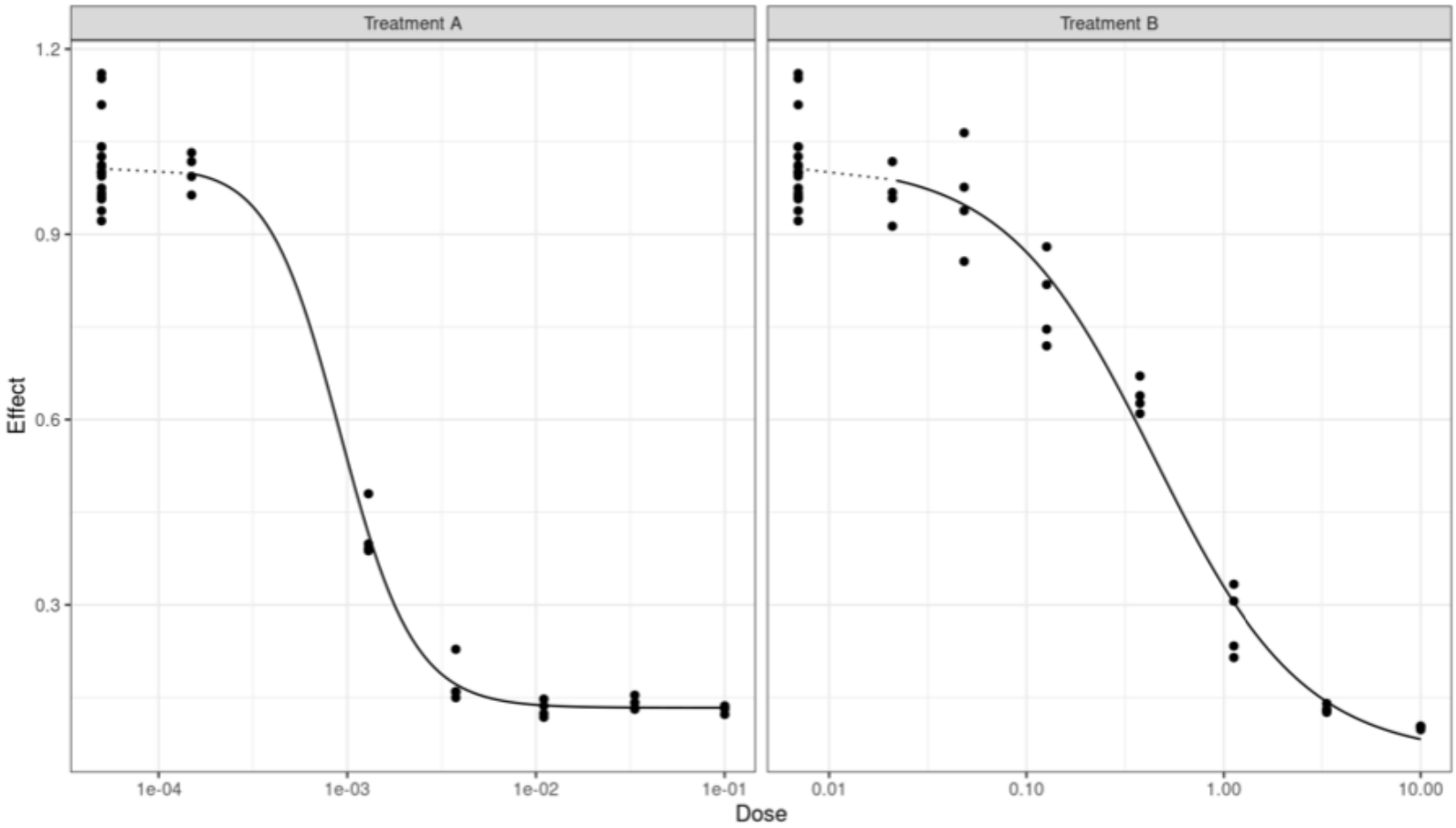
Simulation Study: effect size

The coverage of the effect size confidence intervals (nominal coverage of 95%) have been (are being) evaluated under the same scenarios as the tests.

Scenario 1

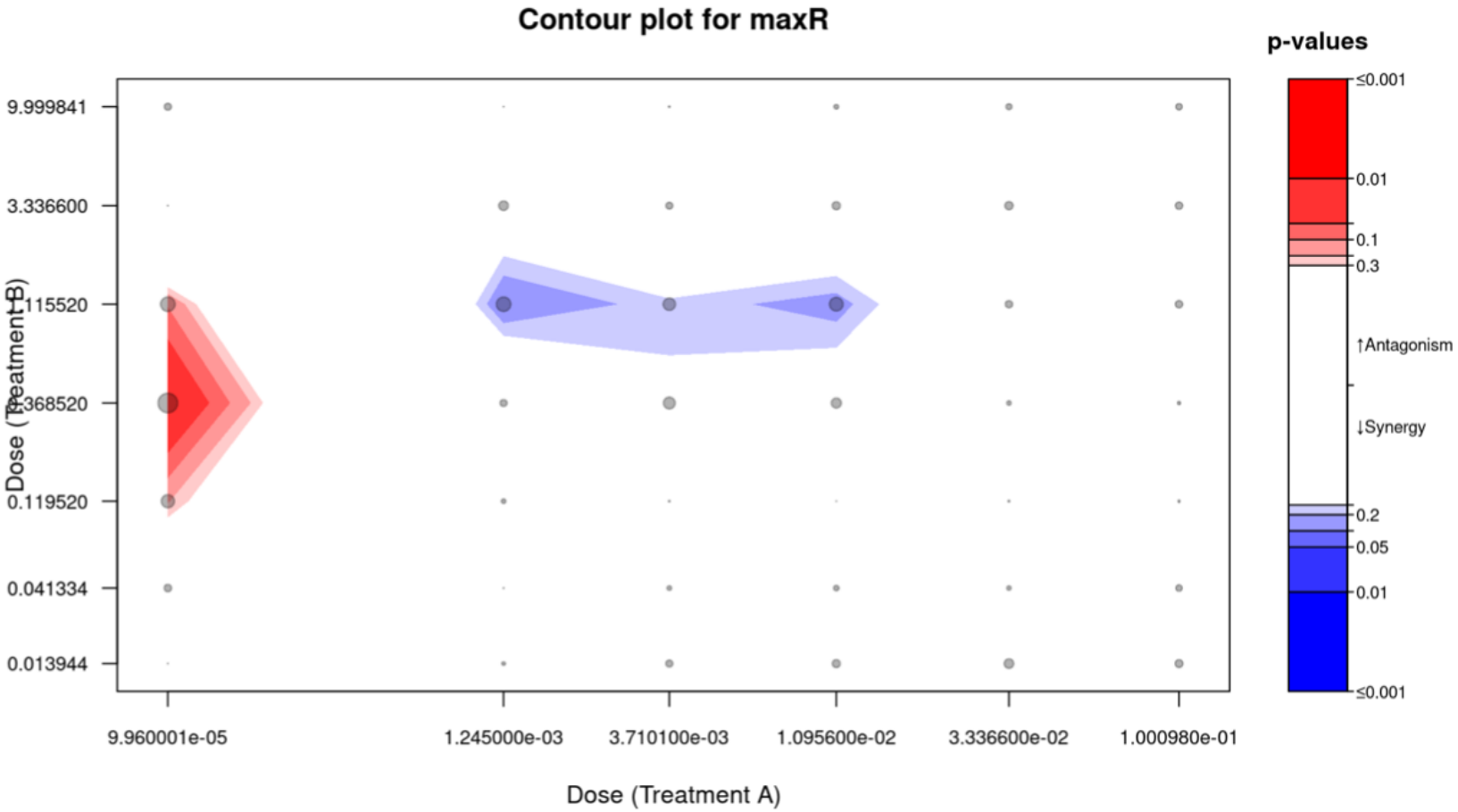


Example



Example

Results for DAM=Loewe



Example

Results for DAM=Loewe

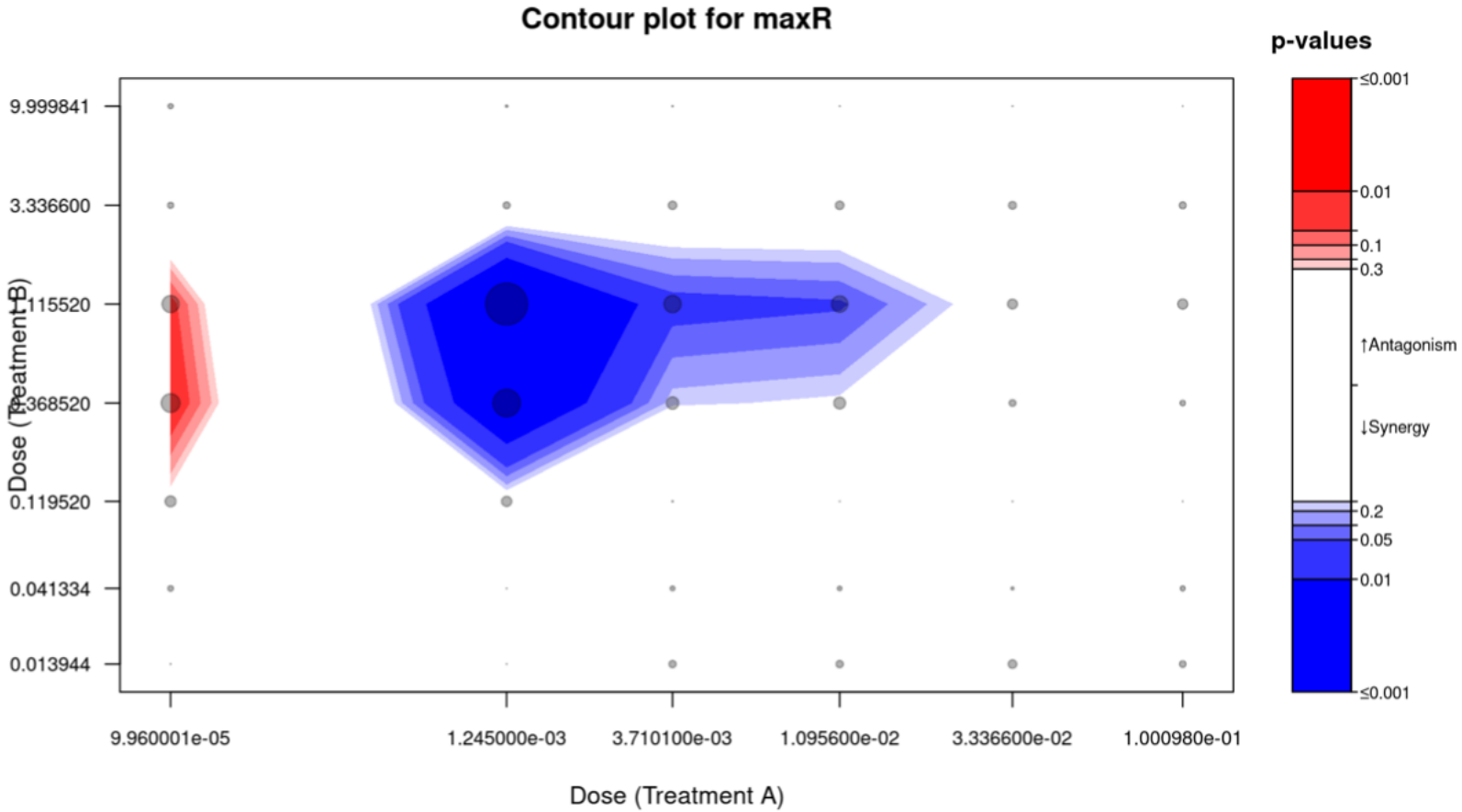
d2 \ d1	0.0000996	0.0012450	0.0037101	0.0109560	0.0333660	0.1000980
9.99984	0.02 (-0.04, 0.07)	0.00 (-0.05, 0.05)	-0.01 (-0.07, 0.04)	-0.03 (-0.08, 0.01)	-0.04 (-0.08, 0.00)	-0.04 (-0.08, 0.00)
3.33660	-0.01 (-0.06, 0.05)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, 0.00)
1.11552	0.12 (0.00, 0.23)	-0.09 (-0.15, -0.02)	-0.06 (-0.10, -0.02)	-0.05 (-0.09, -0.01)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, 0.00)
0.36852	0.26 (0.10, 0.42)	-0.08 (-0.18, 0.02)	-0.07 (-0.11, -0.02)	-0.05 (-0.08, -0.01)	-0.04 (-0.08, 0.00)	-0.04 (-0.08, 0.01)
0.11952	0.21 (0.03, 0.38)	-0.07 (-0.19, 0.05)	-0.05 (-0.11, 0.02)	-0.03 (-0.07, 0.01)	-0.02 (-0.07, 0.02)	-0.02 (-0.07, 0.03)
0.04133	0.12 (-0.06, 0.30)	-0.03 (-0.15, 0.08)	-0.02 (-0.09, 0.05)	-0.01 (-0.06, 0.04)	-0.02 (-0.07, 0.03)	-0.01 (-0.06, 0.04)
0.01394	0.01 (-0.16, 0.19)	-0.04 (-0.16, 0.07)	-0.01 (-0.08, 0.06)	-0.00 (-0.06, 0.05)	0.01 (-0.05, 0.06)	-0.00 (-0.06, 0.05)

call: additive antagonism

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Example

Results for DAM=HSA



Example

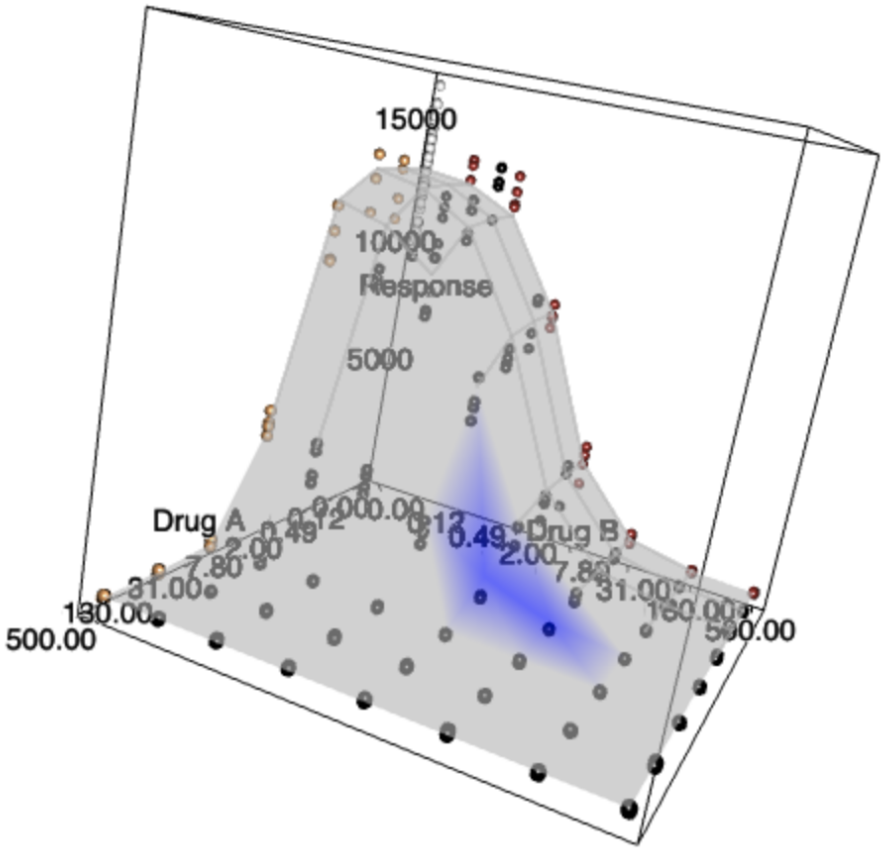
Results for DAM=HSA

d2 \ d1	0.0000996	0.0012450	0.0037101	0.0109560	0.0333660	0.1000980
9.99984	0.02 (-0.04, 0.07)	0.01 (-0.04, 0.07)	0.01 (-0.04, 0.06)	0.01 (-0.04, 0.06)	0.01 (-0.04, 0.06)	0.01 (-0.04, 0.06)
3.33660	-0.01 (-0.06, 0.05)	-0.05 (-0.09, -0.00)	-0.05 (-0.09, -0.01)	-0.05 (-0.08, -0.02)	-0.04 (-0.08, -0.01)	-0.04 (-0.08, -0.01)
1.11552	0.11 (-0.00, 0.22)	-0.18 (-0.24, -0.12)	-0.08 (-0.13, -0.02)	-0.05 (-0.09, -0.02)	-0.04 (-0.08, -0.00)	-0.04 (-0.08, -0.00)
0.36852	0.23 (0.07, 0.38)	-0.20 (-0.29, -0.12)	-0.07 (-0.13, -0.02)	-0.05 (-0.08, -0.01)	-0.04 (-0.08, 0.00)	-0.04 (-0.08, 0.01)
0.11952	0.17 (-0.00, 0.34)	-0.12 (-0.22, -0.02)	-0.05 (-0.11, 0.01)	-0.03 (-0.07, 0.01)	-0.02 (-0.07, 0.02)	-0.02 (-0.07, 0.02)
0.04133	0.09 (-0.08, 0.26)	-0.05 (-0.16, 0.05)	-0.02 (-0.09, 0.04)	-0.01 (-0.06, 0.03)	-0.02 (-0.06, 0.03)	-0.01 (-0.06, 0.04)
0.01394	-0.01 (-0.17, 0.16)	-0.05 (-0.16, 0.06)	-0.01 (-0.08, 0.06)	-0.00 (-0.05, 0.05)	0.01 (-0.05, 0.06)	-0.00 (-0.06, 0.05)

call: additive antagonism synergy

Example

Interactive 3D plots



Conclusions

- Generic: meanR and maxR can be used with all null models
  No need for a model for synergy (i.e. nonparametric)
- Control of false positive calls:
  - very good if DAM=DGM
  - slightly liberal, but acceptable, if DAM$\neq$DGM
  - if DAM=HSA, tests are too liberal; if DGM=HSA, too many antogonism calls
- Confidence intervals for effect sizes:
  Preliminary conclusion: good coverage control
- Check model fits: As for all statistical modelling: make plots to check model fits (dose-response curves, variance model) and data quality
- R package: BIGL on CRAN

References

This talk was based on the following references:

- Van der Borght, K., Tourny, A., Bagdziunas, R., Thas, O., Nazarov, M., Turner, H., ... and Ceulemans, H. (2017). BIGL: Biochemically Intuitive Generalized Loewe null model for prediction of the expected combined effect compatible with partial agonism and antagonism. *Scientific reports*, 7(1), 1-9.
- Thas, O., Tourny, A., Verbist, B., Hawinkel, S., Nazarov, M., Mutambanengwe, K. and Bijnens, L. (2022). Statistical detection of synergy: New methods and a comparative study. *Pharmaceutical Statistics*, 21(2), 345-360.