Safety assessment of GMOs – How to perform combined difference and equivalence testing?

EFSA Stakeholder workshop, 14 March 2019, Brussels

WAGENINGEN
UNIVERSITY & RESEARCH

Hilko van der Voet

Why equivalence testing?

Regulated products, examples:

- Pharmaceutical drugs
- Pesticides
- Genetically modified organisms (GMOs)

Safety assessment (ideal):

- Comparative approach: compare new product/application to established product/application(s) with a history of safe use
- Try to prove that there is no difference → SAFE!

But: using limited data and statistical methods, one never can prove the null hypothesis H_0: $\Delta = 0$

Moreover, statistical significance is not the same as relevance (EFSA 2017)

Therefore **equivalence testing** is needed: prove that differences are small enough considering uncertainties

Selected literature on GMO safety assessment 2010-2019

Food and Feed RA			ERA
Field study compositional data		Animal study	Field study
Traditional data	Omics data	Toxicological data	Abundance data
EFSA 2010a			**EFSA 2010b**
van der Voet 2011 **EFSA 2011a**		**EFSA 2011b**	
Ward 2012		Séralini 2012	
EU Impl. Reg. 2013		**EU Impl. Reg. 2013**	
Kang 2014	van Dijk 2014	**EFSA 2014**	Goedhart 2014
			van der Voet 2015
Vahl 2016		Schmidt 2016	
Vahl 2017		Schmidt 2017 van der Voet 2017 Hong 2017	
Paoletti 2018	Kok 2018, **EFSA 2018** Engel 2018	van der Voet 2018	
Jiang 2019 Engel *in prep.*		van der Voet 2019b Steinberg 2019	van der Voet 2019a van der Voet 2019b

GRACE project

G-TwYST project

AMIGA project

EFSA Guidance for GMO risk assessment 2010 - 2011

European Food Safety Authority
EFSA Journal 2010;8(11):1879

SCIENTIFIC OPINION

Guidance on the environmental risk assessment of genetically modified plants[1]

European Food Safety Authority
EFSA Journal 2011; 9(5):2150

SCIENTIFIC OPINION

Guidance for risk assessment of food and feed from genetically modified plants[1]

European Food Safety Authority
1 February 2010. Statistical considerations for GMOs safety

SCIENTIFIC OPINION

Scientific Opinion on

Statistical considerations for the safety evaluation of GMOs [1]

European Food Safety Authority
EFSA Journal 2011; 9(5):2149

SCIENTIFIC OPINION

Guidance on selection of comparators for the risk assessment of genetically modified plants and derived food and feed[1]

European Food Safety Authority
EFSA Journal 2011;9(12):2438

SCIENTIFIC OPINION

Guidance on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed [1]

EU law 2013

Official Journal of the European Union

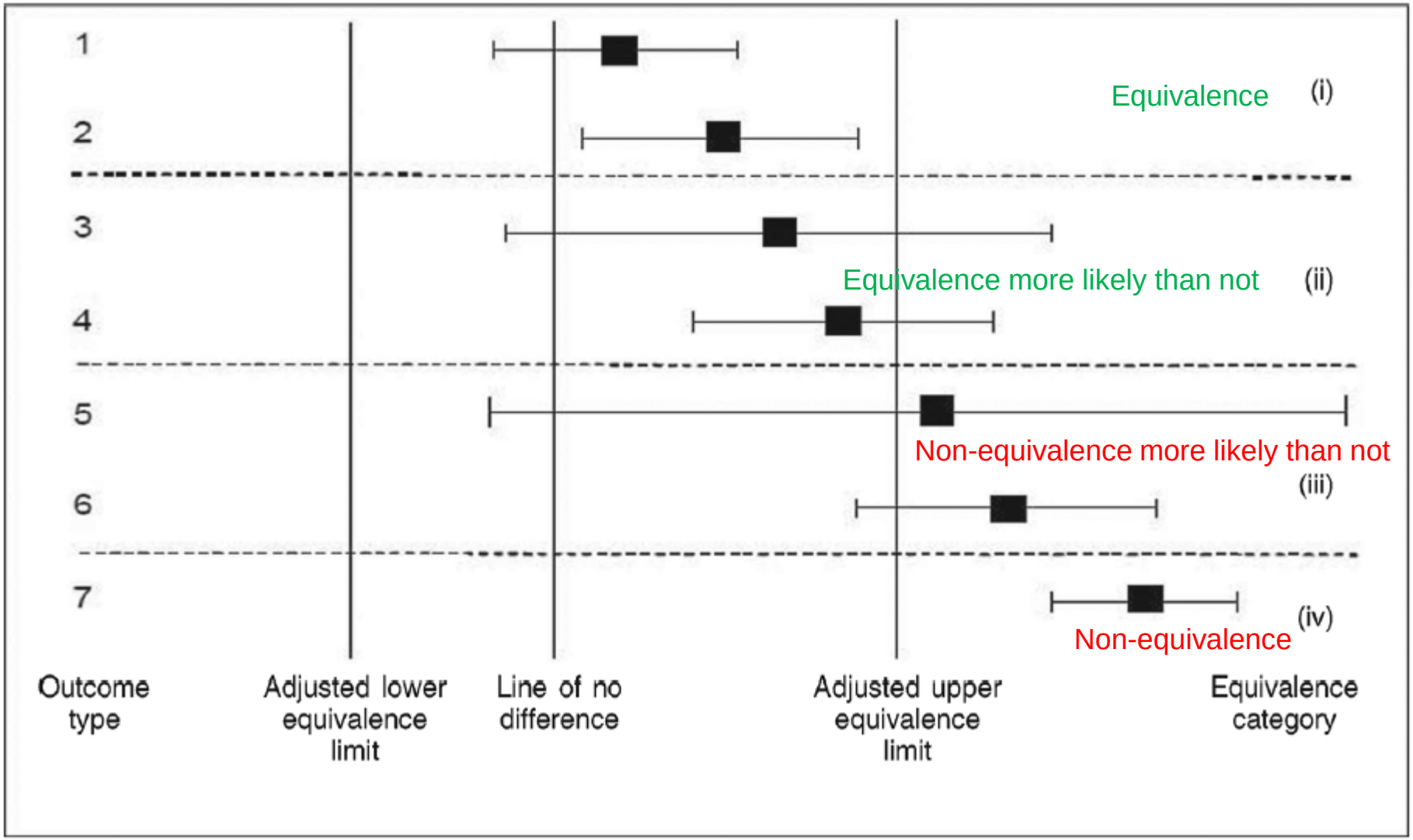
COMMISSION IMPLEMENTING REGULATION (EU) No 503/2013

of 3 April 2013

on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006

Equivalence testing introduced 2010-2013

Comparing GMO to COMP



COMMISSION IMPLEMENTING REGULATION (EU) No 503/2013

EFSA GMO Analysis software 2014

https://www.efsa.europa.eu/en/applications/gmo/tools

efsa
European Food Safety Authority

Home > Applications helpdesk > GMO > Tools

GMO applications: Tools

In the European Union, the use of genetically modified organisms (GMOs) is regulated through a legal framework. According to EU legislation (Regulation (EC) No 1829/2003, Regulation (EU) 503/2013 and Directive 2001/18/EC), GMOs can only be authorized for placing on the EU market following a scientific assessment of any risks that they may pose to human and animal health and the environment.

An integral part of such scientific assessment is the evaluation of field trials performed for the comparative assessment of the compositional, agronomic and phenotypic characteristics of GM plants. In 2009, EFSA's GMO Panel adopted an opinion on Statistical considerations for the safety evaluation of GMOs to provide detailed guidance on the performance of such field trials and their statistical analysis. The principles, concepts and data requirements presented in the above mentioned opinion are endorsed in the GMO Panel's Guidance for risk assessment of food and feed from genetically modified plants and the more recent implementing Regulation (EU) 503/2013.

- GMO Analysis software: installer (11.3 Mb)
- GMO Analysis software: user manual (0.9 Mb)

Mixed model

$$y_{ijkl} = m + e_i + r_{ij} + t_k + g_l + \varepsilon_{ijkl}$$

- *e*: environment (site) i (Note: site might be location x year)
- *r*: replicate (block) j
- *t*: treatment group k

  k=1: comparator; k=2: GMO; k=3: references
- *g*: genotype (variety, line) l

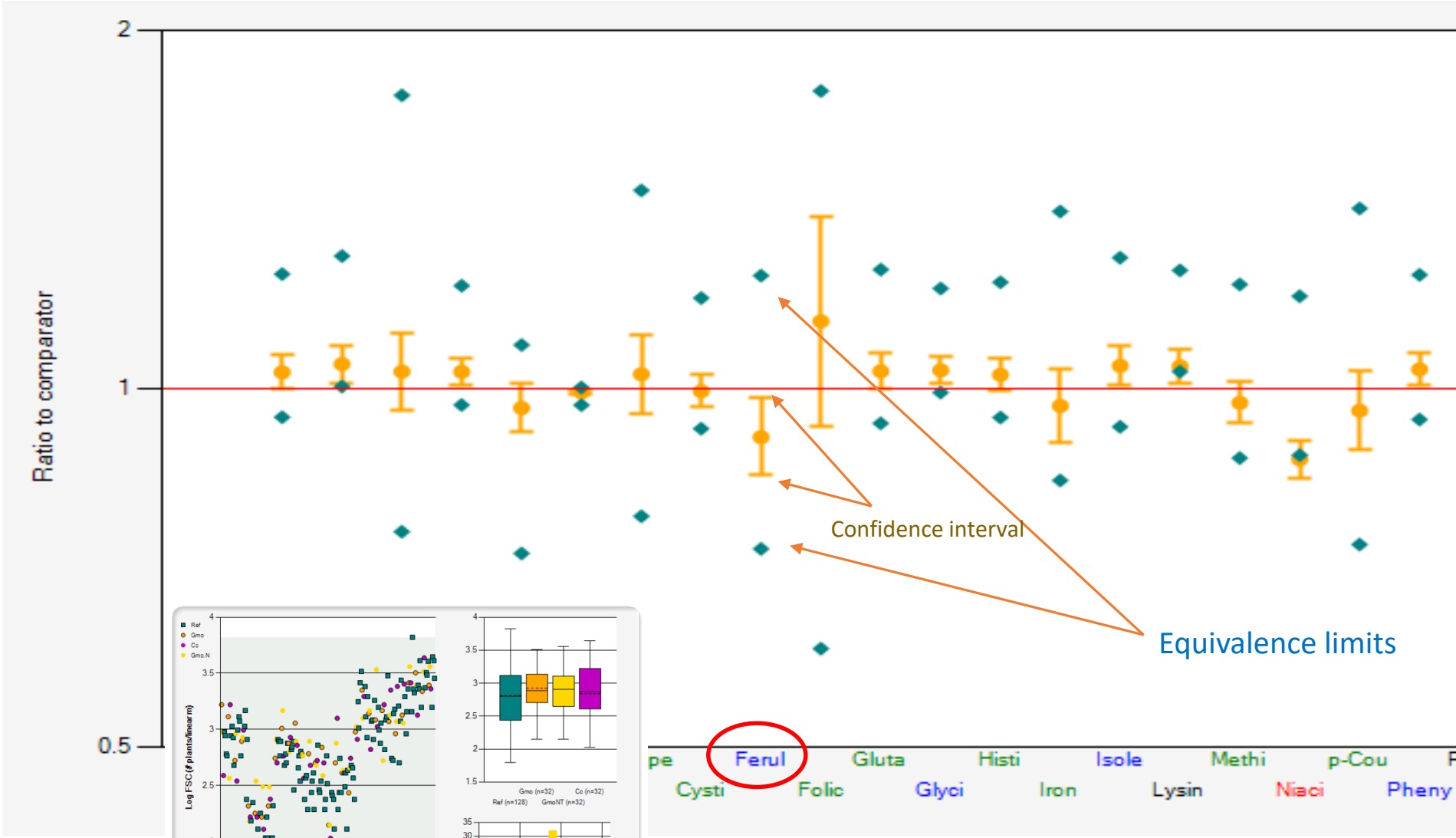
  l=1: comparator; l=2: GMO; l=3,...,K: references
- *m* and *t* fixed factors, others factors are random

  interest is in:

  $d_1 = t_2 - t_1$ (difference test)

  $d_2 = t_2 - t_3$ (equivalence evaluation)
- variance components V_e, V_r, V_g, V_0

Example output GMO Analysis (overview all endpoints)



Developments in test approaches

- Environmental risk assessment
  - How to set equivalence limits for nontarget taxa counts?
  - How to test in hierarchies?
- Food and feed risk assessment
  - Compositional analysis
    - How to optimise equivalence tests for field trials with references?
    - Should GxE interaction be addressed?
    - Can omics data be used?
  - Animal feeding studies
    - Are they needed? [→ Steinberg]
    - How to apply equivalence tests (few references in study)?
    - Is there a role for Standardised Effect Sizes?
- General
  - What are the right comparator(s)?
  - What sample sizes are needed? (power analysis)
  - How to address multiplicity (many tests at the same time)

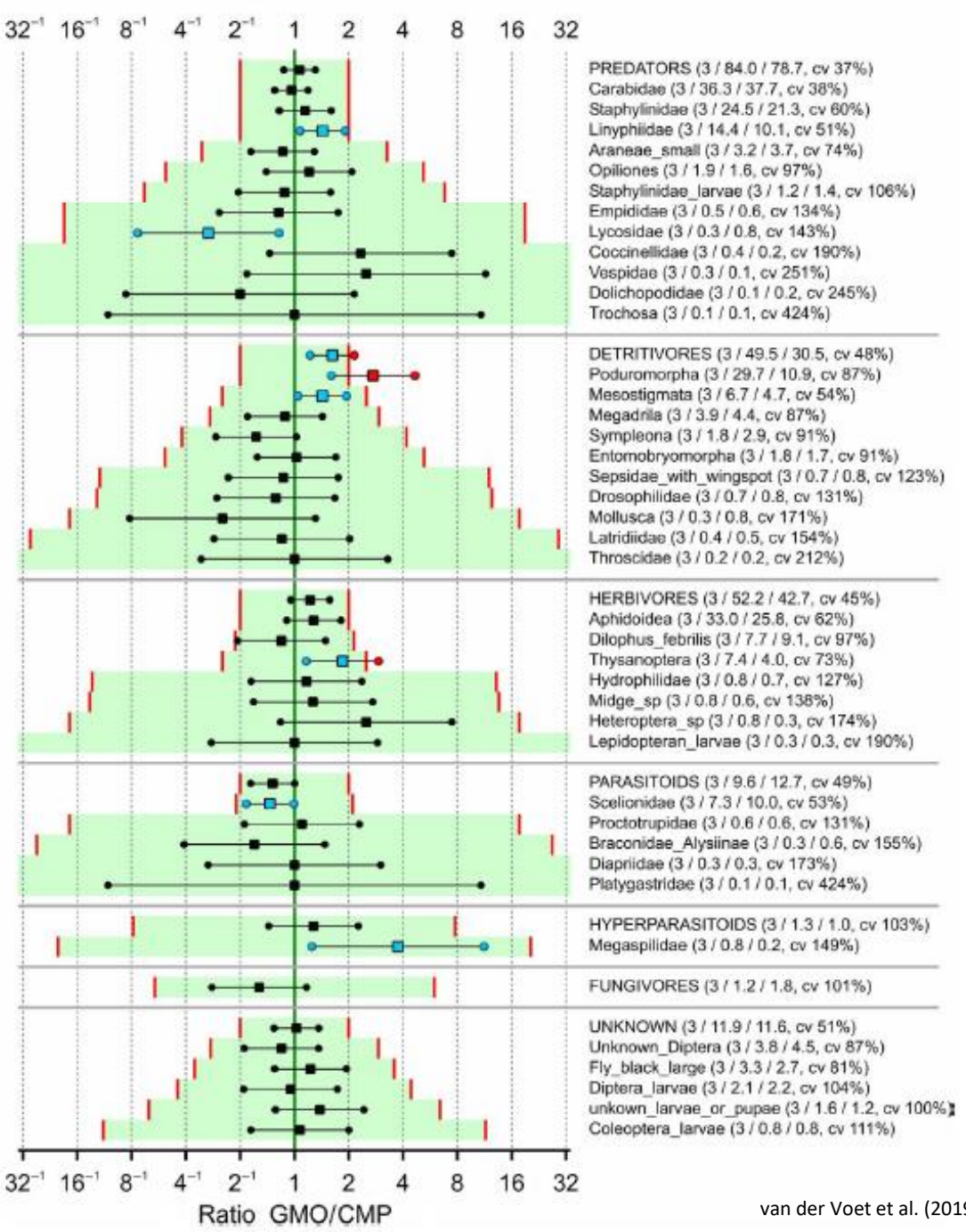
ERA example

Field trial Ireland 2013 (AMIGA project)

- Cisgenic late blight resistant potato line, *vs.*
- Conventional comparator
- Many nontarget organisms
- Often low counts

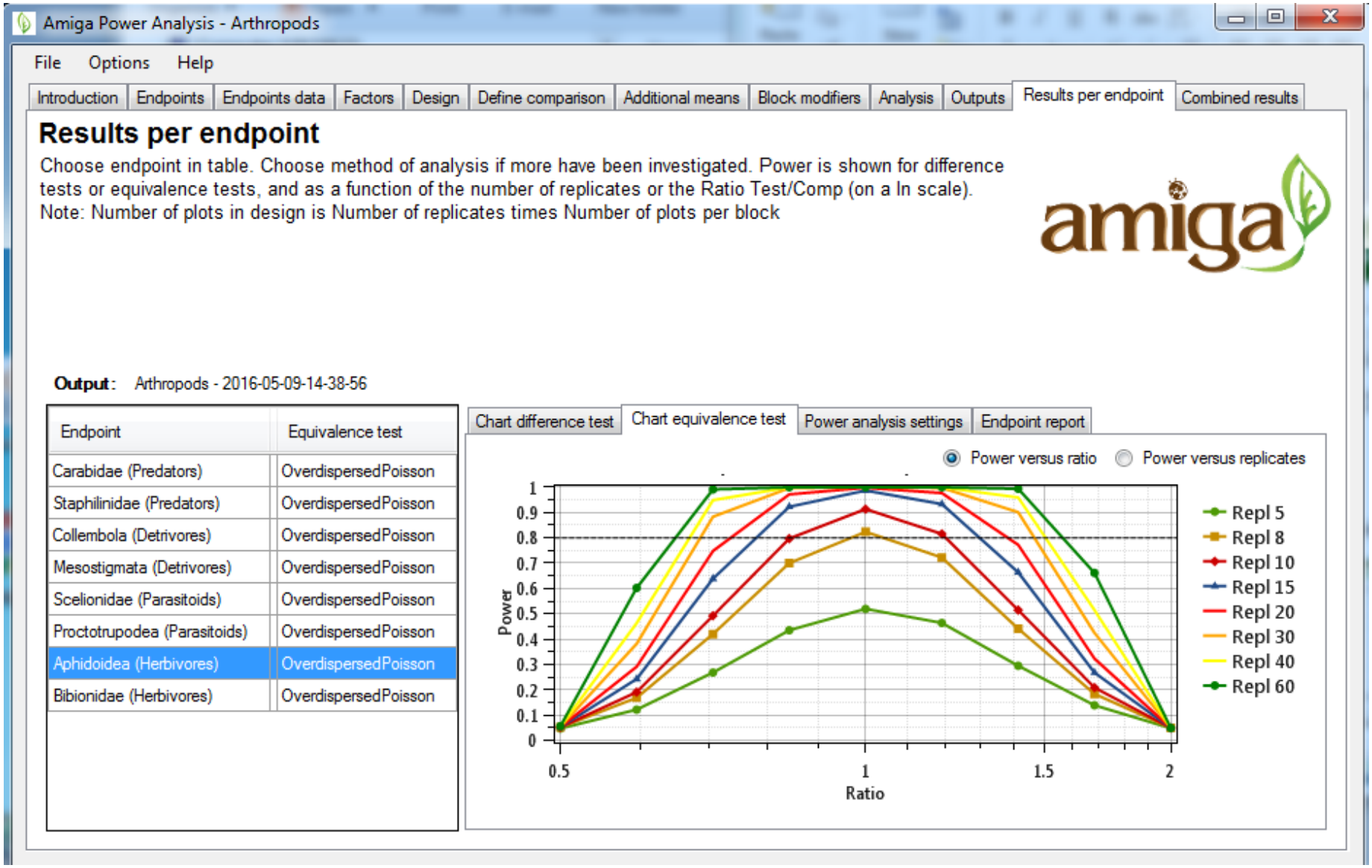
Statistical approach

- Equivalence limits (or Limits of Concern) set to two-fold for $n \geq 10$
- Wider for $n < 10$ based on Poisson distribution
- NTOs grouped by function
- Confidence intervals in graph with LoCs
- DIfferent ways to address the hierarchy of endpoints



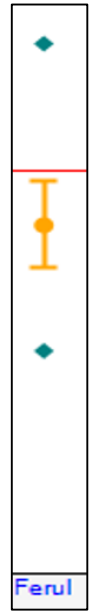
van der Voet et al. (2019)

How many plots? Power vs. ratio GMO / CMP



Kruisselbrink et al. (2016). Software for power analysis and analysis of data from field studies. Deliverable 9.5, AMIGA project. http://edepot.wur.nl/455506

Compositional and animal study data: Estimation of equivalence limits from reference varieties



Current algorithm EFSA is a 2-step approach:

1. Estimate equivalence limits (ELs)
2. Use equivalence limits for testing the GMO

Step 2 assumes ELs to be fixed, this is only an approximation

- **Field study compositional data:**
  All data come from one experiment. An improved statistical analysis can be performed by including the uncertainty of the ELs in the equivalence test in a 1-step approach (Kang & Vahl 2014, Vahl & Kang 2016, Engel et al., paper and R software *in prep.*)
- **Animal feeding study** (G-TwYST project):
  Method of Vahl & Kang has been adapted for animal studies with use of historical data for references (van der Voet et al. 2017, 2019, Steinberg et al. 2019)

References for safety

Group differences expressed as differences on a logarithmic scale (or equivalently as ratios on the original scale)

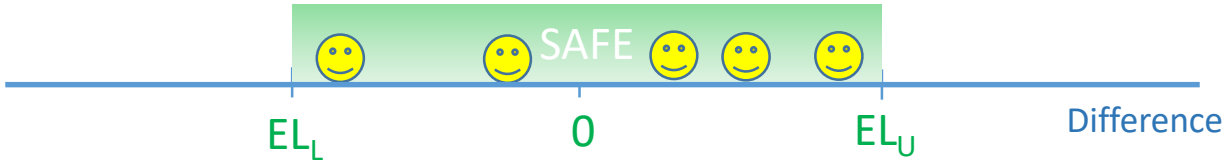
Often no generally accepted lower or upper toxicity limits (TL) are available. However, TLs are not needed to assess equivalence

Sometimes (unofficial) lower and/or upper equivalence limits (EL) are available, e.g. Hong et al. (2017) for 9 animal endpoints

Another approach is to estimate $[EL_L, EL_U]$ from non-GM (=safe) reference groups in historical data

What is a typical group difference for non-GM groups?

Compare to historical references



= difference between 2 reference groups (history of safe use)

We have used historical non-GM data as references

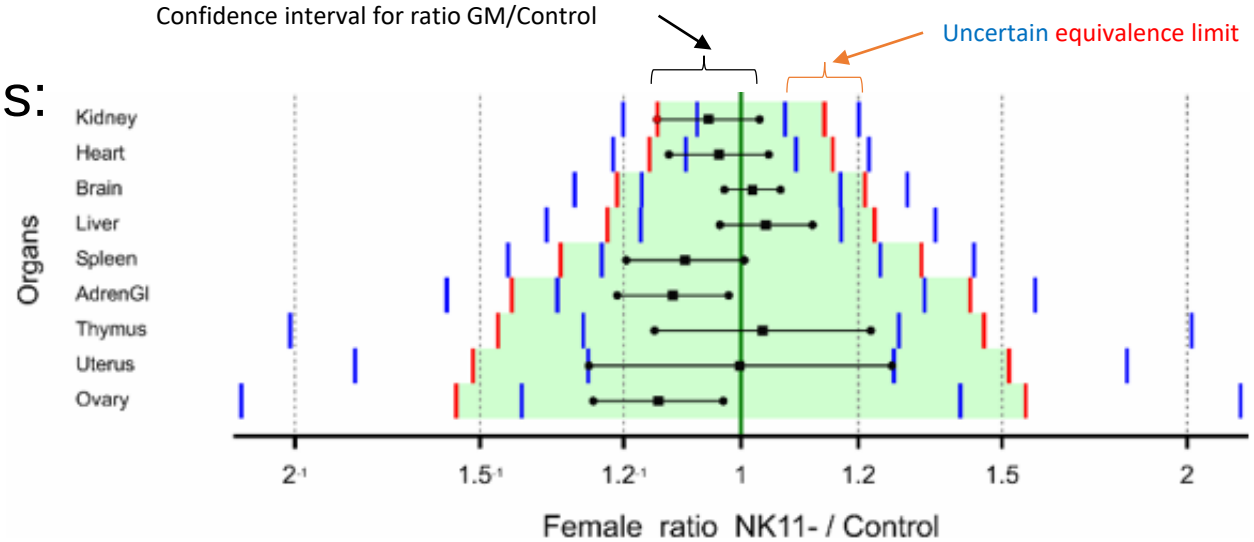
Limited historical data, therefore the estimates of EL will be uncertain

Limited current data on GM and Control feeds, therefore the estimates of current values are also uncertain, use confidence interval ⟷

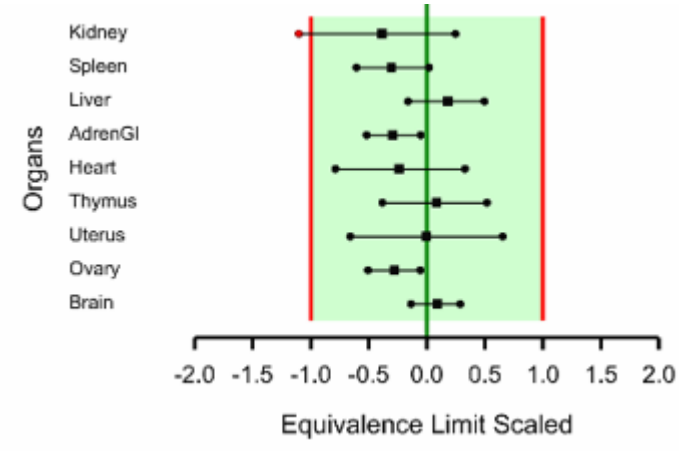
In the proposed method, the uncertainties are combined in a single measure, the Equivalence-Limit Scaled Difference (ELSD)

Example: rat feeding study, female organ weights

Group differences:



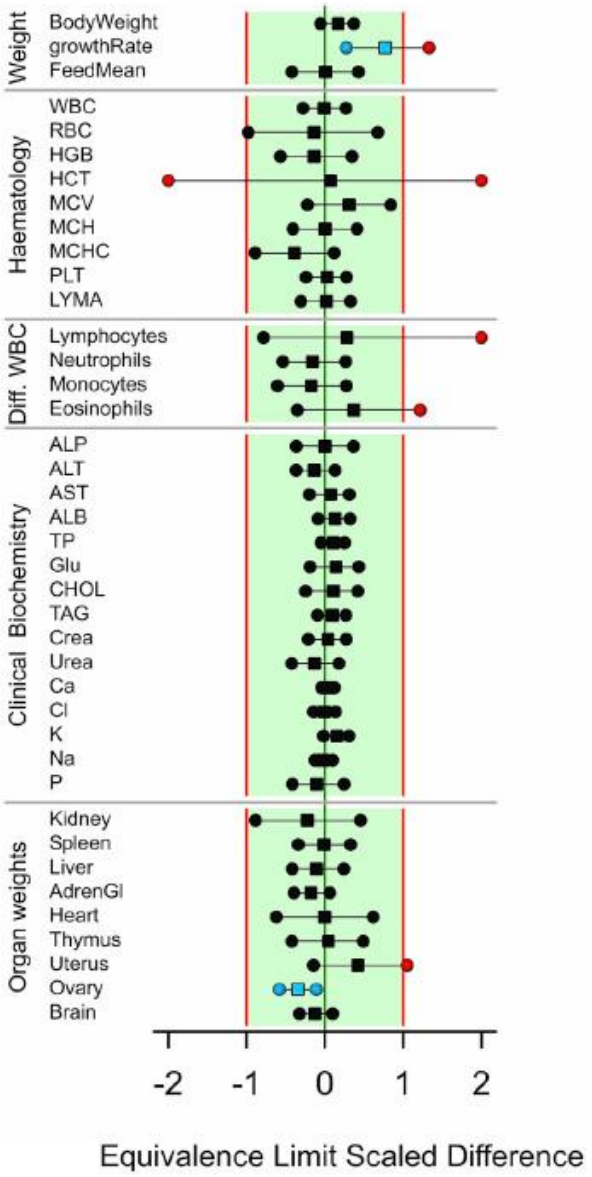
Scaled:



- Interval in [-1,+1] = ‘proof of equivalence’
- Central point in [-1,+1] = ‘Equivalence more likely than not’ (EFSA 2011)

Difference and equivalence tests for animal feeding study data, example

- 33% NK603 maize (sprayed with RU) vs. non-GM reference varieties, female rats, 90 days
- Equivalence limits are set so that a variety similar to the references would be judged equivalent with 95% power
- ELSD confidence intervals are scaled w.r.t. the equivalence limits, so [-1,+1] is the safe area.
- 2 statistically significant differences
- 35 endpoints are equivalent, 5 endpoints are equivalent more likely than not



van der Voet et al. (2019)

Multiplicity of tests

- Food composition: 40 - 80 endpoints
- Animal studies : 50 – 150 endpoints
- Nontarget organisms ERA: 10 - 100 taxa

With $\alpha = 0.05$ and no differences, expect 5 significant test results per 100 endpoints.

Probability of a false detection for *m* independent tests: $\mathrm{P(FP)} = 1 - (1 - \alpha)^m$

- e.g. for $\alpha = 0.05$ and $m = 50$, $\mathrm{P(FP)} = 1 - (1 - 0.05)^{50} = 0.92$

Correction for multiplicity

- Many methods exist for difference testing
- A popular method is False Discovery Rate (FDR), e.g. by adjustment of p values:

  FDR-adjusted p values are obtained by multiplication of the raw p values with factors between 1 and m, where m is the number of tested hypotheses. Let $p_1 \leq \ldots \leq p_m$ be the ordered p values for m endpoints. Then, the FDR adjusted p values according to a linear step-up algorithm are sequentially calculated as $\tilde{p}_{(m)} = p_{(m)}$; $\tilde{p}_j = \min(\tilde{p}_{(j+1)}, \left(\frac{m}{j}\right) p_j)$, for $j = m - 1, \ldots, 1$.

Multiplicity of tests

- Hong et al. (2017) analysed up to 146 endpoints and applied a multiplicity correction for difference testing (FDR-adjusted p values)
  - This controls the probability (rate) of finding a significant difference when in fact there is no difference (false discoveries)
  - However, this adjustment is completely inappropriate for equivalence testing
  - For equivalence testing, one would like to control the probability of concluding equivalence when in fact there is no equivalence
- FDR-adjusted p values are higher than unadjusted p values, and consequently the power of the tests is lower.
  - However, Hong et al. (2017) ignored FDR adjustment in their power analyses and present misleading conclusions

Use of omics?

Untargeted approach

→Obtain holistic / system-wide overview of biology of plant

→Monitoring for many unintended effect(s)

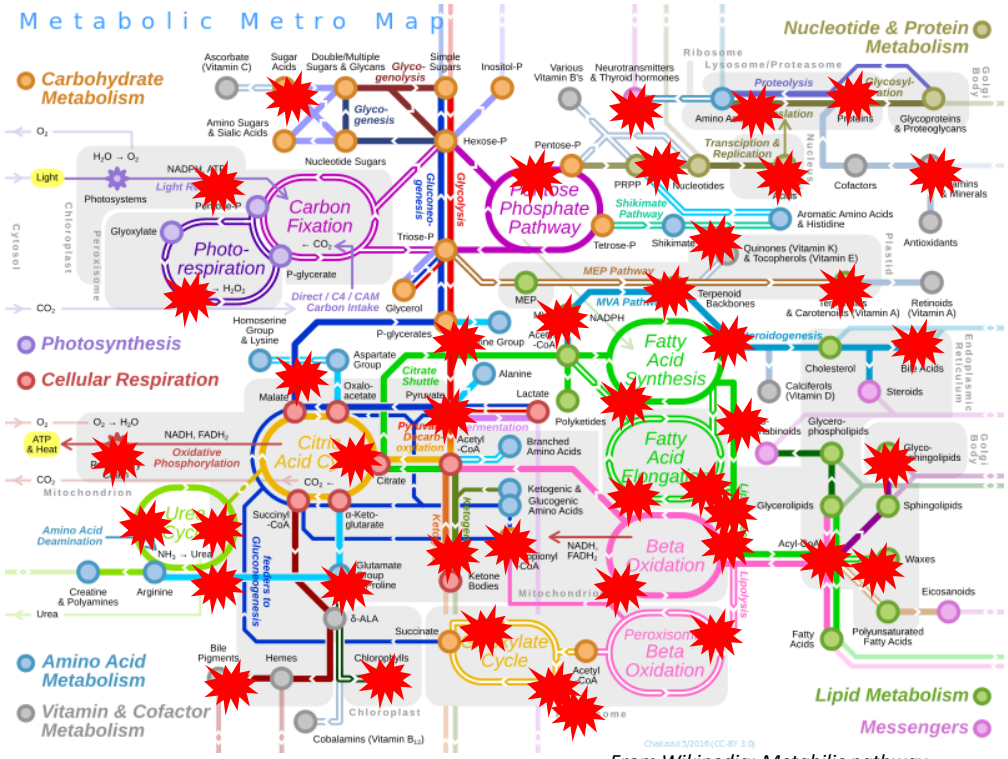
efsa
European Food Safety Authority

EVENT REPORT

APPROVED: 13 November 2018

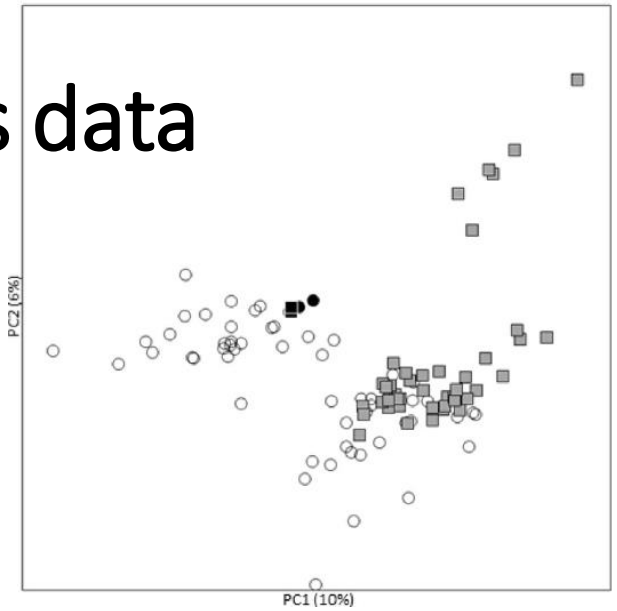
doi:10.2903/sp.efsa.2018.EN-1512

EFSA Scientific Colloquium 24 – 'omics in risk assessment: state of the art and next steps



From Wikipedia: Metabilic pathway

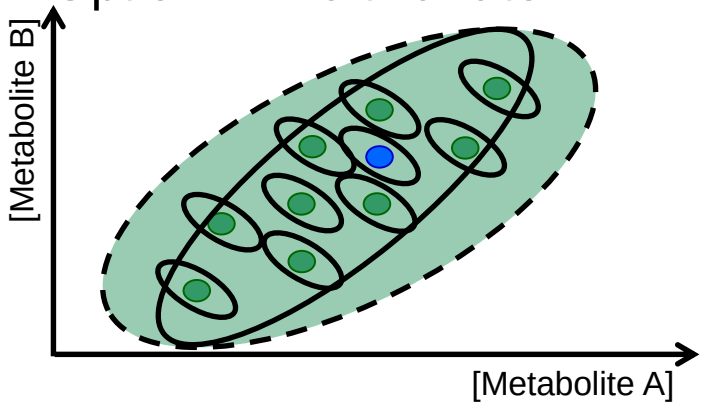
Safety assessment with omics data



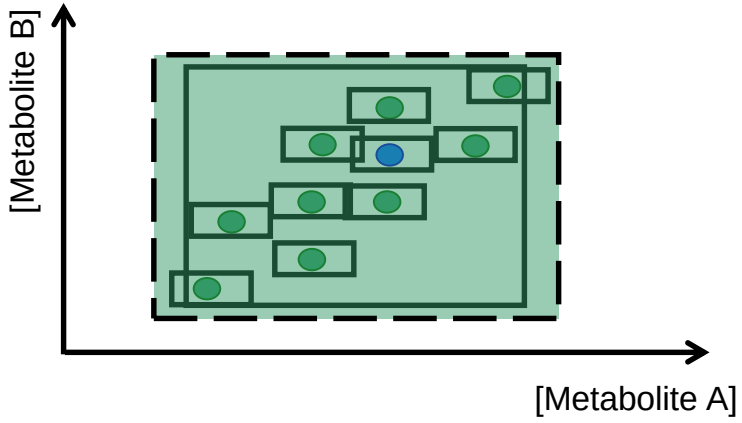
- SIMCA one-class model was proposed by van Dijk et al. (2014)
- Used in GRACE project (Corrujo et al. 2018 for maize, Kok et al. 2018 for potato)

- Evaluation and comparison with new proposals (Engel & van der Voet 2018, and ongoing work)
  - Adapt for equivalence testing framework
  - Compare to alternative models

Option 1: multivariate



Option 2: univariate



Choice of comparators

Definition: *similar food or feed produced without the help of genetic modification and for which there is a well-established history of safe use*

Commission Implementing Regulation (EU) 503/2013:

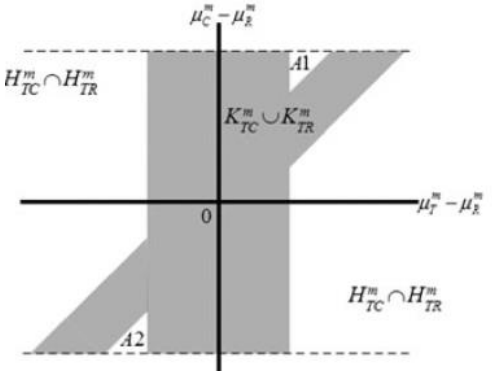
- For the difference test, the component of the fixed factor of interest is the single degree-of-freedom contrast between the GM plant and its conventional counterpart.
- For the equivalence test, the component of the fixed factor of interest is the single degree-of-freedom contrast between the GM plant and the set of non-GM reference varieties.

Criticism by Jiang et al. (2019):

- “the result of EFSA equivalence testing often has little or nothing to do with the GM trait effect, which should be the sole focus of the comparative assessment”

Can we devise equivalence testing approaches that consider `both conventional counterpart and references?

- e.g. a proposal was made by Vahl & Kang (2017):



GxE interactions

	E_1	E_2	E_3	…	E_n
GMO	10	10	**100**	10	10
Control	10	10	10	10	10
Ref_1	10	10	10	10	10
Ref_2	10	**100**	10	10	10
…	10	10	10	10	10
Ref_k	10	10	10	10	10

Risk assessment can adjust the models for interactions

However, the real questions are for risk management:

- What are possible outcomes of risk assessment?
  - Should it be possible to declare a GMO equivalent in some environments but not in others?
  - If so, how to define the environments: case-by-case? as a random factor while using further modelling, e.g. soil and meteo descriptors?

More interaction RA/RM needed to address interactions

Conclusions / Points for discussion

- New scientific methods for GMO risk assessment have become available since 2010/2011

- Some methods are ready for practical application
  - Replace the 2-step approach for equivalence testing by a 1-step approach
  - Implement Equivalence Limit Scaled Differences as the preferred statistic (suitable for all areas)

- Some subjects need further work but are important for the future
  - Handling of multiplicity
  - Handling of GxE interaction
  - Use of omics data