

# OVERVIEW OF INTEGRATED TEST STRATEGIES METHODS IN APPLICATION TO IN VITRO DATA— INDUSTRY EXPERIENCE

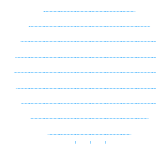
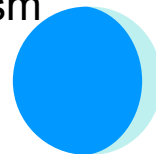
*Joanna Jaworska*

Central Product Safety - Procter & Gamble Eurocor  
Strombeek - Bever, Belgium

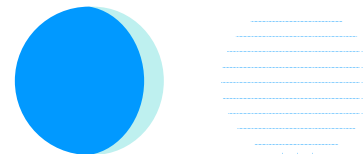


# Introduction

- Emerging need in chemical management is Development of Integrated Testing Strategies (ITS)
  - Gain more comprehensive basis to make a decision
  - Better reuse existing data to make ITS resources efficient
- New type and larger amounts of data are changing paradigms of data evaluation
  - ongoing refinement of in vitro tests - experimental design (i.e. concentration at which chemicals are tested),
  - Many new in vitro tests that provide new insights on mechanistic basis of the endpoint evaluated
  - How to integrate information from several tests each addressing a different mechanism
  - Increasing reliance on vitro data requires development of methods for data integration
  - Methods for in vitro data integration need to be fitting to generalize integration with other data, Adme, other biologically related endpoints, Mechanism

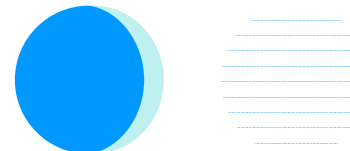


# Part 1. Towards data integration strategy – understanding what we have and where we are going to



# Strategy of data integration from in vitro data in application to assess in vivo endpoint

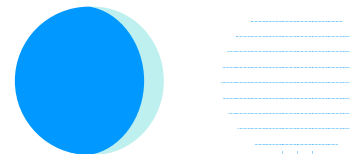
- Replacement of in vivo tests by a single in vitro study is not realistic given the complexity of mechanisms involved in the in vivo test.
- Combinations of several in vitro tests, covering relevant mechanistic steps if possible and organized in a logical, hypothesis driven decision scheme are needed to make efficient use of generated data.
- The development of new in vitro tests addressing improvements in understanding of the toxicity mechanisms are subject of active research.
- The field that needs equal attention and further development are methodologies to integrate multiendpoint data to provide:
  - transparent,
  - structured,
  - consistent
- And hypothesis driven interpretation to support a decision (OECD Workshop on Integrated Testing Approaches, 2007).



# How to meet criteria to be consistent, transparent, structured ? –

## We need a formal framework

- Evolution of the framework of narrative to qualitative to quantitative
  - To increase rational and coherent interpretation
  - To quantify uncertainty
    - can assess value of information of individual tests and batteries,
    - eventually guide testing –avoid duplication of information, stop testing when desired/or maximum possible reliability/ uncertainty reduction is achieved.
- Provide objective ways to deal with complex and conflicting information
  - Mix of categorical and continuous data
- Methods that allow to combine history, expert opinion, experiments and model results and reflect hierarchy of the testing strategy
  - Initial hypothesis is revised based on the new evidence to generate updated hypothesis



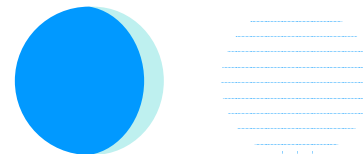
# Possible integration approaches

- Frameworks for data integration that meet criteria put forward by the OECD are different flavours of quantitative weighing schemes.
- Scoring - Among them scoring schemes are the easiest to apply and many were developed (Calabrese et al. 2007). They are very useful for relative ranking. The weights can be
  - Heuristic ( test 1 – 5, test 2 – 3 etc.)
  - Developed by datamining and fitting to a particular model structure ( eg. Linear model – the decision of the model is again heuristic)
- Weighing schemes can also be probabilistic.



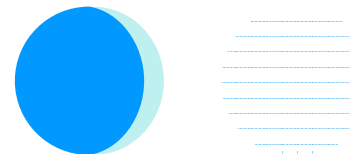
# Test battery

- The goal is to combine results of individual tests to achieve greater predictivity than predictivity of individual tests and therefore increase confidence in overall assessments.
- This presentation discusses Bayesian interpretation of tests battery results
  - probabilistic reasoning



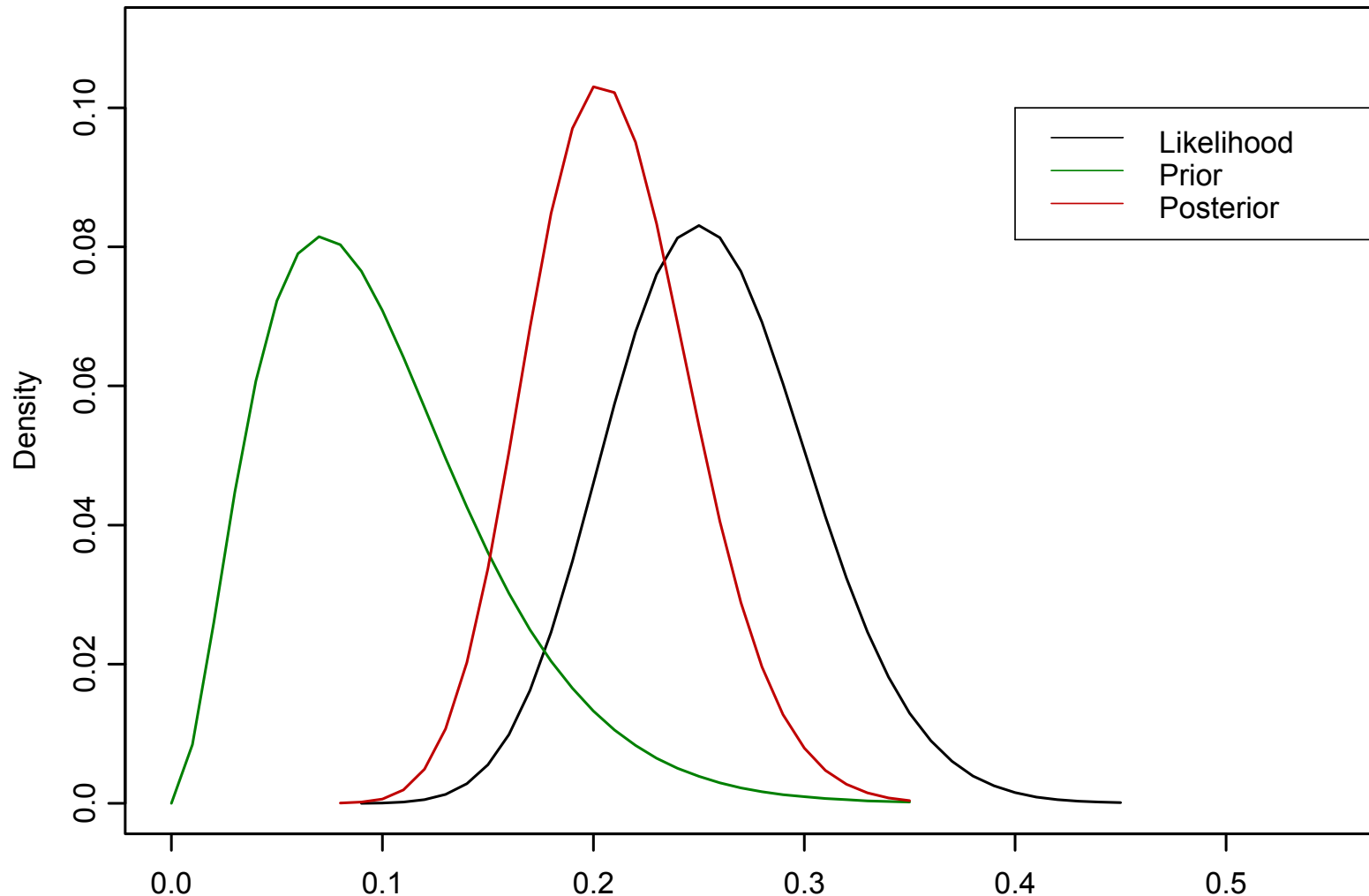
# Bayesian interpretation of test battery

- Formal logical tool to combine complex information into one framework by probabilistic reasoning  
formally generates one result: probability that a chemical is/is not active based on a specific battery outcome.
- Meets all the desired characteristics
  - structured,
  - transparent,
  - objective,
  - Quantitative
  - Updates hypothesis as new evidence arrives





# Example of Bayes' Theorem



- The location (mean) of the posterior distribution is between the ones for the prior and the likelihood, i.e. it is a weighted average between the two parameter estimates (from prior and likelihood).
- The location of the posterior distribution is closer to the location of the likelihood curve because there is more data in the new study (80 subjects) versus the “data” from the prior (30 subjects).
- The curve for the posterior is narrower than any of the other two curves.

# What is predictivity ?

For a 2-state condition (C+, C-) performance characteristics of a test (T+, T-) are

- Sensitivity  $\Pr(T+|C+)$ 
  - $\Pr(T+|S-)$  is high for many in vitro tests
- Specificity  $\Pr(T-|C-)$

**Characteristics of the validation set. Cannot tell much about state of the chemical.**

For Bayesians predictivity is assessed by Posterior Probability  $\Pr(C|T)$

- Predictive value positive

$$\Pr(C+ | T+) = \frac{\Pr(C+) \cdot Se}{\Pr(C+) \cdot Se + \Pr(C-)(1 - Sp)}$$

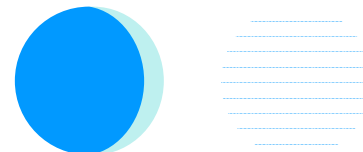
- Predictive value negative

$$\Pr(C- | T-) = \frac{\Pr(C-) \cdot Sp}{\Pr(C+) \cdot (1 - Se) + \Pr(C-)(Sp)}$$

- **Note reversal of the conditionality: given an outcome of the test (+;-) what is the probability a chemical is +/-**
- **Predictive value depends on the prior probability  $\Pr(S)$  –this is place to introduce chemical specific information : like prevalence of C+ in the chemical class**

# Part 2.

Case study – Bayesian interpretation  
of in vitro genetox battery to assess  
carcinogenicity



# Bayesian evaluation of carcinogenicity potential with in vitro genetox battery case study – data set

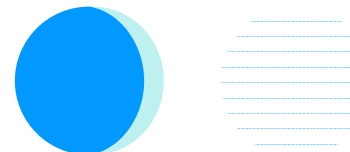
•Kirkland D., Aardema M., Henderson L. Müller L. 2005 Evaluation of the ability of a battery of 3 *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity Mut. Res. 584, 1-256 of over 700 chemicals

• <u>Test</u>	Sensitivity ( $\text{Pr}(T+ C+)$ )	Specificity ( $\text{Pr}(T- C-)$ )
Ames	0.59	0.74
MLA	0.73	0.39
CA	0.66	0.45

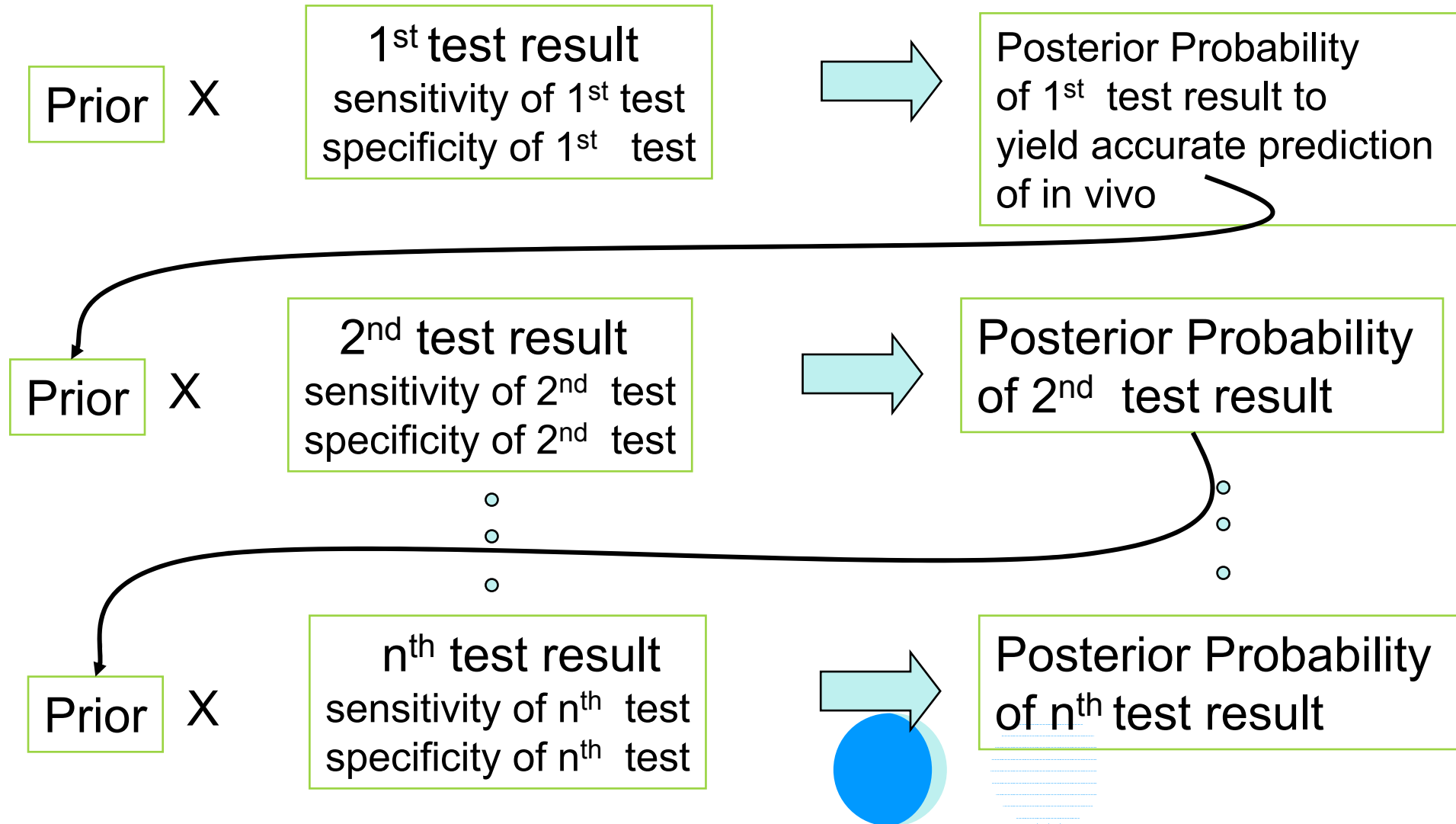
•T+ is a positive test result, C+ chemical is a rodent carcinogen *in vivo*, T- is a negative test result, C- chemical is not a rodent carcinogen *in vivo*.

Sensitivity= #carcinogens positive in the invitro genetox assay/#carcinogens evaluated;

Specificity= #noncarcinogens negative in the invitro genetox assay/#noncarcinogens evaluated;



# Bayesian calculations of a battery's predictivity



Bayesian interpretation of results from 3 genotoxicity *in vitro* tests: Ames (A), Mouse Lymphoma assay (MLA) and Chromosomal Aberration assay (CA) – strategy for analyses

- Influence of the prevalence or prior knowledge about the chemical property
- Increase predictivity
- Guide testing by quantifying uncertainty
  - Comparison of results from dependent test battery vs. independent test battery



# Predictive value of individual tests

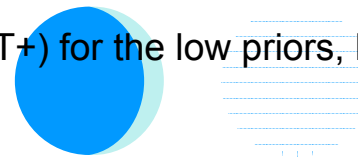
	P (C+ T)				
Prior	0.01	0.1	0.2	0.5	0.8
A+,	0.022	0.2	0.36	0.69	0.9
A-	0.006	0.06	0.12	0.36	0.69
MLA+	0.012	0.024	0.23	0.55	0.83
MLA-	0.007	0.014	0.15	0.41	0.74
CA+	0.012	0.024	0.23	0.55	0.83
CA-	0.008	0.015	0.16	0.43	0.75

- This type of analysis allows us to
  - gives perspective on the “check-box” approach;
  - encourages to develop a prior- i.e. develop a hypothesis before we test
  - Clearly, one must be careful not to conflate the quite distinct notions of the “truth” of the data and the “significance” of the data,
- The chance that a chemical is C+ after A+ increases about 2 times
- The chance that a chemical is C+ after A- decreases about 2 times
- Predictive values of MLA and CA are about the same across chemical classes with different prevalence of C+

# Predictive value of 2 tests Ames and MLA

	C+	P(C-)	C+	P(C-)	C+	P(C-)	C+	P(C-)
P(C+)	0.1		0.2		0.5		0.8	
P(C+ A+)	0.2	0.8	0.36	0.64	0.69	0.31	0.9	0.1
P(C+ MLA+, A+)	0.25	0.75	0.43	0.57	0.75	0.25	0.92	0.08
P(C+ MLA-, A+)	0.09	0.91	0.19	0.81	0.48	0.52	0.79	0.21
P(C+ A-)	0.06	0.94	0.12		0.36		0.69	
P(C+ MLA+, A-)	0.08	0.92	0.16	0.84	0.43	0.57	0.75	0.25
P(C+ MLA-, A-)	0.04	0.96	0.08	0.92	0.27	0.73	0.60	0.40

- MLA + after Ames+ improves our belief that a chemical is carcinogenic only by maximum 6%. This might be seen unexpected to the readers used to consensus counts who would interpret positive result in 2 tests doubles the probability that a chemical is carcinogen compared to positive result in 1 test.
- MLA- after Ames- is more informative because in the range of priors 0.5-0.8 we gain about 10% in certainty that a chemical is not a carcinogen.
- Predictive values of MLA+, A- and MLA-, A+ are about the same across all priors.
- If we cut off 0.7 for accepting the conclusion that a chemical is a carcinogen then for a prior 0.5 MLA+, Ames + would suffice to reach this conclusion, for the prior of 0.8 MLA+, Ames+ and MLA-, Ames +, would suggest that a chemical is a carcinogen. For the prior 0.5 MLA-, Ames +, MLA+, Ames- predictive values suggests that we need to generate more data to refine our assessment.
- It is important to note a high number of false positives  $P(C-|T+)$  for the low priors, but also a high number of false negatives  $P(C+|T-)$  for high priors

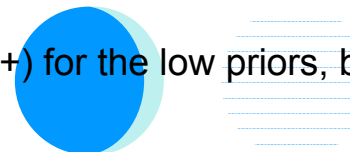




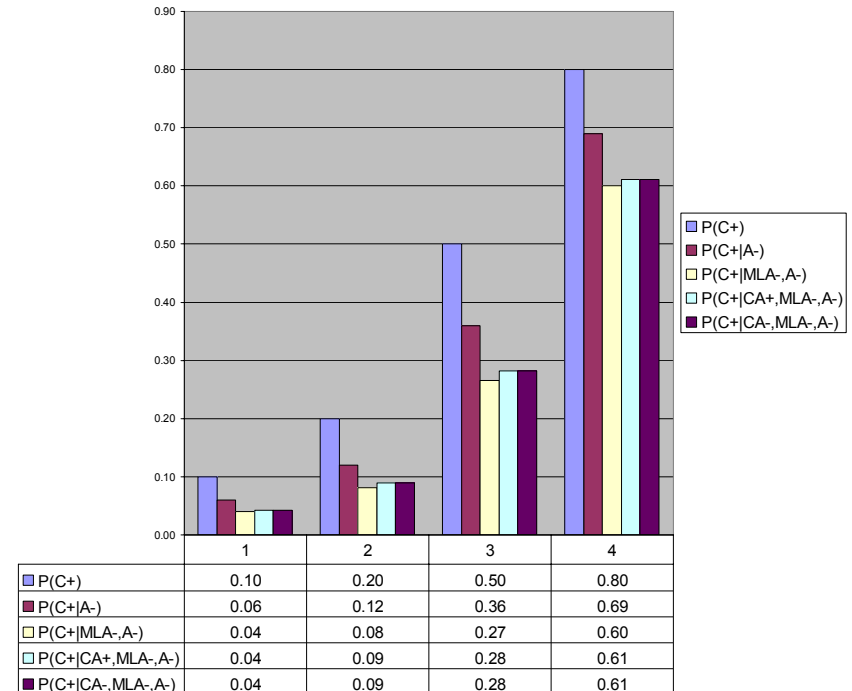
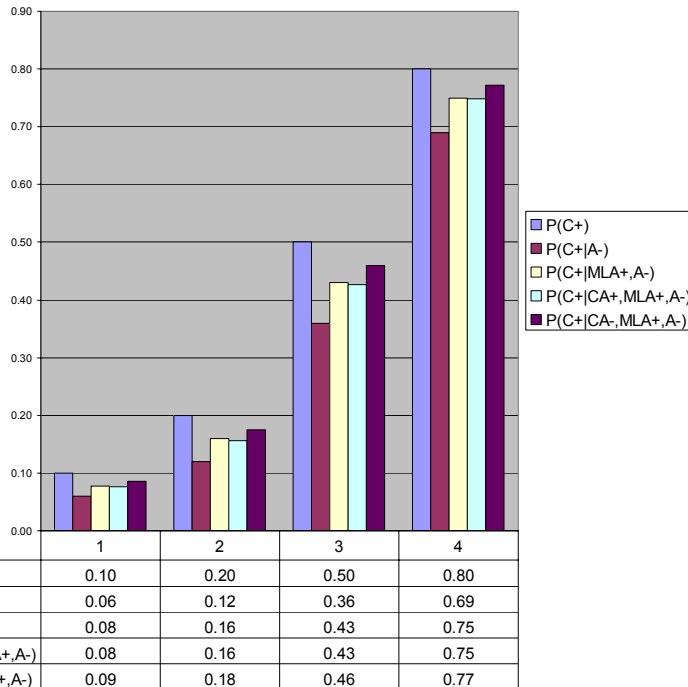
# Predictive value of 2 tests – Ames and CA

	C+	P(C-)	C+	P(C-)	C+	P(C-)	C+	P(C-)
P(C+)	0.1		0.2		0.5		0.8	
P(C+ A+)	0.2	0.8	0.36	0.64	0.69	0.31	0.9	0.1
P(C+ CA+, A+)	0.24	0.76	0.42	0.58	0.74	0.26	0.92	0.08
P(C+ CA-, A+)	0.08	0.92	0.16	0.84	0.43	0.57	0.75	0.25
P(C+ A-)	0.06	0.94	0.12		0.36		0.69	
P(C+ CA+, A-)	0.07	0.93	0.14	0.86	0.39	0.61	0.72	0.28
P(C+ CA-, A-)	0.06	0.94	0.12	0.88	0.36	0.64	0.70	0.30

- CA + after Ames+ improves our belief that a chemical is carcinogenic by maximum 5%. This again might be seen unexpected to the readers used to consensus counts who would interpret positive result in 2 tests doubles the probability that a chemical is carcinogen compared to positive result in 1 test.
- CA - after Ames- is not informative because for small priors the predictive value is the same as predictive value of Ames negative, and for larger priors the difference is just over 1%.
- When resolving conflicting data: CA+, A- or CA-, A+ we can see a difference compared to MLA and Ames conflicting results. Probability that a chemical is a carcinogen given CA-, Ames+ is 3-4% higher compared to CA+, Ames- results for the priors in the 0.5-0.8 range. The result CA-, Ames + is marginally more conclusive than CA+, Ames-.
- Following our 0.7 cut-off rule for CA, Ames battery and 0.5 prior we would only accept CA+, Ames+ as sufficiently conclusive. Other CA, Ames battery outcomes are not conclusive and need to generate more data to refine our assessment.
- It is important to note a high number of false positives  $P(C-|T+)$  for the low priors, but also a high number of false negatives  $P(C+|T-)$  for high priors.



# Predictive value of 3 tests – Ames-, MLA, CA



- Adding MLA after A- is valuable because if result is negative then the chance is further reduced by 10%, if positive the chance that a chemical is C+ increases but by only 5%.
- Conducting CA is not bringing significant refinement to the evaluation of the carcinogenicity potential.

Adding CA after A-, MLA- brings no further refinement  
 Adding CA after A-, MLA+ brings no further refinement



# Resolving conflicting results

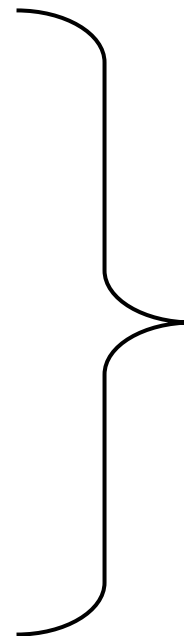
- The approach helps to weigh uncertainty when resolving conflicting results:

- (+, +, -)

- $P(C+|CA-, MLA+, A+)=0.65$
- $P(C+|CA+, MLA-, A+)=0.64$
- $P(C+|CA+; MLA+, A-)=0.41$
- $P(C+|CA+, MLA+, A+)=0.74$

- (-, -, +)

- $P(C+|CA+, MLA-, A-)=0.27$
- $P(C+|CA-, MLA+, A-)=0.45$
- $P(C+|CA-, MLA-, A+)=0.37$
- $P(C+|CA-, MLA-, A-)=0.27$

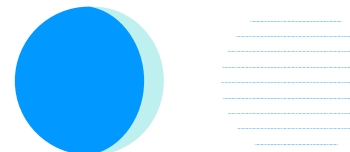


$\Pr(C+)=0.5$



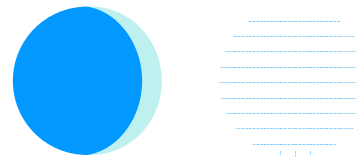
# Conditional dependence between tests

- Tests have overlapping mechanisms
- Tests are used to predict the same endpoint
- Varying degree of conditionality was observed for different tests outcomes
- $P(\text{CA-}, \text{MLA-}, \text{A+}) \text{ i-d} = 0.52 - 0.34 = 0.17$
- Assuming independence we tend to overpredict
- When tests are more predictive conditionality will deflate predictive values of the battery
- Useful perspective when thinking how many, and how good tests we need to reach a decision based on the battery

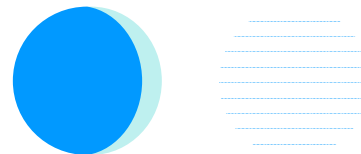


# What about hypothesis driven approach

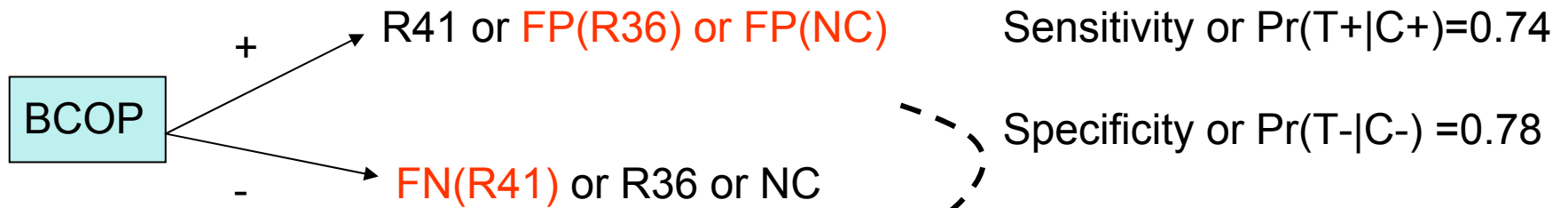
- For the genetox battery need to add a test checking if a chemical is genotoxic or epigenetic carcinogen
  - Experimental test
  - Chemoinformatic approach ( Toxtree)
- In general
  - As we understand more and more about molecular pathways we will be developing hierarchical integration schemes



# Part 3. Exploratory work with eye irritation data



# Decision schema using BCOP and NI/I

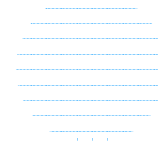
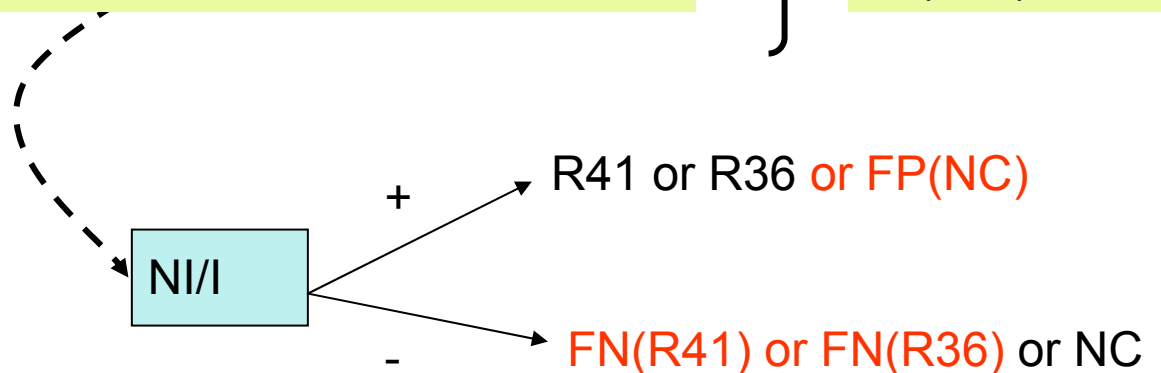


$$\Pr(R36 \text{ or } NC | BCOP+) = \Pr(R36 | BCOP+) + \Pr(NC | BCOP+) \\ \Pr(R41 | BCOP+)$$

$$\Pr(R41 \text{ or } R36) \\ \Pr(NC)$$

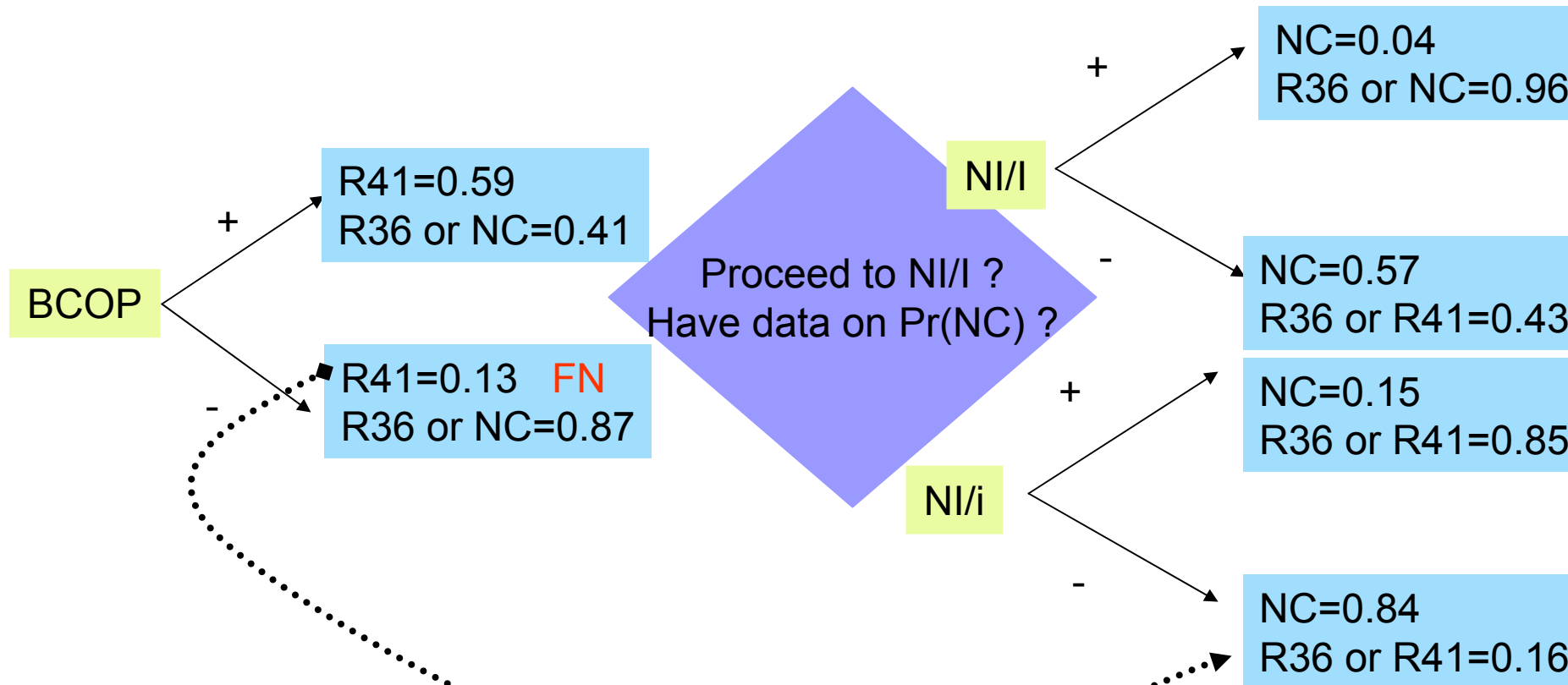
Sensitivity or  $\Pr(T+|C+)=a$

Specificity or  $\Pr(T-|C-)=b$



# Building decision tree

for anionic surfactant



By conducting NI/I after BCOP we may reduce probability of FN  
but quantitative answer requires knowledge of R36/R41 ratio



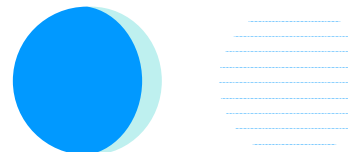
# Summary

Application of a Bayesian battery approach meets multiple criteria for integration strategy

- allows to quantify uncertainty propagation in a tiered mode
- Refined predictivity of the battery compared with simple consensus approach
- Transparent resolution of conflicting information

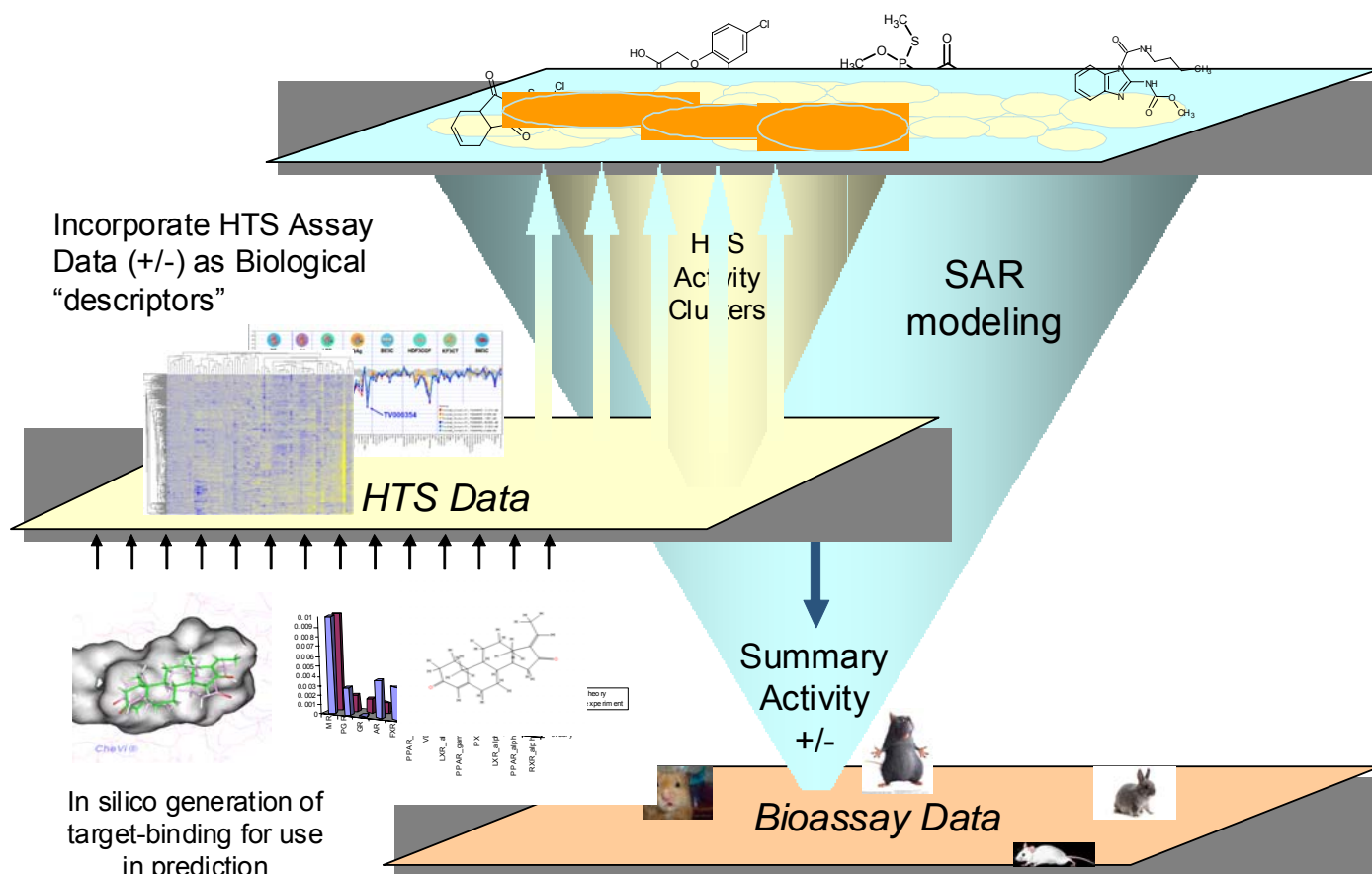
Large differences in interpreting result on activity/lack of it depending on the prior –

Develop the hypothesis!



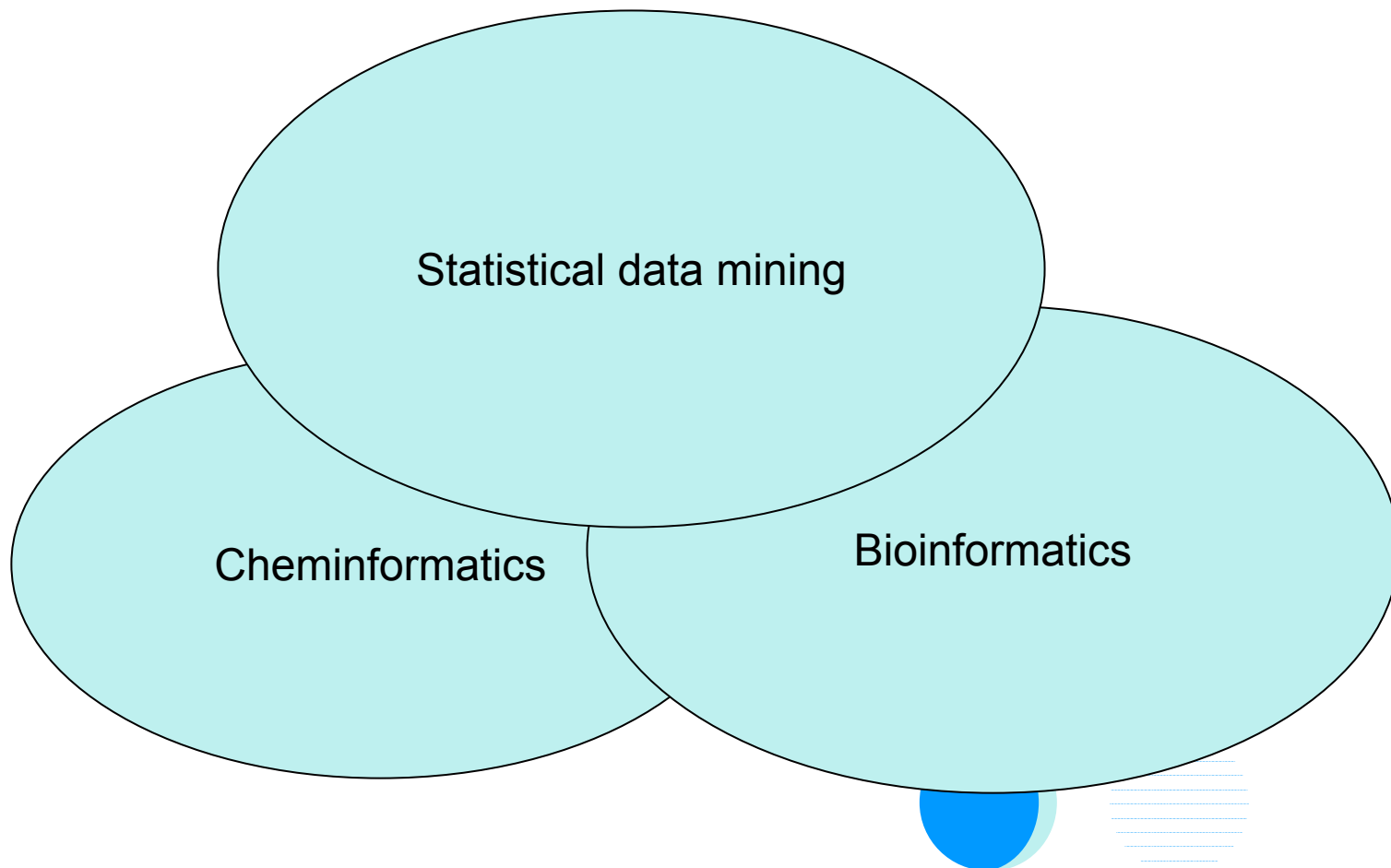
# Comment on the prior from chemical class to mechanism based

## Structure-Activity Approaches to Toxicity Prediction



# Outlook

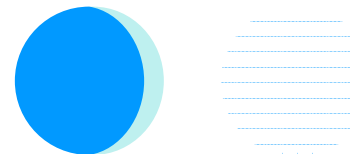
This analysis was done on the data set level the next step is to do it on the compound (mechanistic hypothesis, good prior) - “individualized “ toxicology ?



# Outlook

This approach can be applied in development of/ and even validation of in vitro testing strategies

Can be also helpful in setting up realistic policies by quantifying FN, or FR rates based on the tests available



# Acknowledgements

Tom Aldenberg RIVM

Robert McDowell USDA

EU 6<sup>th</sup> Framework Integrated Project Osiris

