Advances in multi-parameter optimisation: Targeting the "best" profile for your project's objectives

ACS Spring National Meeting, March 16th 2013

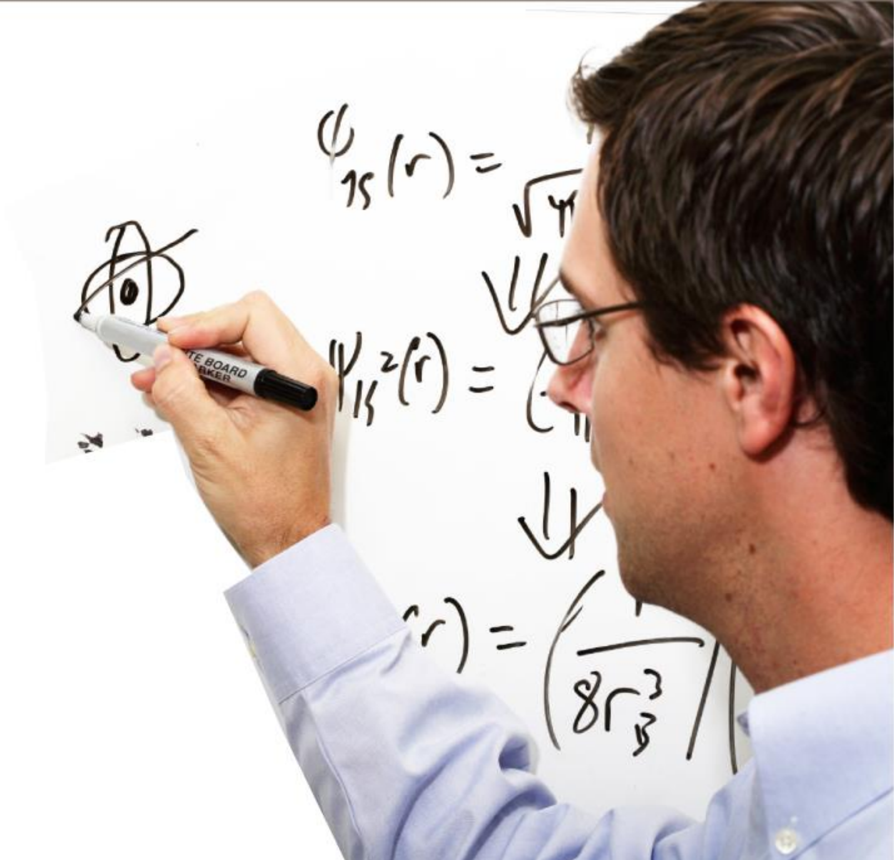
Matthew Segall, Edmund Champness

Overview

- Multi-parameter optimisation in drug discovery
- Finding the ‘best’ profile for your project’s objective
  - Example: Selection to reduce toxicity risk
- ‘Hard’ vs. ‘soft’ boundaries
  - Example: Selection for CNS indications
- Testing the robustness of your decisions
  - Sensitivity analysis
- Conclusions

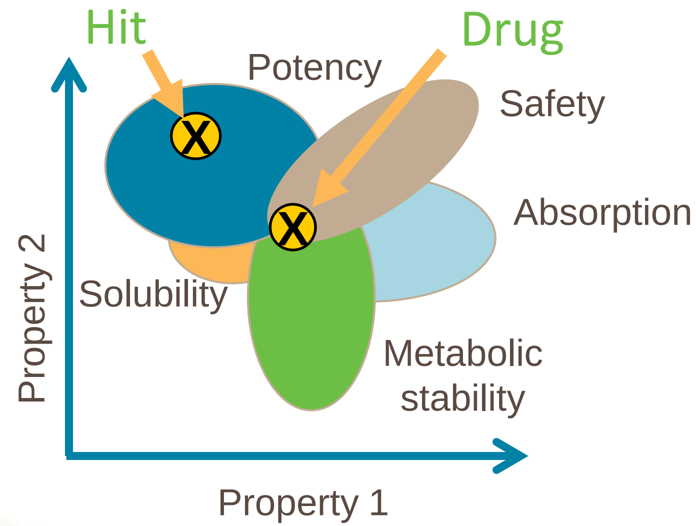
Multi-parameter Optimisation in Drug Discovery

optibrium

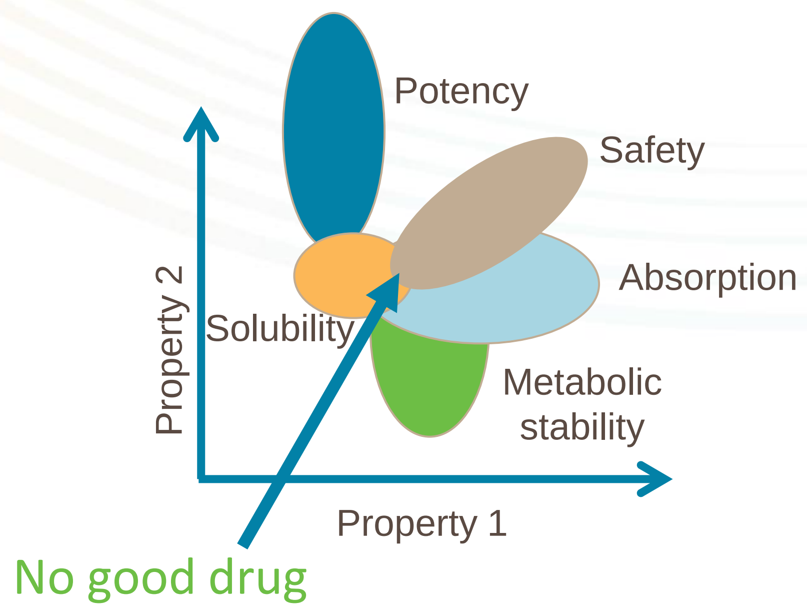


The Objectives
Multi-parameter optimisation

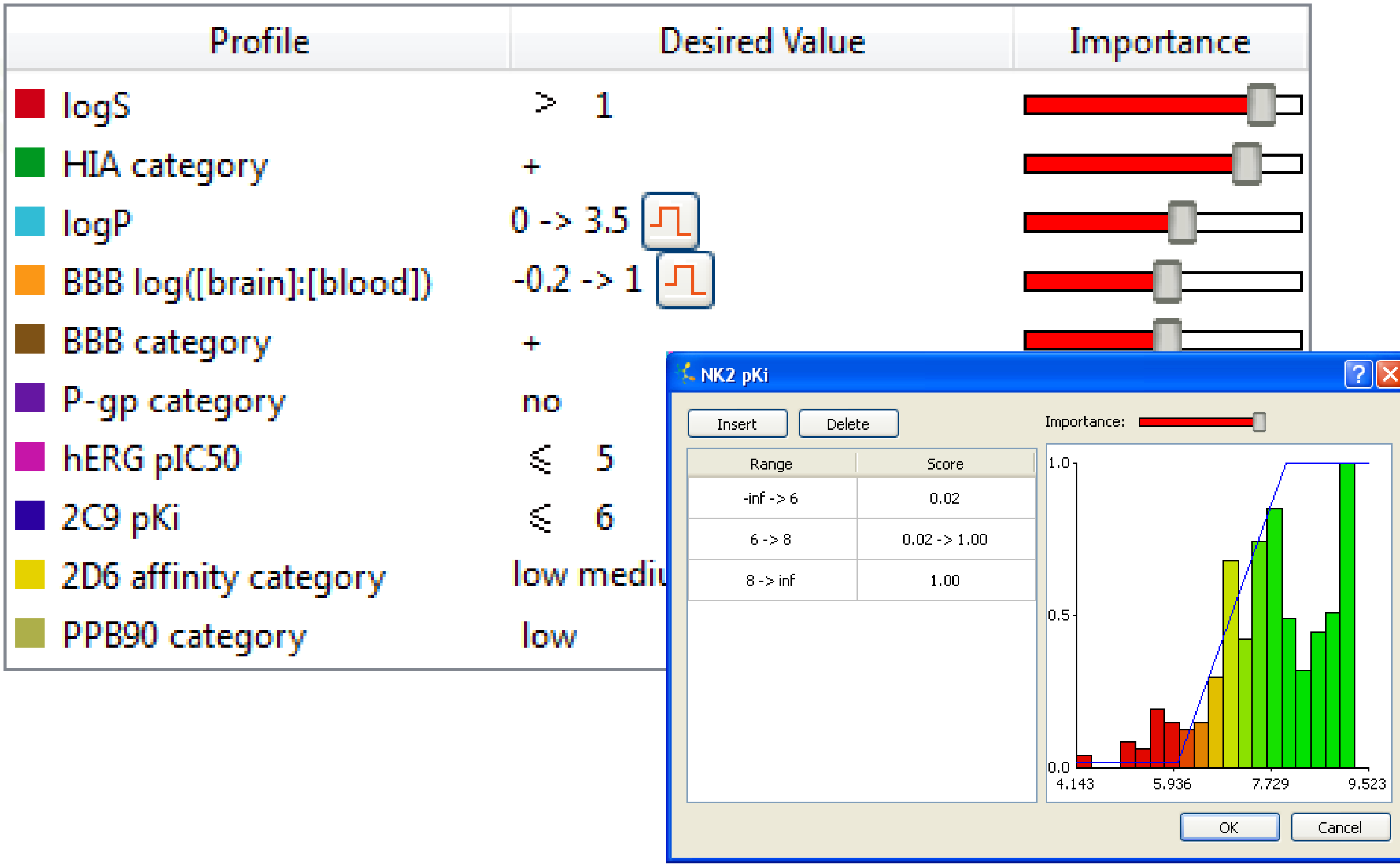
- Identify chemistries with an optimal **balance** of properties



- Quickly identify situations when such a balance is not possible
  - –Fail fast, fail cheap
  - –Only when **confident**



Multi-parameter Optimisation
Probabilistic Scoring*

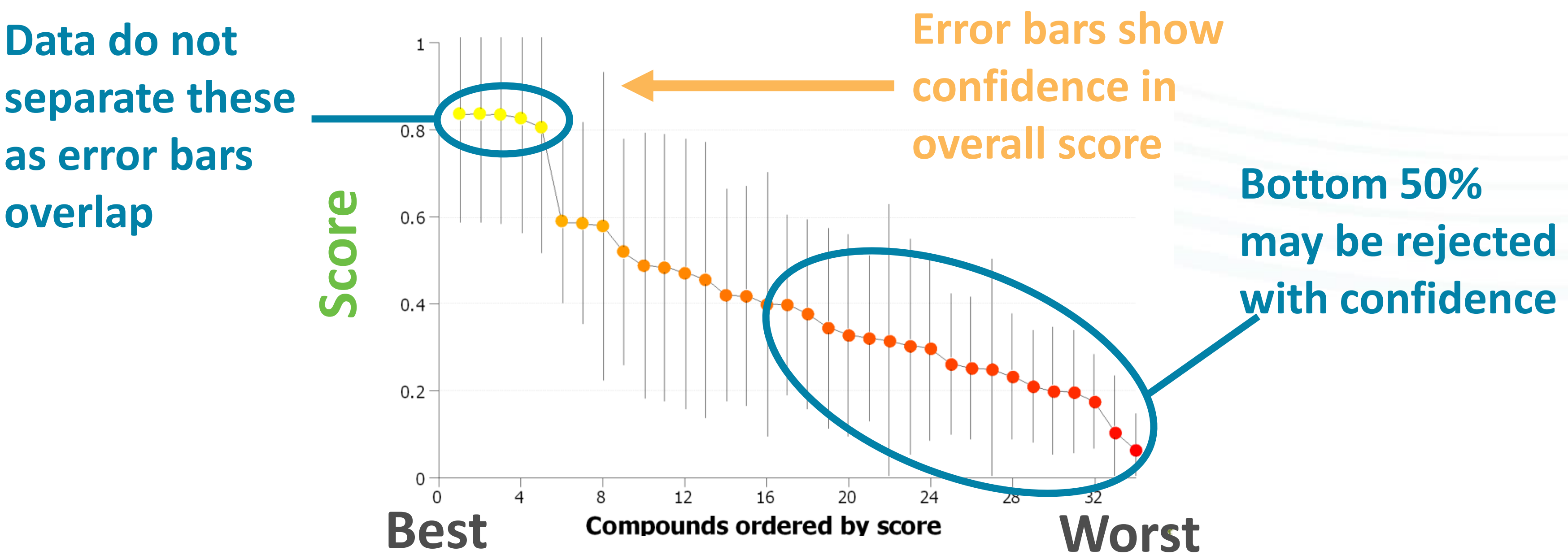


*M.D. Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

Multi-parameter Optimisation
Probabilistic Scoring*

- **Property data**
  - Experimental or predicted
- **Criteria for success**
  - Relative importance
- **Uncertainties in data**
  - Experimental or statistical

- **Score (Likelihood of Success)**
- **Confidence in score**

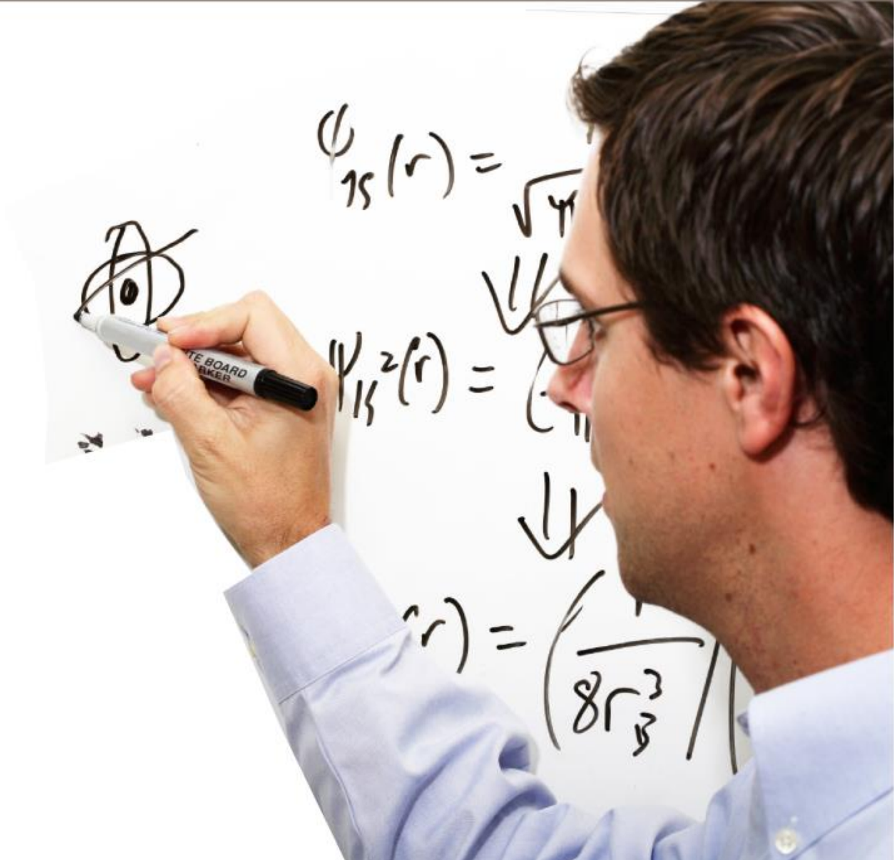


*M.D. Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

Finding the ‘Best’ Profile for your Project Objectives

Patent pending

optibrium



Finding Tailored Profiles
Objectives

- Use existing data to find scoring profiles that identify compounds with improved chance of success
  - Any drug discovery objective, e.g. clinical, PK, toxicity...
  - Once developed, a profile can be applied prospectively to find new compounds
- Identify most important data with which to distinguish between successful and unsuccessful compounds
  - Any data can be used as input, calculated or experimental
- Explore multi-parametric data
  - Consider properties simultaneously, not individually
  - Avoid ‘over counting’ of correlated factors
- Rules must be interpretable and modifiable
  - Avoid black boxes
  - Synergy between computer and experts

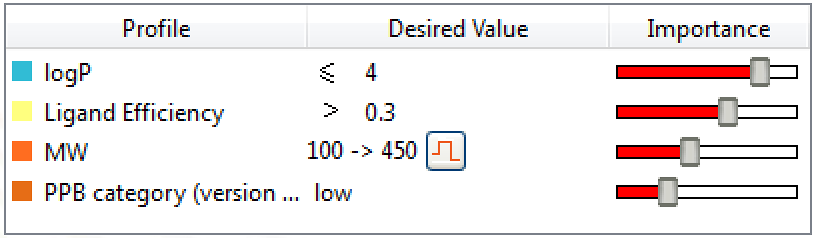
*Patent pending

I. Yusof *et al.* (2014) Drug Discov. Today DOI: 10.1016/j.drudis.2014.01.005

What is a Rule?

- A **Rule** is a set of property criteria that in **combination** identify ‘good’ compounds, e.g.

logP < 4
Ligand efficiency > 0.3
100 < MW < 450
PPB category = low

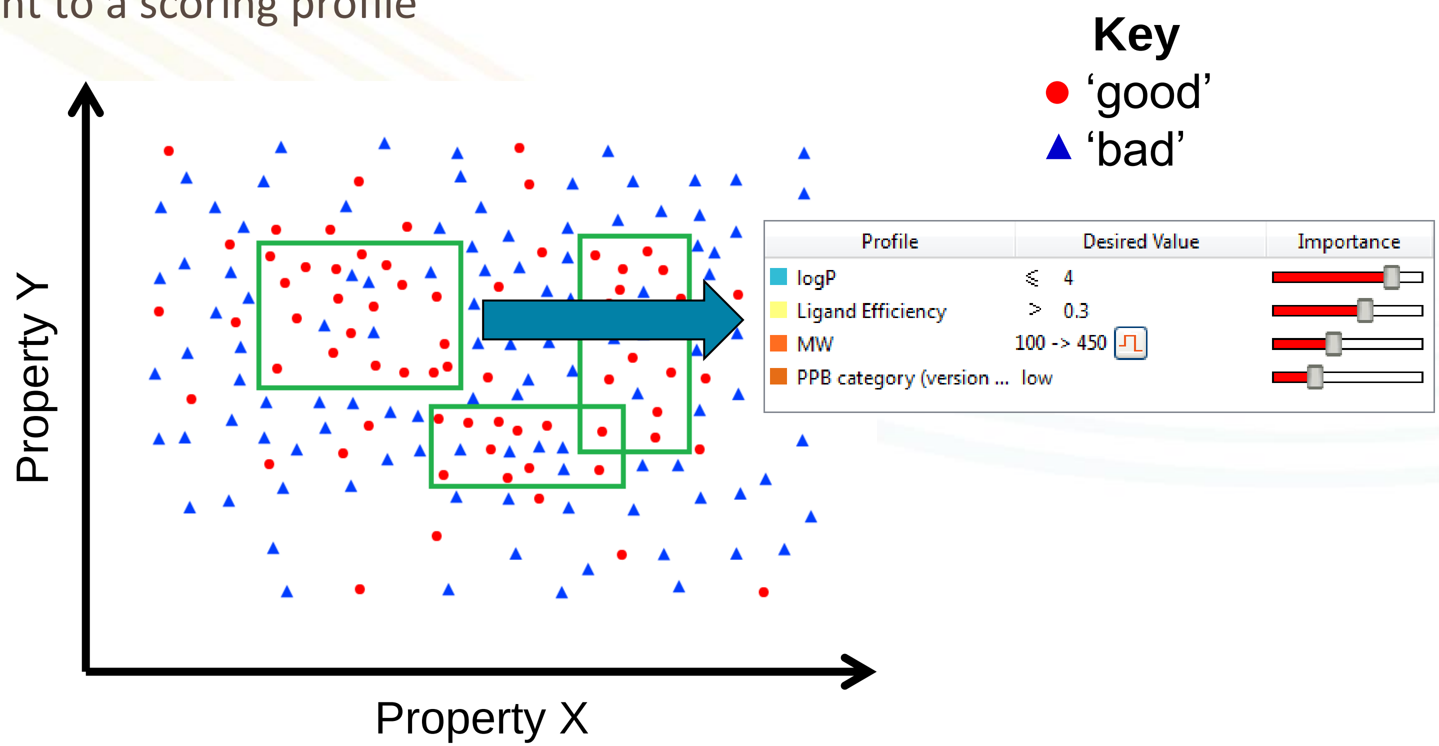


- For example, Lipinski RoF:

logP<5	MW<500
HBD<5	HBA<10

Finding Rules with PRIM

- A **Rule** is a box in multi-dimensional property space containing significantly more ‘good’ than ‘bad’ compounds
  - Use Patient Rule Induction Method (PRIM) by Friedman and Fisher* find rules in multi-dimensional data
  - Equivalent to a scoring profile



* Friedman & Fisher Stat. and Comp. **9**(2), p. 123 (1999)

Example Application
Finding rules for selection of non-toxic compounds

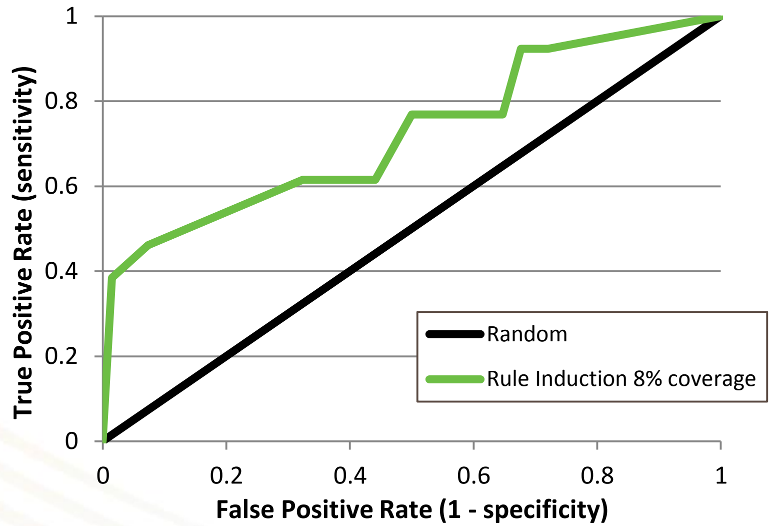
- *In vitro* assay data from CEREP Bioprint®
  - Percentage inhibition of 185 targets including GPCF, kinase, NR, P450s...
- Drugs labelled as ‘cardiotoxic’, ‘hepatotoxic’ or ‘clean’
  - Based on FDA Adverse Event Reporting System
  - Reporting odds ratio (ROR) of 2.5 or above at System Organ Class level in MeDRA Ontology
  - Cardiotoxicity set: 408 ‘cardiotoxic’ ,66 ‘non-cardiotoxic’
  - Heptotoxicity set: 302 ‘hepatotoxic’ , 168 ‘non-hepatotoxic’
- Data sets divided into training, validation and test sets
  - Ratio 70:15:15

 I. Yusof *et al.* (2014) Drug Discov. Today DOI: 10.1016/j.drudis.2014.01.005

Example Application
Cardiotoxicity results

Profile	Desired Value	Importance
H2	-0.01 -> 2.02	
5HT1A	≤ 8.08	
A1	≤ -0.99	

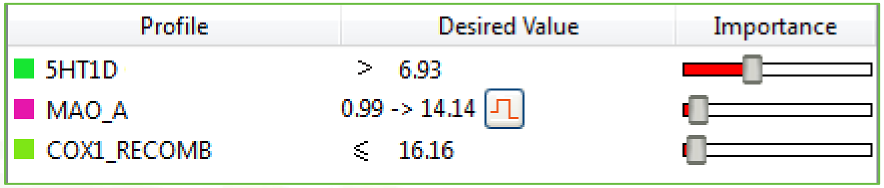
Set	Mean Improvement (%)	Support (%)
Train	233	9.3
Val	173	10
Test	419	7.4



- Selected only 3 targets from 185
  - Rules ‘make sense’: Targets identified have known CV side effects
- 5/6 compounds meeting all criteria are non-cardiotoxic (83%)
- 19/20 compounds failing all criteria are cardiotoxic (95%)

 I. Yusof *et al.* (2014) Drug Discov. Today DOI: 10.1016/j.drudis.2014.01.005

Example Application
Hepatotoxicity results

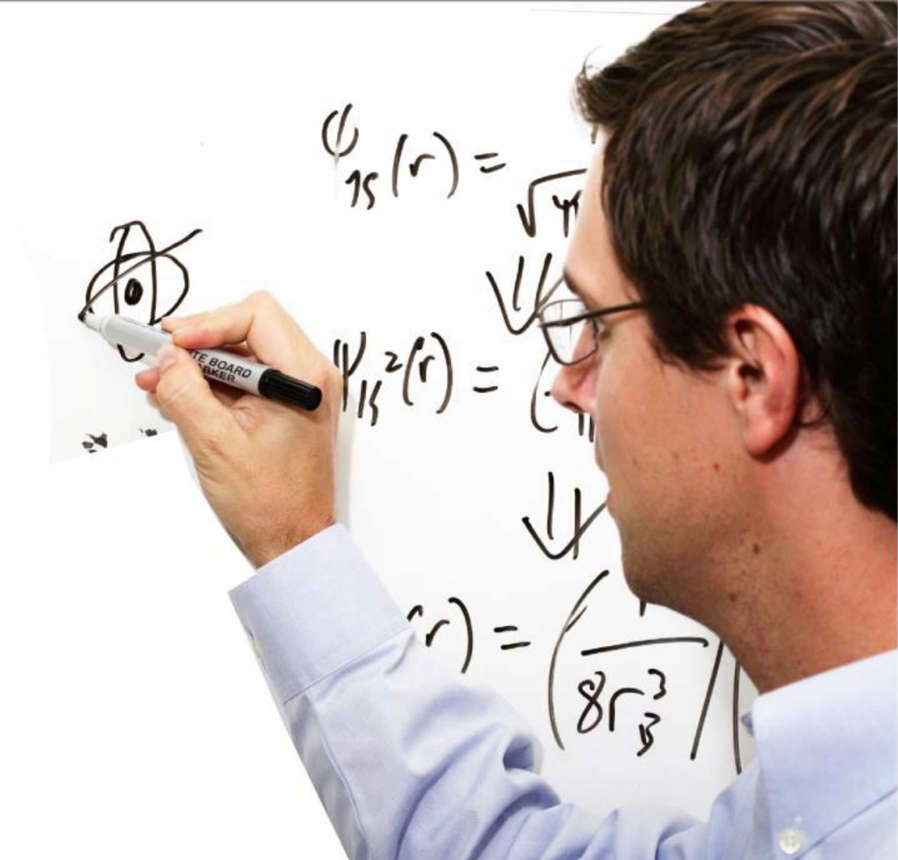


Set	Mean Improvement (%)	Support (%)
Train	51	12
Val	56	14
Test	39	11

- Rules are (just) statistically significant, but don’t ‘make sense’
  - Rules appear to be result of noise in small data set
- Large majority of the targets in data set are not known to relate with hepatotoxicity
  - In few examples, e.g. PPARγ there are a statistically insignificant number of inhibitors in the data set
- Non ‘black-box’ method highlights limitations of data set

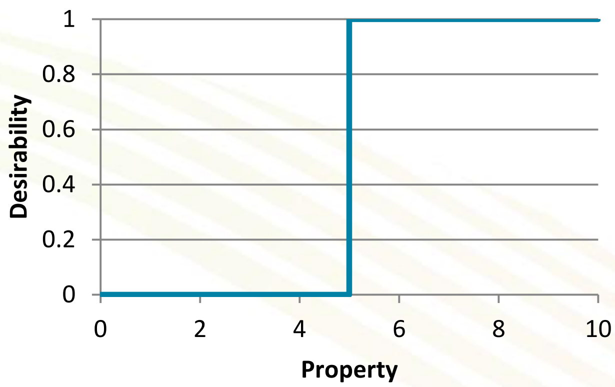
‘Hard’ vs. ‘Soft’ boundaries

Patent pending



Desirability Functions*

- Relate property values to how ‘desirable’ the outcome



Simple filter: >5

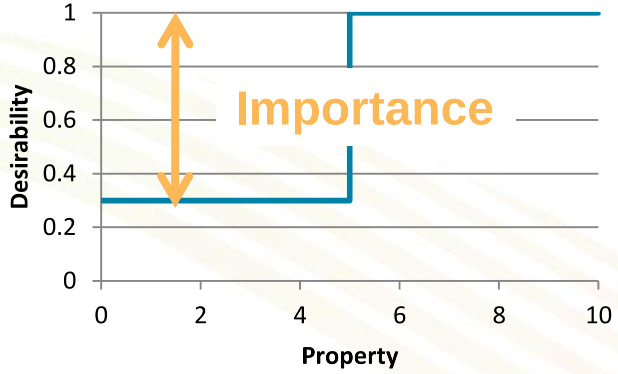
- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add ‘soft’ boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. **21** p. 494

Desirability Functions*

- Relate property values to how ‘desirable’ the outcome



Desired value: >5

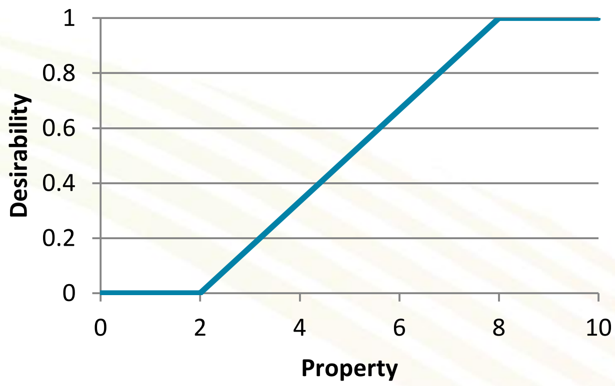
- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add ‘soft’ boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. **21** p. 494

Desirability Functions*

- Relate property values to how 'desirable' the outcome



Trend: >8

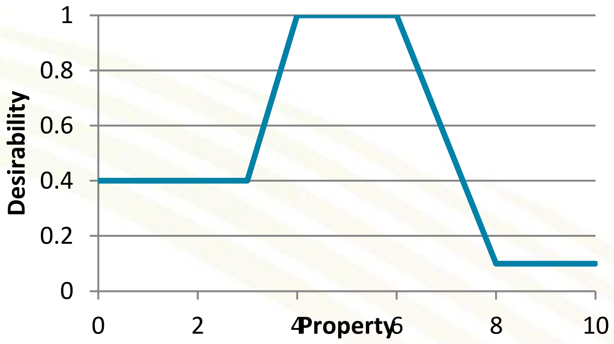
- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add 'soft' boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. **21** p. 494

Desirability Functions*

- Relate property values to how 'desirable' the outcome



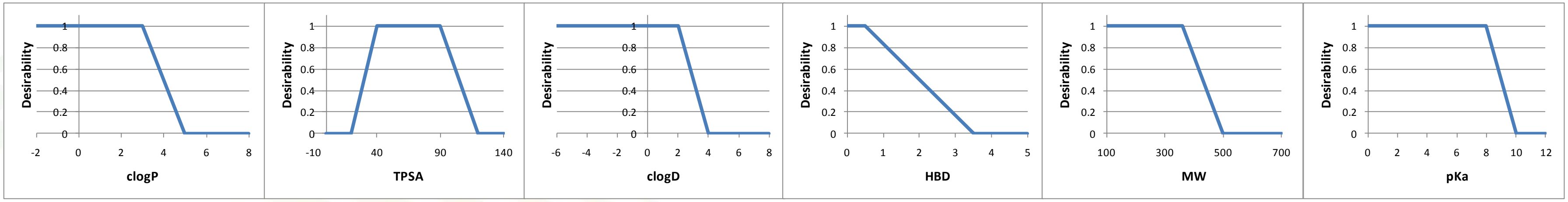
'Soft' range: 4-6

- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add 'soft' boundaries to ideal ranges

Desirability Functions
Example: CNS MPO*

clogP **TPSA** **clogD** **HBD** **MW** **pKa**



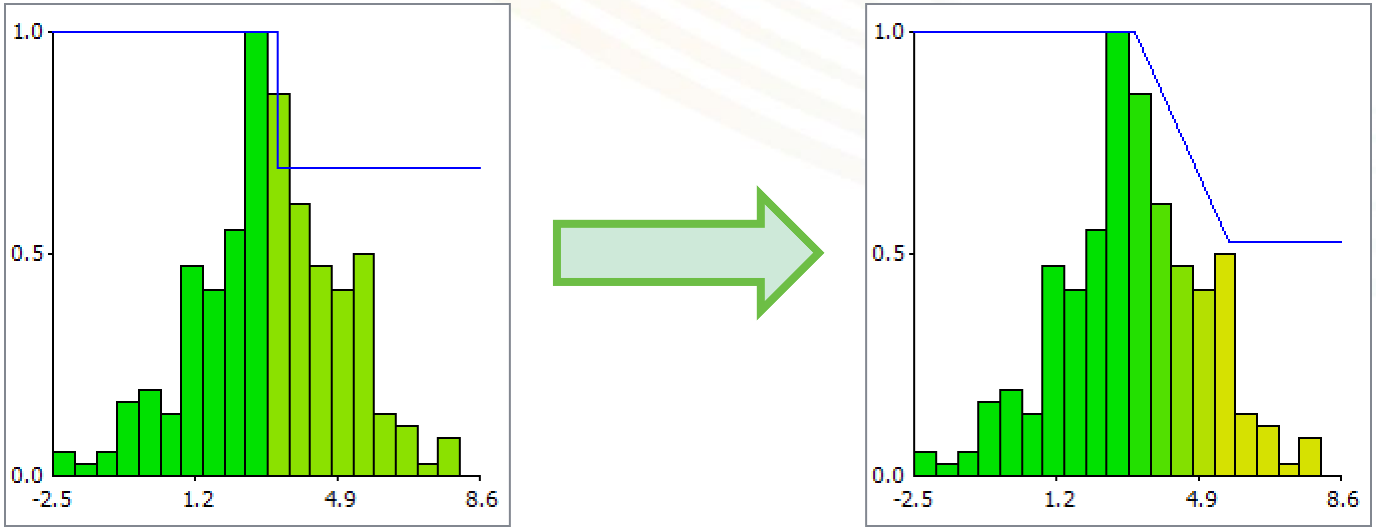
CNS MPO = sum of desirabilities for each parameter

- 74% of marketed CNS drugs achieved CNS MPO > 4 vs. 60% of Pfizer candidates
- Correlations observed between high CNS MPO score and good *in vitro* ADME properties, e.g. MDCK P_{app}, HLM stability, P-gp transport

*Wager *et al.* (2010) ACS Chem. Neurosci. **1** p. 435

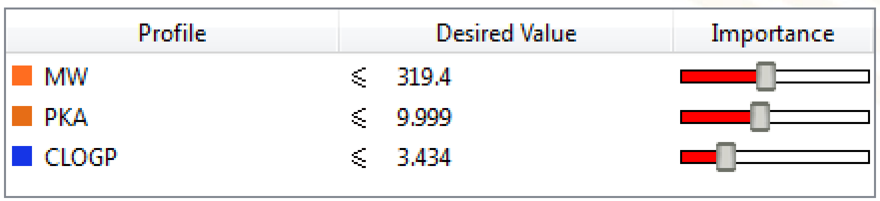
Determining ‘Soft’ Box Boundaries

- Box bounds previously only output as hard cut-offs
- Sensitivity analysis of box bounds to data sampling
  - Particularly important for sparse data
  - Incorporate uncertainty into the generated box bounds
  - Cross validation between training/validation sets



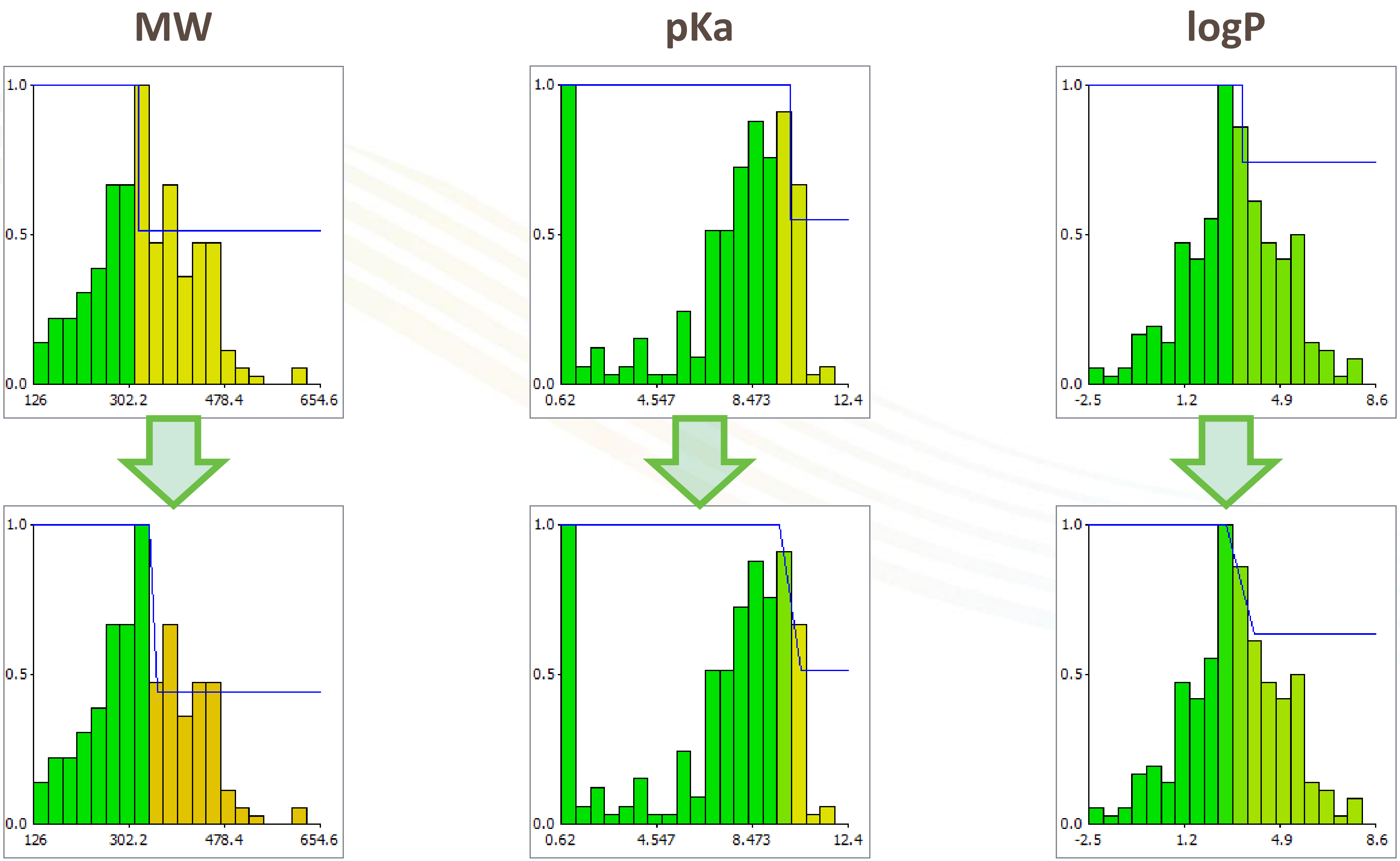
Example Application
CNS Drugs

- Data set of 119 CNS Drugs and 108 Candidates published by Wager *et al.* in CNS MPO paper
- Divided into training, validation and test sets (55:25:20)
- Rule with hard cut-offs:



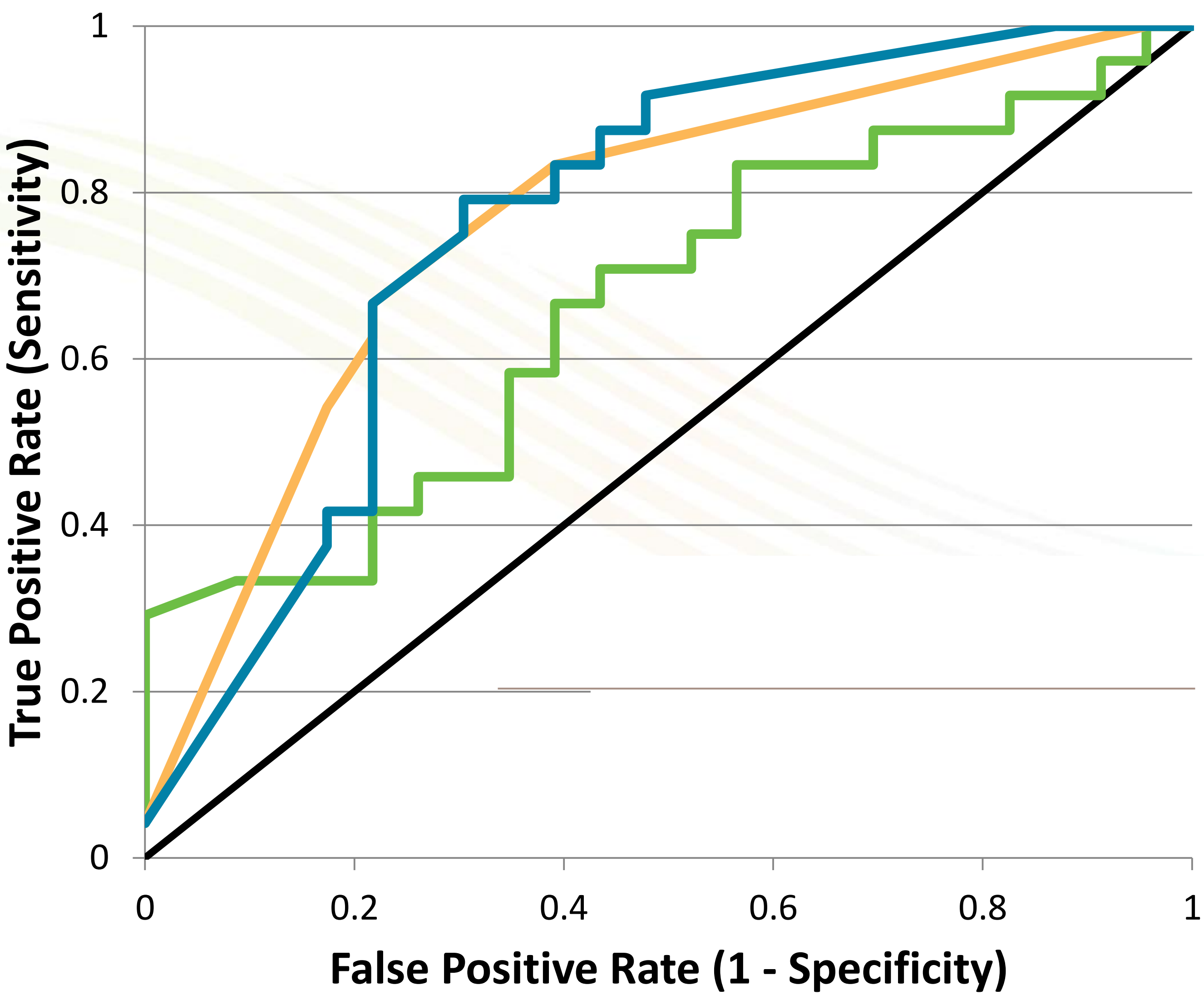
Set	Mean Improvement (%)	Support (%)
Train	42	28
Val	56	32
Test	47	34

Example Application
CNS Drugs – Introducing ‘soft’ boundaries



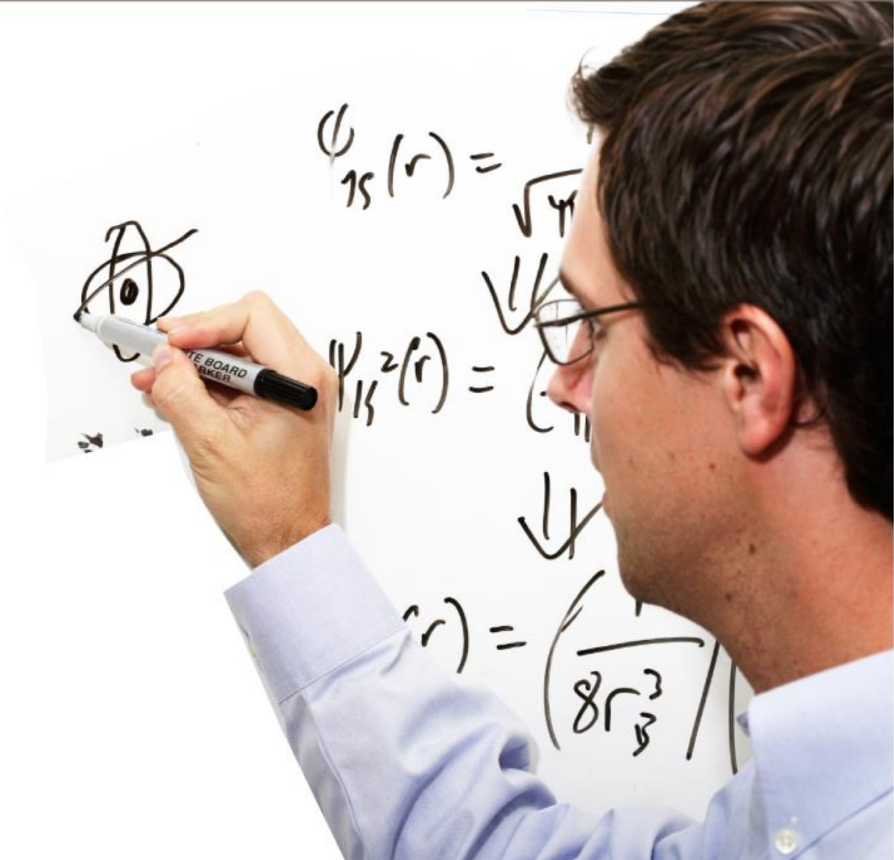
Example Application

CNS Drugs – Comparison of ROC curves for test set



Testing the Robustness of Your Decisions

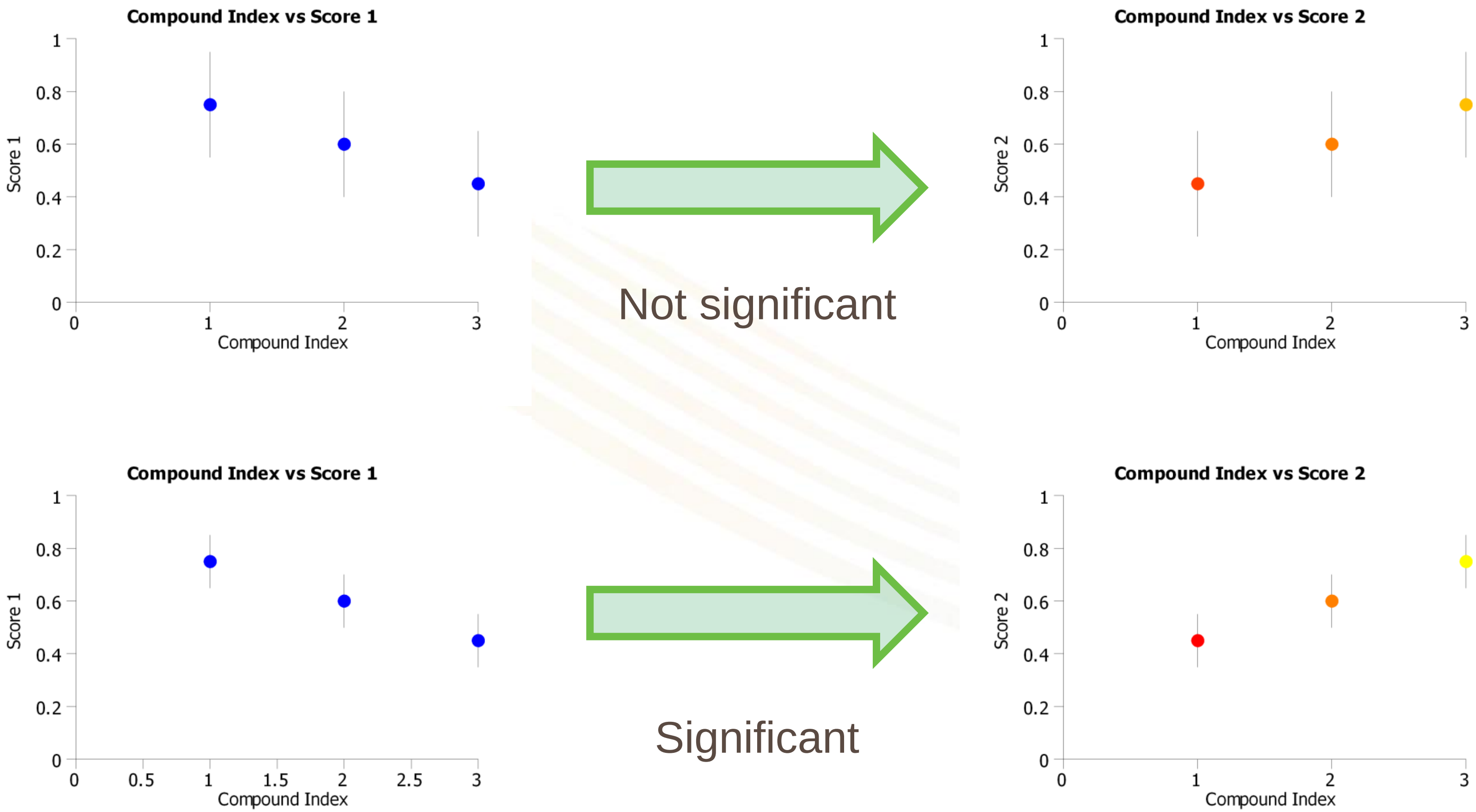
Patent pending



Sensitivity Analysis

- What impact would changing a property criterion have on the *decision* we would make?
  - How large a change is necessary to have a significant impact?
- To which property criteria is compound priority most *sensitive*?
  - Which criteria/importance will, if modified, significantly change the order of compound priority?
- Highlight new avenues for exploration and avoid missed opportunities
- Considerations
  - Need to consider statistical significance of reordering (given uncertainties in scores)
  - Interested in changes to high-ranked compounds. Reordering of rejected compounds is not relevant

Sensitivity Analysis
Importance of uncertainty

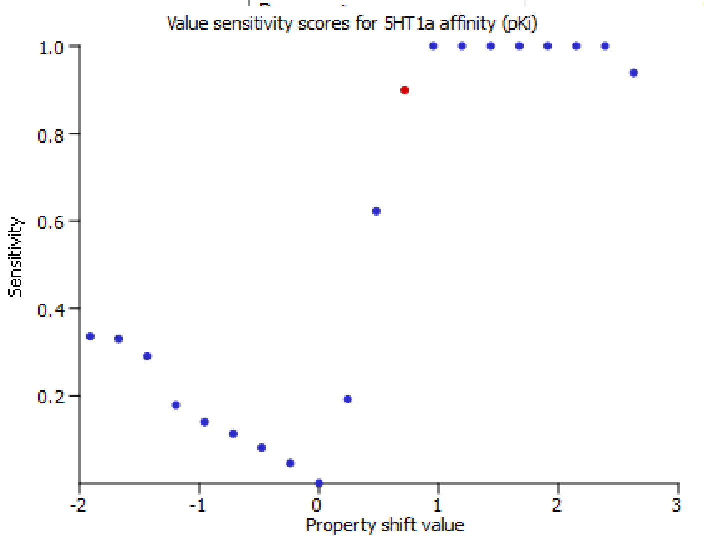


Modified Spearman's rank correlation coefficient accounts for uncertainty

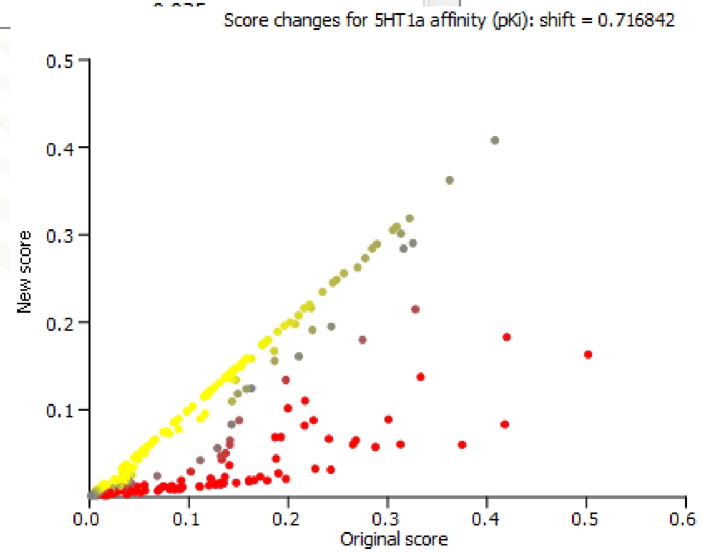
Example Output
Sensitive parameter

What parameters are most sensitive?

	Value Sensitivity	Importance Sensitivity
5HT1a affinity (pKi)	1.000	0.008
logP	0.310	0.096
BBB log([brain]:[blood])	0.249	0.015
hERG pIC50	0.096	0.207
2D6 affinity category	N/A	0.107
BBB category	N/A	0.055
logS	0.040	0.002
[illegible]	N/A	[illegible]



What magnitude of change has a significant impact?

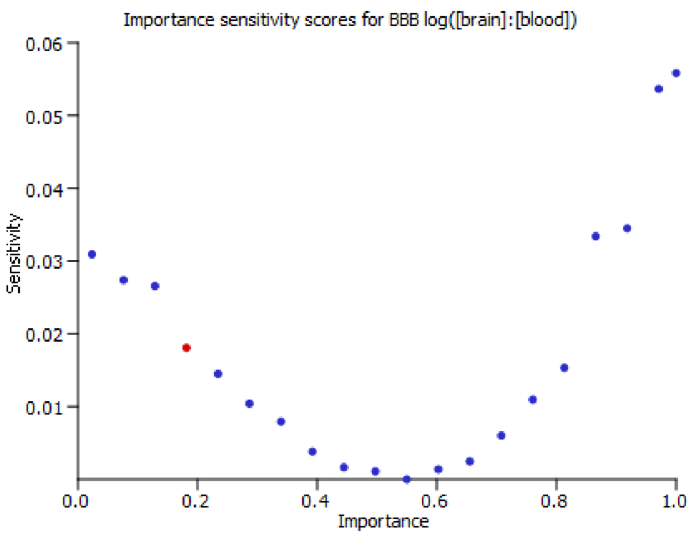


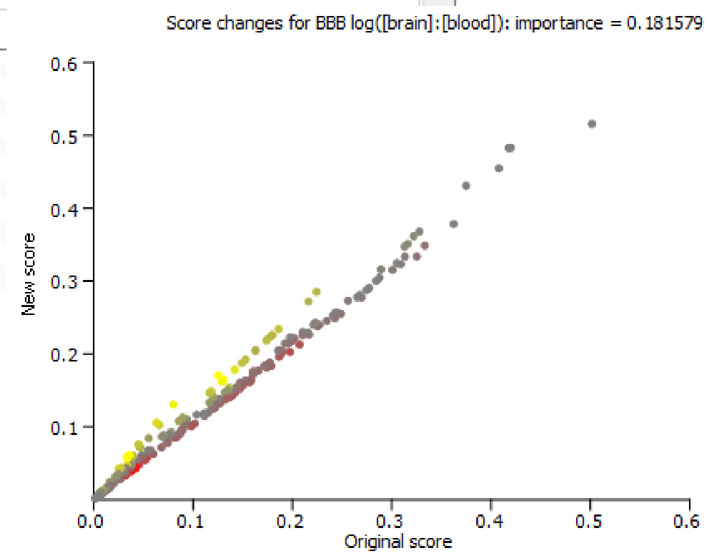
What compounds are most affected?

Example Output
Insensitive parameter

What parameters are most sensitive?

	Value Sensitivity	Importance Sensitivity
5HT1a affinity (pKi)	1.000	0.008
logP	0.310	0.096
BBB log([brain]:[blood])	0.249	0.015
hERG pIC50	0.096	0.207
2D6 affinity category	N/A	0.107
BBB category	N/A	0.055
logS	0.040	0.002
[illegible]	N/A	





What magnitude of change has a significant impact?

What compounds are most affected?

Conclusion

- Rule induction can generate interpretable parameter scoring profiles tailored project objectives

- 'Soft' boundaries provide more subtle distinctions between compounds

- Sensitivity analysis of scoring criteria is important to avoid missed opportunities due to the criteria we have chosen

- Reference (Rule induction): Yusof *et al.* Drug Discov. Today (2014)
  - 10.1016/j.drudis.2014.01.005
  - www.optibrium.com/community/publications

- See a live demo at Optibrium booth #1516