Data is Like Paint… It Does No Good Until it is Applied

ACS Fall Meeting 2018

Matthew Segall – matt@optibrium.com

© 2018 Optibrium Ltd.
Optibrium™, StarDrop™, Auto-Modeller™, Card View™ and Glowing Molecule™ are trademarks of Optibrium Ltd.

Overview
Confidence

- Making decisions based on data
  - Uncertainty everywhere!
- How good do our models need to be?
- Conclusions

Knowledge is like paint… It does no good until it is applied

- Doe Zantamata

© 2018 Optibrium Ltd.

Sources of Uncertainty
Statistical

- Experimental variability/error
  - Single measurements: assay variability
    - pK_i/pIC_{50} ~ 0.3 – 0.7 log units (factor of 2-5 in K_i/IC_{50})
  - Multiple replicates: mean and standard error in mean
- Statistical uncertainty in predictions
  - Standard error of prediction (assessed from validation)
    - logP ~ 0.4 -0.5 log units
    - logS ~ 0.7 – 0.8 log units
    - pK_i ~ 0.9 – 1.0 log units
  - Need to consider domain of applicability

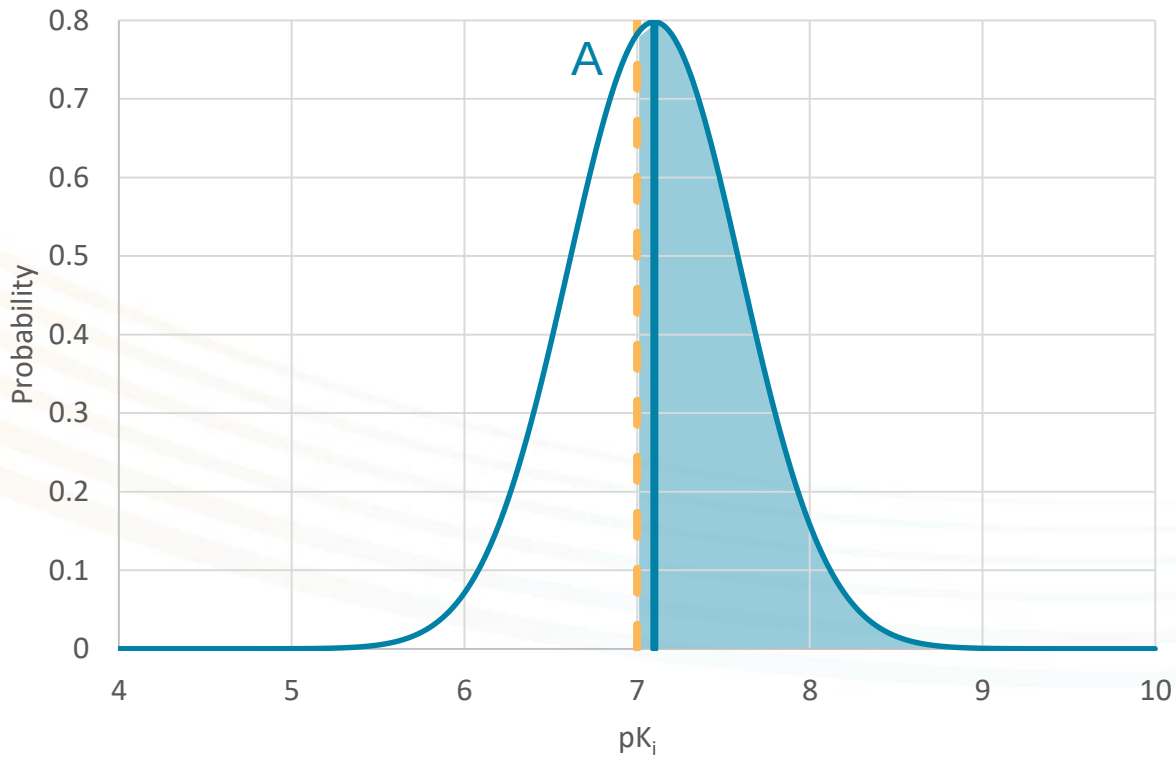
© 2018 Optibrium Ltd.

Sources of Uncertainty
Statistical

- Measured pK_i of compound A:
  $x_A = 7.1 \pm 0.5$ (1 SD)

- Does compound A meet criterion $pK_i > 7$ (better than 100 nM)?

$$X_A \sim N(x_A, \sigma^2) = N(7.1, 0.25)$$
$$\Rightarrow P(X_A > 7) = 0.58$$



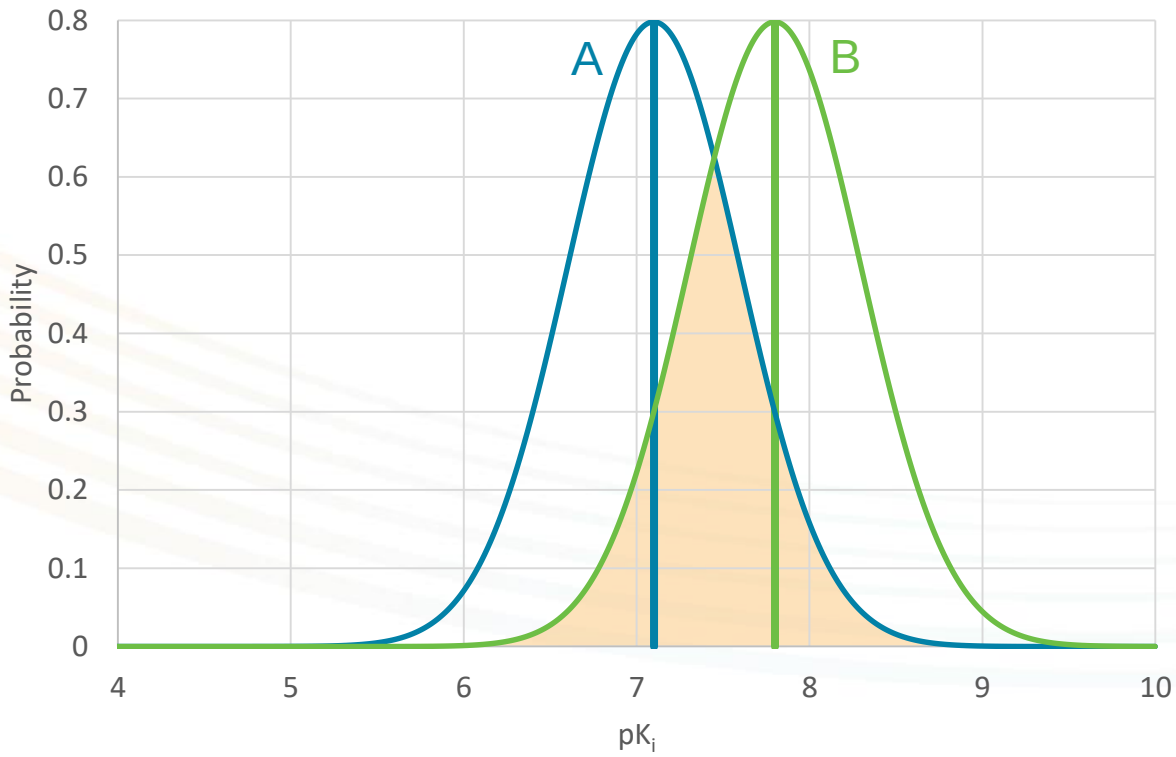
© 2018 Optibrium Ltd.

Sources of Uncertainty
Statistical

- Measured pK_i of compound B :
  $x_B = 7.8 \pm 0.5$

- Is compound B 'better' than compound A?

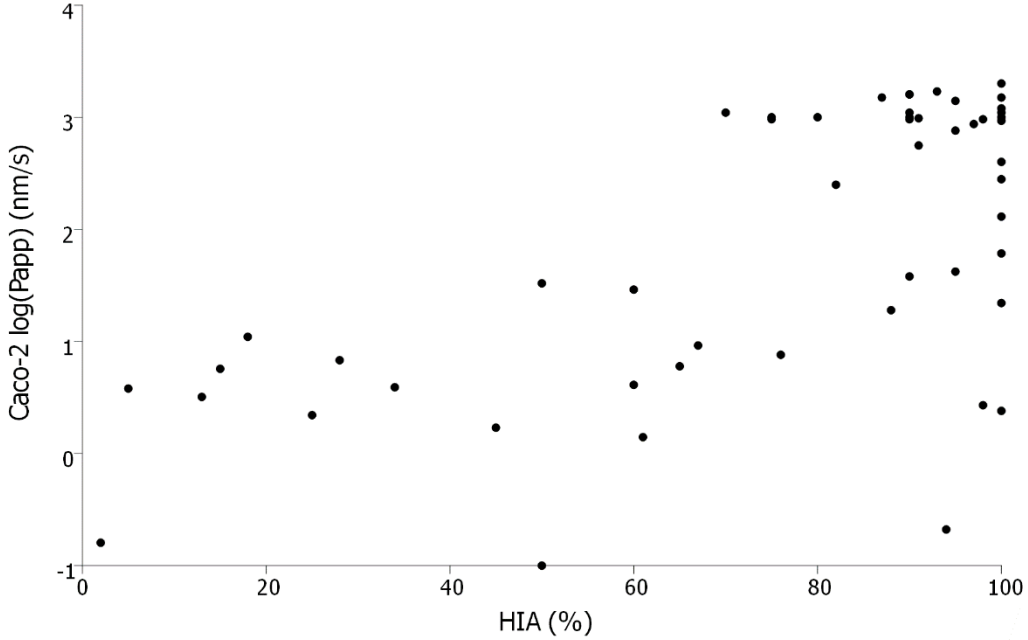
$$(X_B - X_A) \sim N(x_B - x_A, \sigma_A^2 + \sigma_B^2)$$
$$= N(0.7, 0.5)$$
$$\Rightarrow P(X_B > X_A) = 0.84$$



© 2018 Optibrium Ltd.

Sources of Uncertainty
Relevance

- All sources of data in drug discovery are **models** of the ultimate human patient
  - *In vivo*, *in vitro* or *in silico*
  - Inference/translation
- For example, Caco-2 permeation (model of absorption):

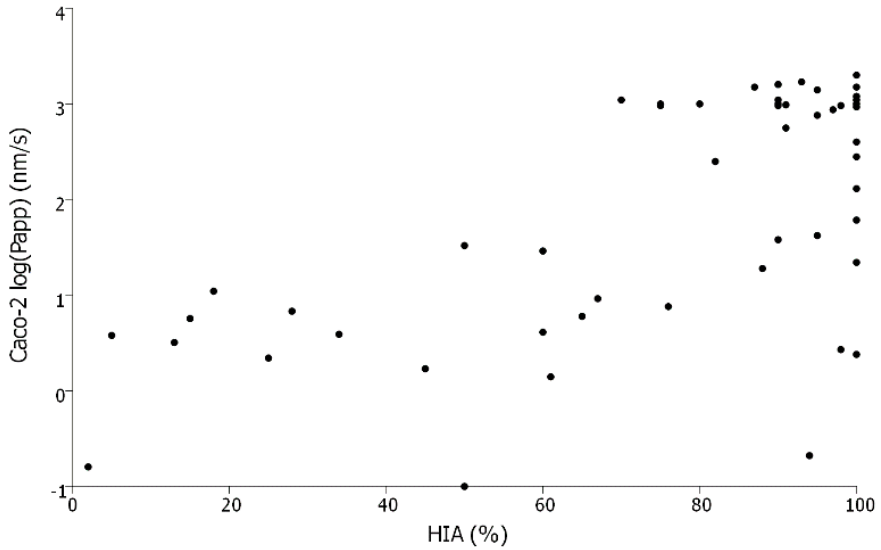


Irvine *et al.* (1999) J. Pharm. Sci. **88** pp. 28-33

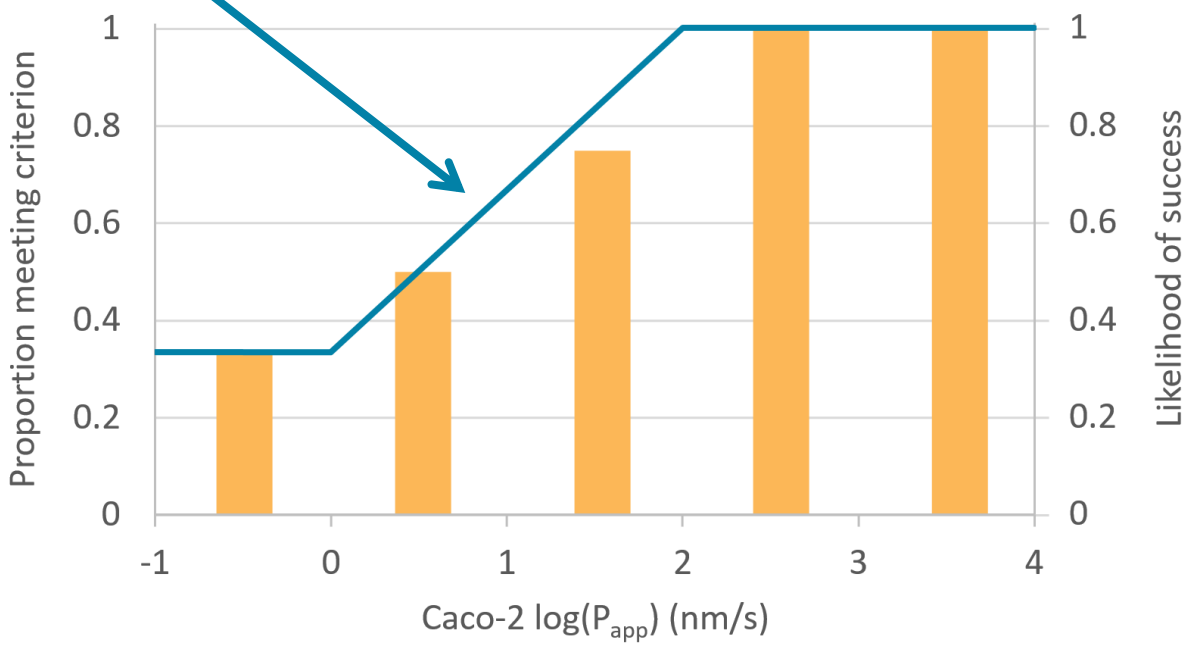
© 2018 Optibrium Ltd.

Sources of Uncertainty
Relevance

- What is the impact of data on a compound's chance of success?
  - E.g. What is chance of a compound achieving human intestinal absorption (HIA) > 50%



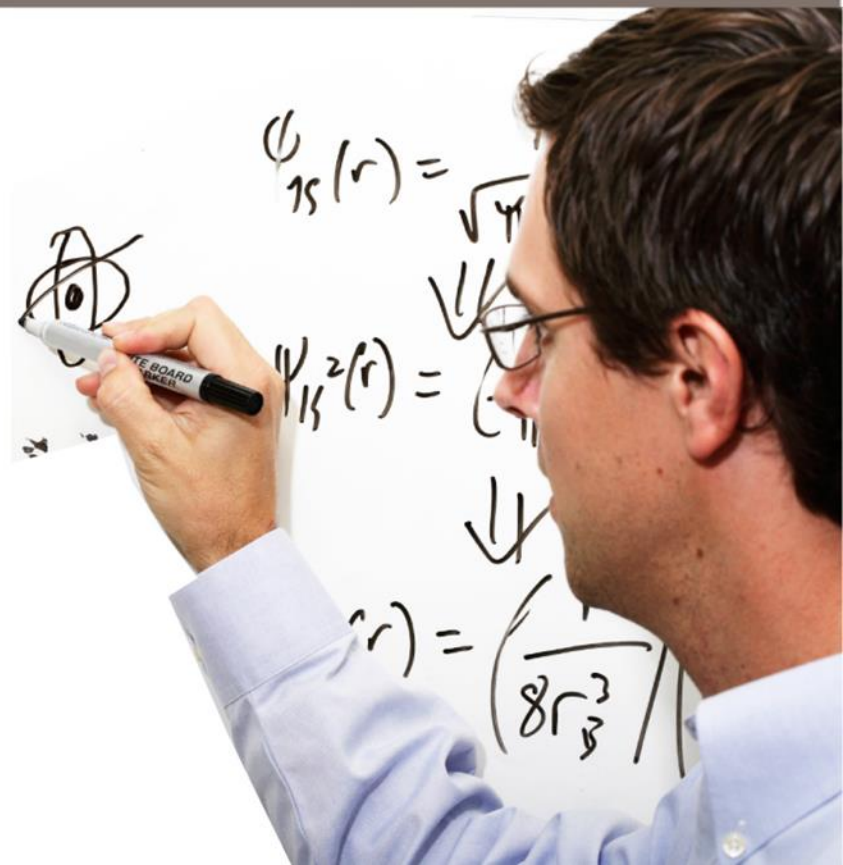
Desirability function



© 2018 Optibrium Ltd.
Segall and Champness (2015) J. Comp.-Aided Mol. Des. **29**(9), pp. 809-816

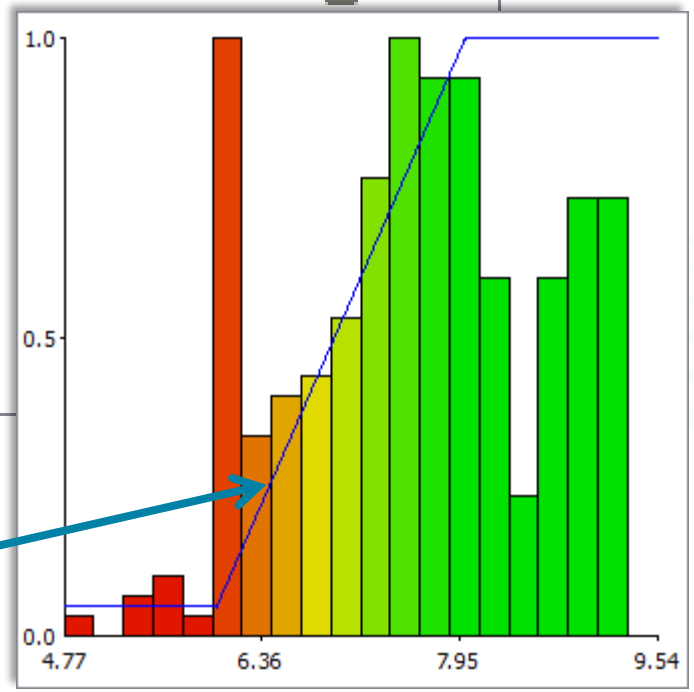
Bringing it Together

optibrium



Probabilistic Scoring
Scoring Profile

Property	Desired Value	Importance
5HT1a affinity (pKi)	8 -> inf	
logS	⩾ 1	
HIA category	+	
logP	0 -> 3.5	
BBB log([brain]:[blood])	-0.2 -> 1	
BBB category	+	
P-gp category	no	
hERG pIC50	⩽ 5	
2C9 pKi	⩽ 6	
2D6 affinity category	low medium	
PPB90 category	low	



Desirability function

© 2018 Optibrium Ltd.
Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

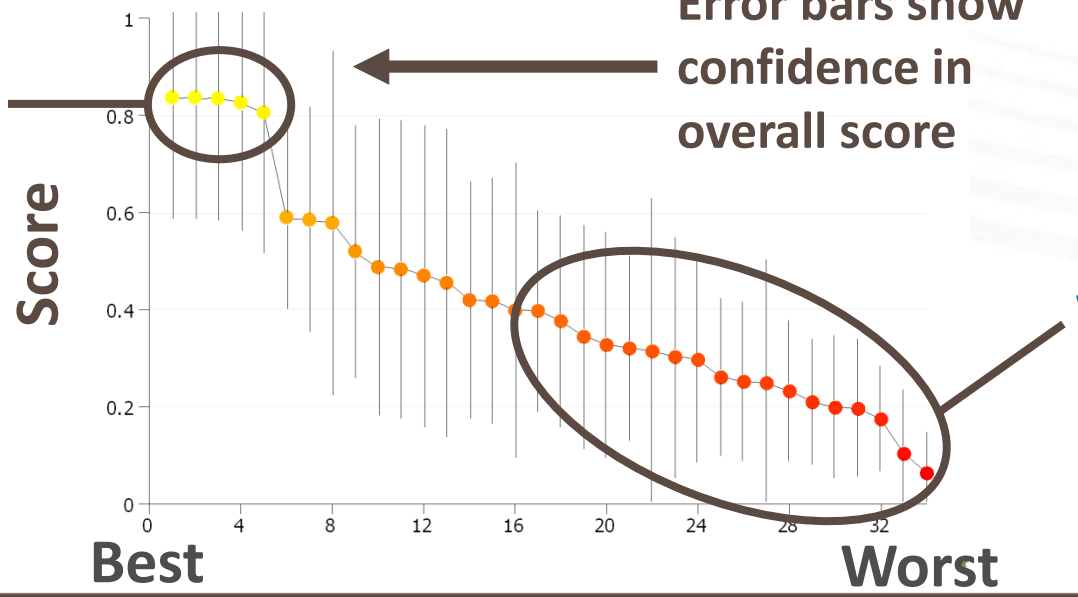
Probabilistic Scoring

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

Score (Likelihood of Success)
Confidence in score

Data do not separate these as error bars overlap

Error bars show confidence in overall score

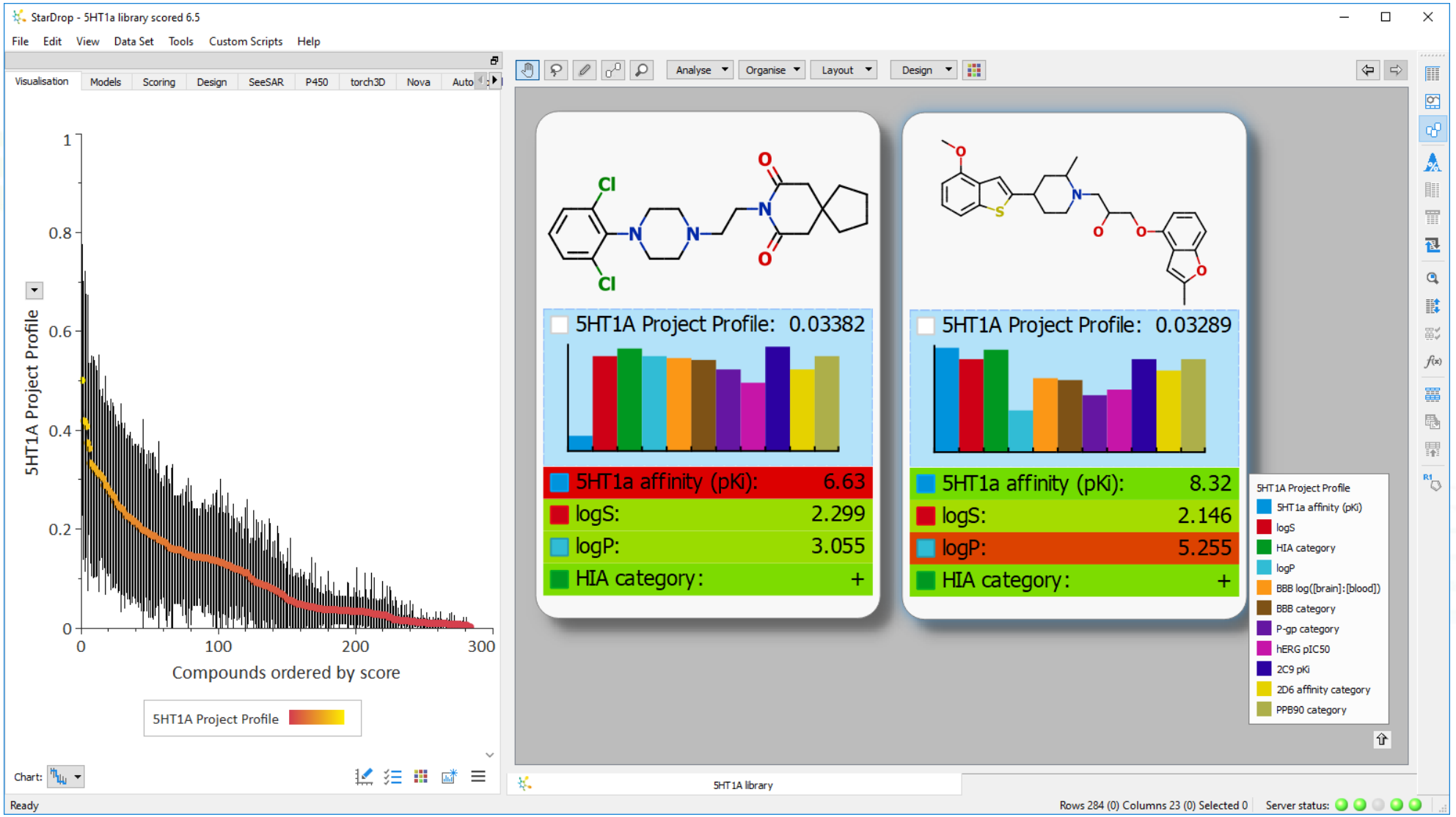


Bottom 50% may be rejected with confidence

© 2018 Optibrium Ltd.

Probabilistic Scoring

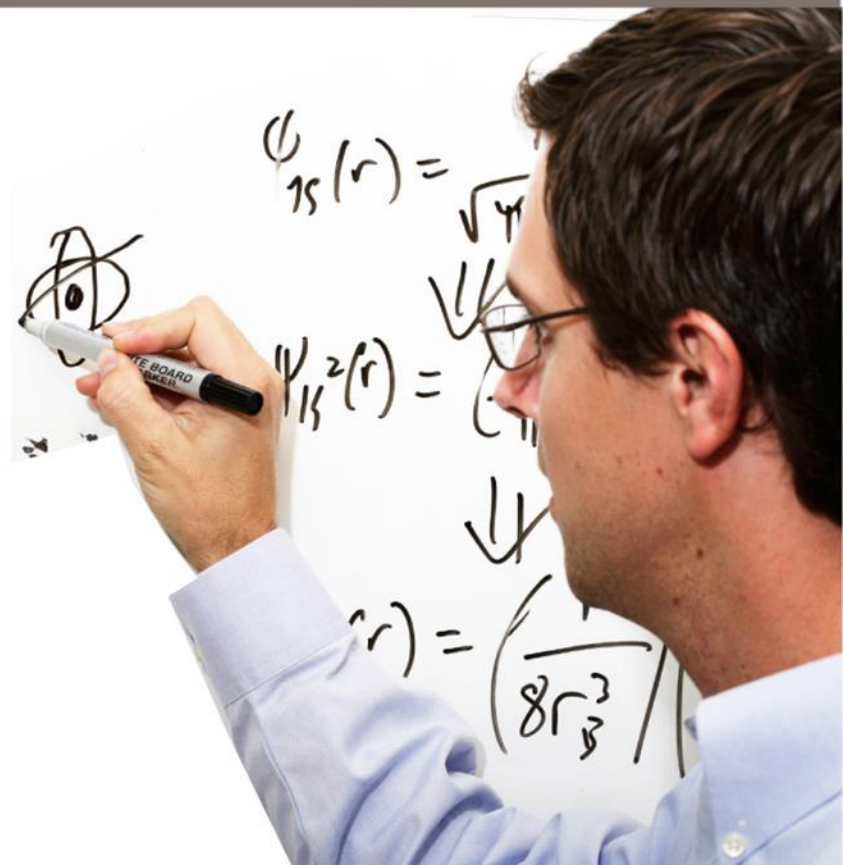
Guide redesign to improve chance of success



© 2018 Optibrium Ltd.
Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

Example

optibrium



Compound Prioritisation
Hard Cut-offs

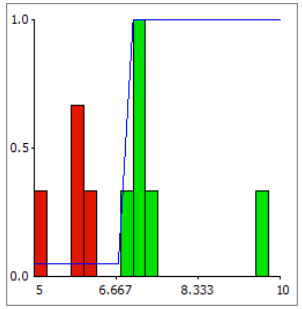
Property	Desired Value	Importance
Potency (pKi)	> 7	
log Selectivity	> 1	
log Solubility (uM)	> 2	

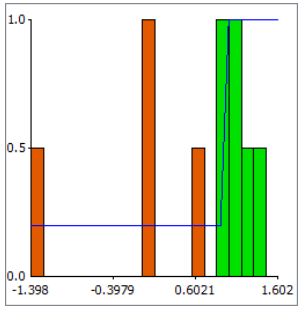
Compound	Potency (pKi)	log Selectivity	log Solubility (uM)
A	10	0.6	1.8
B	7.3	0.9	3.7
C	7.2	0	1.5
D	7	1	2
E	6.9	1.1	1
F	6	1.3	1.7
G	5.9	1	2.4
H	5.8	1.6	1.9
I	5	-1.4	3.9
J	7.1	?	3.2

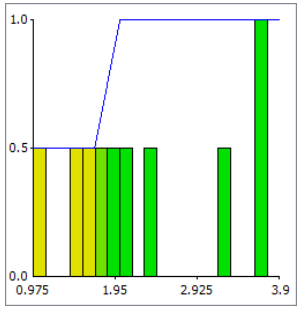
© 2018 Optibrium Ltd.

Compound Prioritisation
Desirability Functions

Property	Desired Value	Importance
Potency (pKi)	7 -> inf	
log Selectivity	1 -> inf	
log Solubility (uM)	2 -> inf	







Compound	Potency (pKi)	log Selectivity	log Solubility (uM)	Score
D	7	1	2	1
E	6.9	1.1	1	0.4
B	7.3	0.9	3.7	0.22
A	10	0.6	1.8	0.14
C	7.2	0	1.5	0.11
G	5.9	1	2.4	0.05
H	5.8	1.6	1.9	0.042
F	6	1.3	1.7	0.027
I	5	-1.4	3.9	0.01
J	7.1	?	3.2	?

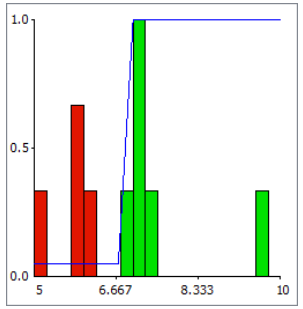
Filter

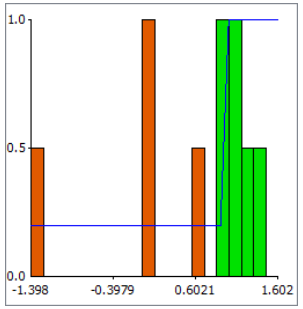
Compound	Potency (pKi)	log Selectivity	log Solubility (uM)
A	10	0.6	1.8
B	7.3	0.9	3.7
C	7.2	0	1.5
D	7	1	2
E	6.9	1.1	1
F	6	1.3	1.7
G	5.9	1	2.4
H	5.8	1.6	1.9
I	5	-1.4	3.9
J	7.1	?	3.2

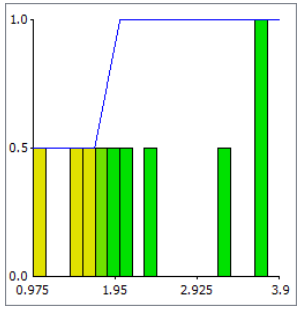
© 2018 Optibrium Ltd.
Segall and Champness (2015) J. Comp.-Aided Mol. Des. **29**(9), pp. 809-816

Compound Prioritisation
Probabilistic Scoring

Property	Desired Value	Importance
Potency (pKi)	7 -> inf	
log Selectivity	1 -> inf	
log Solubility (uM)	2 -> inf	







$$\sigma_{pKi} = 0.3, \sigma_{Sel} = 0.4, \sigma_{logS} = 0.6$$

Filter

Compound	Potency (pKi)	log Selectivity	log Solubility (uM)
A	10	0.6	1.8
B	7.3	0.9	3.7
C	7.2	0	1.5
D	7	1	2
E	6.9	1.1	1
F	6	1.3	1.7
G	5.9	1	2.4
H	5.8	1.6	1.9
I	5	-1.4	3.9
J	7.1	?	3.2

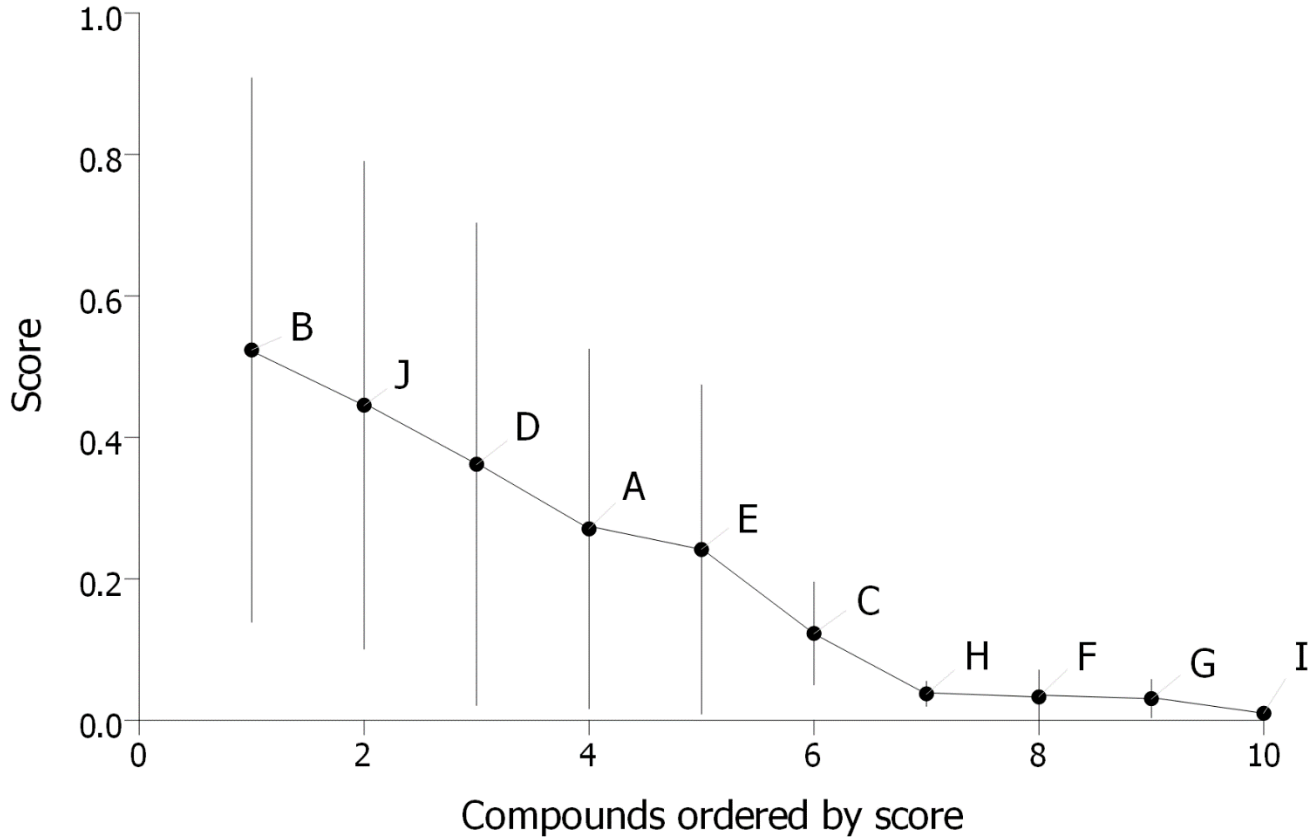
Desirability Function

Compound	Potency (pKi)	log Selectivity	log Solubility (uM)	Score
D	7	1	2	1
E	6.9	1.1	1	0.4
B	7.3	0.9	3.7	0.22
A	10	0.6	1.8	0.14
C	7.2	0	1.5	0.11
G	5.9	1	2.4	0.05
H	5.8	1.6	1.9	0.042
F	6	1.3	1.7	0.027
I	5	-1.4	3.9	0.01
J	7.1	?	3.2	?

Compound	Potency (pKi)	log Selectivity	log Solubility (uM)	Score
B	7.3	0.9	3.7	0.52
J	7.1	?	3.2	0.49
D	7	1	2	0.36
A	10	0.6	1.8	0.27
E	6.9	1.1	1	0.24
C	7.2	0	1.5	0.12
H	5.8	1.6	1.9	0.038
F	6	1.3	1.7	0.033
G	5.9	1	2.4	0.031
I	5	-1.4	3.9	0.01

© 2018 Optibrium Ltd.
Segall and Champness (2015) J. Comp.-Aided Mol. Des. **29**(9), pp. 809-816

Score Distribution

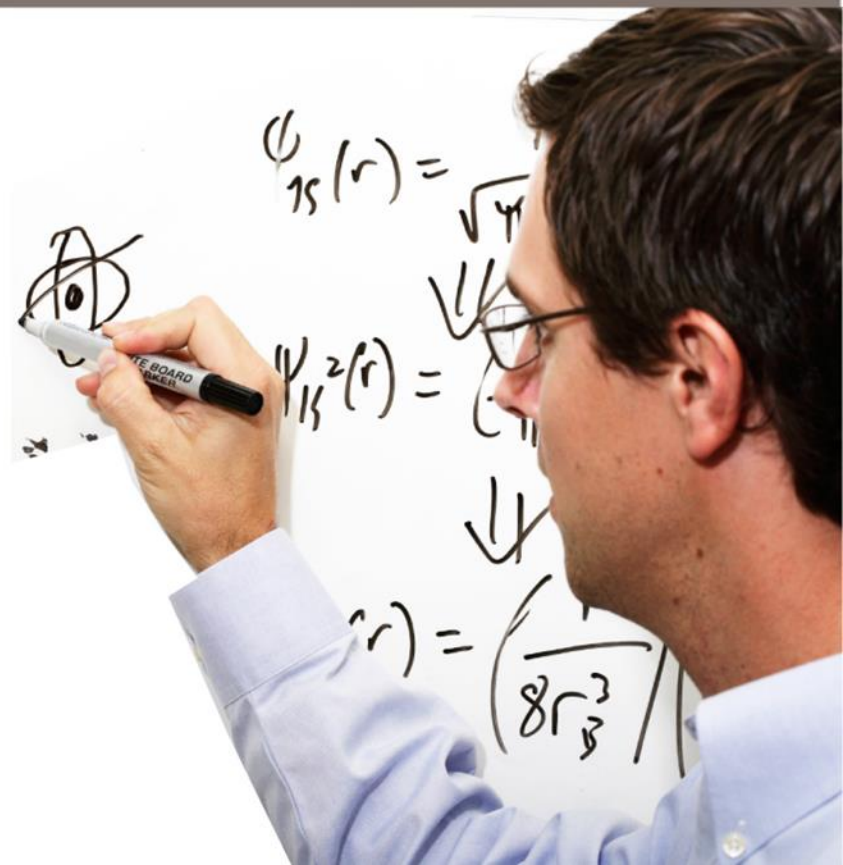


Conclusion: Only compounds H, F, G and I can be confidently rejected…

© 2018 Optibrium Ltd.

How Good to Our Models Have to Be?

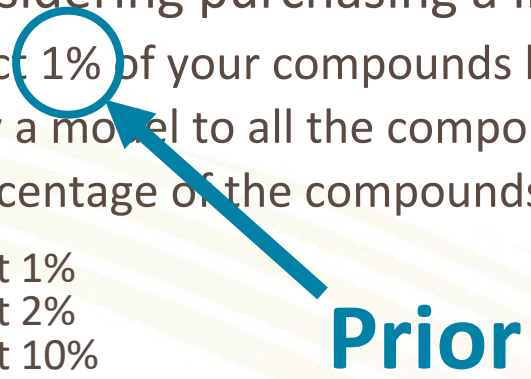
optibrium



How Well Does This Model Conserve Your Options?

- You are considering purchasing a library of compounds:
  - You expect 1% of your compounds have a particular kind of toxicity
  - You apply a model to all the compounds that is 90% reliable (both 90% sensitive and 90% specific)
  - What percentage of the compounds that are predicted to fail, genuinely have the toxicity?

    a) About 1%
    b) About 2%
    c) About 10%
    d) About 50%
    e) About 90%



- Answer?
  - c) Of 1000 compounds, 990 x 0.1 + 10 x 0.9 = 108 would be reported as toxic by the model, of which only 9 really are toxic.

- Easy to overreact to negative results
  - Availability bias (neglect of the prior)*

© 2018 Optibrium Ltd.
*Chadwick and Segall (2010), Drug Discov. Today, **15**(13/14), pp. 561-9

Example Application
Screening Strategy

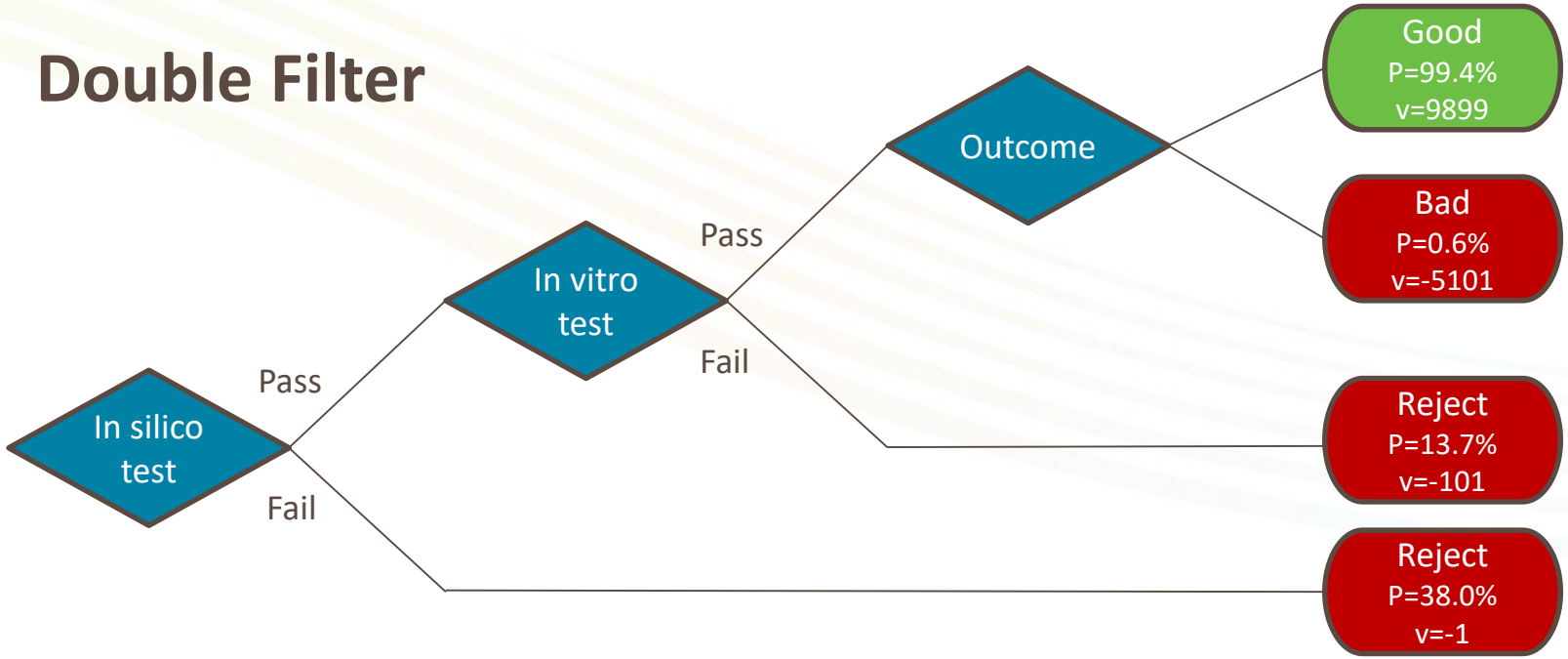
- Two screens for property: *in silico* and *in vitro*
  - *In silico*: cost 1, accuracy 80%
  - *In vitro*: cost 100, accuracy 95%
  - Cost to prove *in vivo* 5,000
  - Net value of good compound 10,000
- 5 Possible screening strategies

© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Two screens for property: *in silico* and *in vitro*

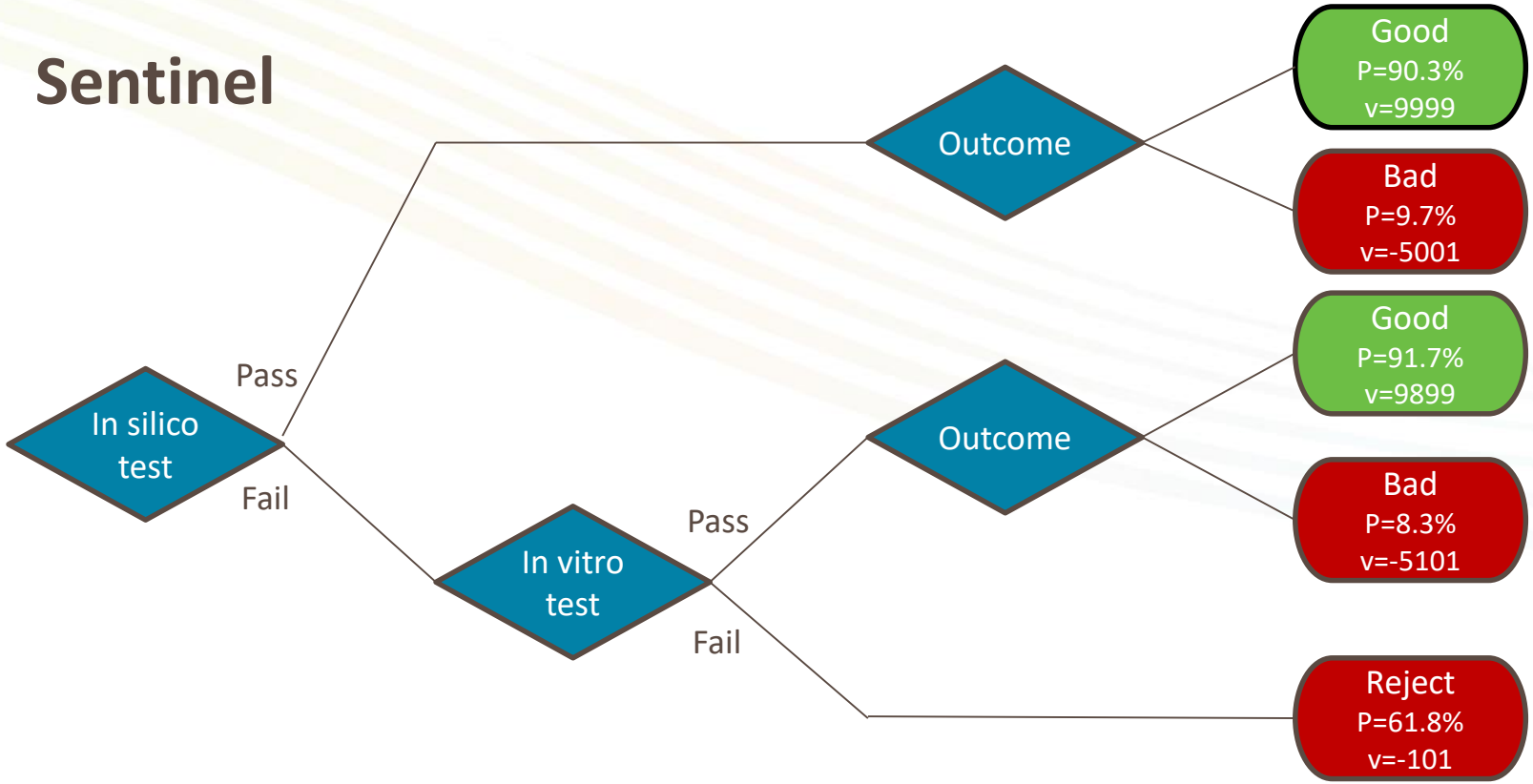
- 5 Possible screening strategies:



© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Two screens for property: *in silico* and *in vitro*
- 5 Possible screening strategies:



© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Two screens for property: *in silico* and *in vitro*

- 5 Possible screening strategies:

In Silico Only

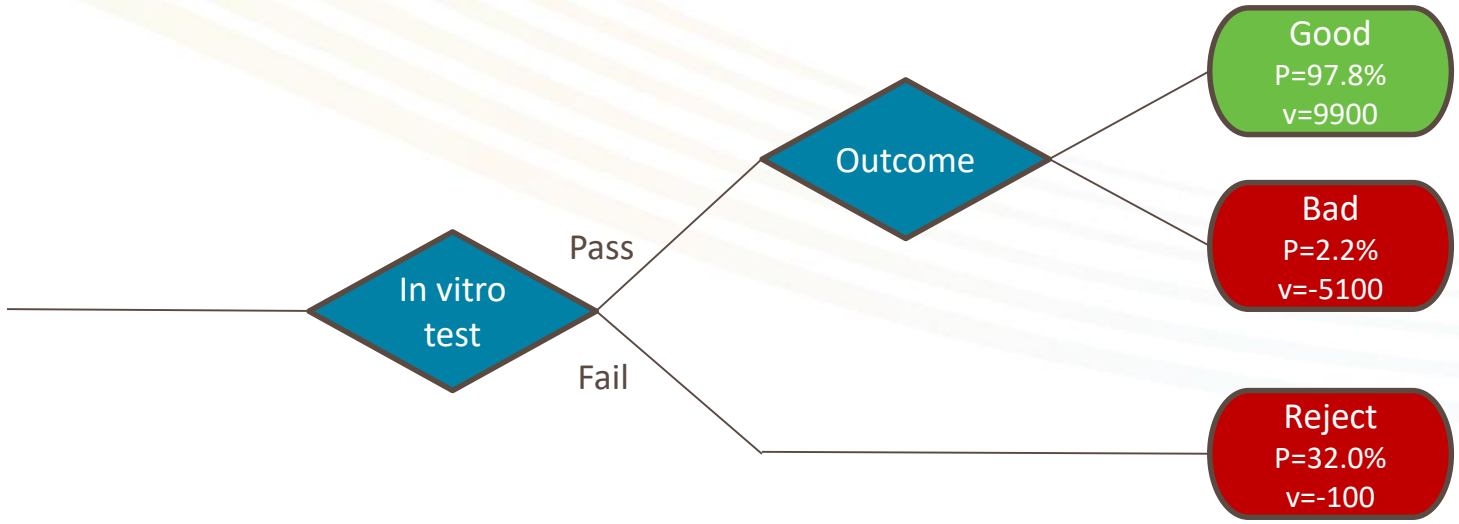
© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Two screens for property: *in silico* and *in vitro*

- 5 Possible screening strategies:

In Vitro Only



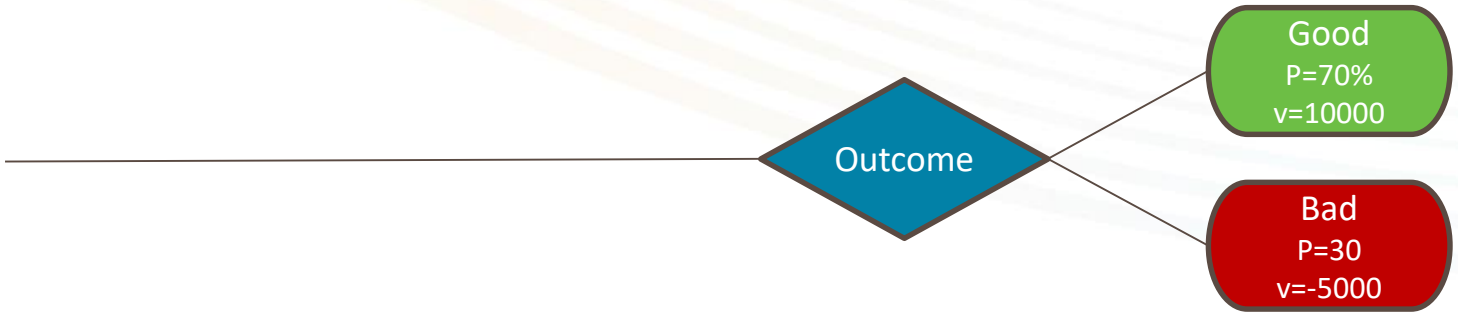
© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Two screens for property: *in silico* and *in vitro*

- 5 Possible screening strategies:

No Screen



© 2018 Optibrium Ltd.

Example Application
Screening Strategy

- Parameters:
  - *In silico*: cost 1, accuracy 80%
  - *In vitro*: cost 100, accuracy 95%
  - Cost to confirm 5,000; Net value of good compound 10,000

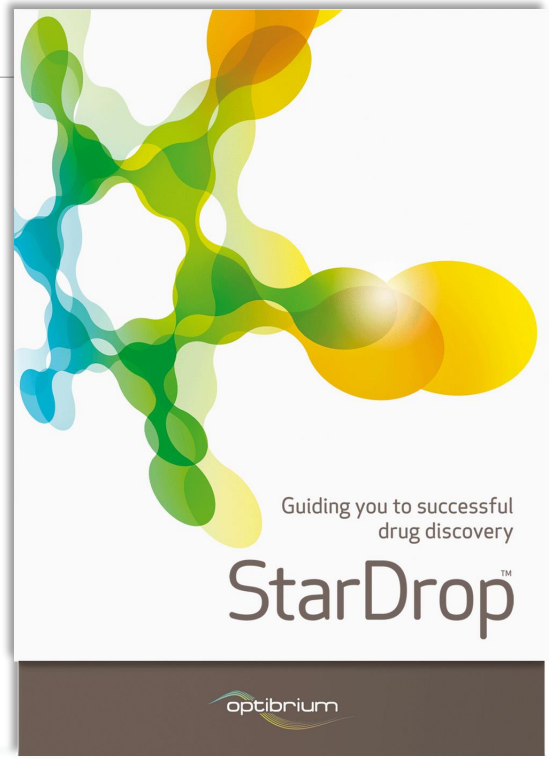
Strategy	Value (Prior for risk 30%)	Value (Prior for risk 40%)
Double filter	5242	4483
Sentinel	6531	5415
In silico only	5299	4399
In vitro only	6475	5500
No screen	5500	4000

Interactive example http://www.tessella.com/screening-strategy-explorer

© 2018 Optibrium Ltd.

Conclusions

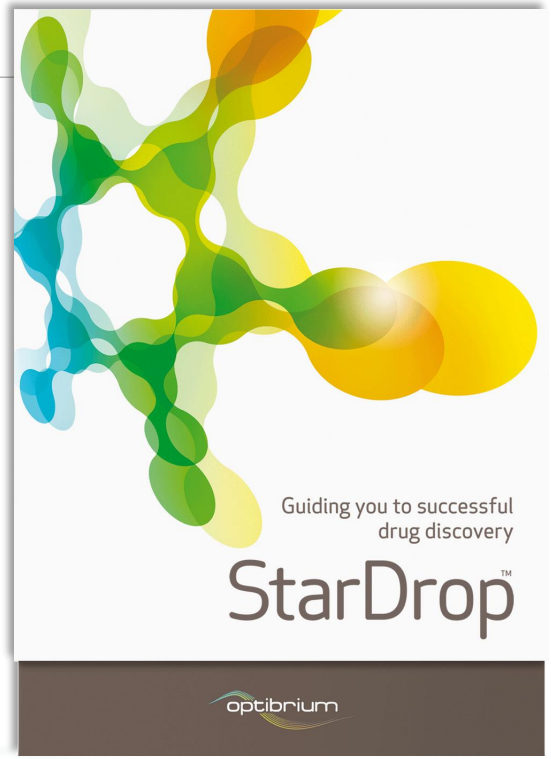
- Data only add value when used to make good decisions in the context of a discovery project
- The value of data can only be assessed when we understand its confidence
  - Avoid wasted effort and missed opportunities
- We can only know if our assays/models add value when we know the priors for the risks we are addressing
  - **BIG opportunity for pre-competitive collaboration**
- For more information and references, please visit:
  - www.optibrium.com/stardrop/
  - www.optibrium.com/community/



© 2018 Optibrium Ltd.

Acknowledgements

- The Optibrium team, in particular Ed Champness
- Andrew Chadwick (formerly of Tessella)



© 2018 Optibrium Ltd.