optibrium

# The Challenges of Decision Making Using Uncertain Data

**ACS Fall National Meeting, August 2014**

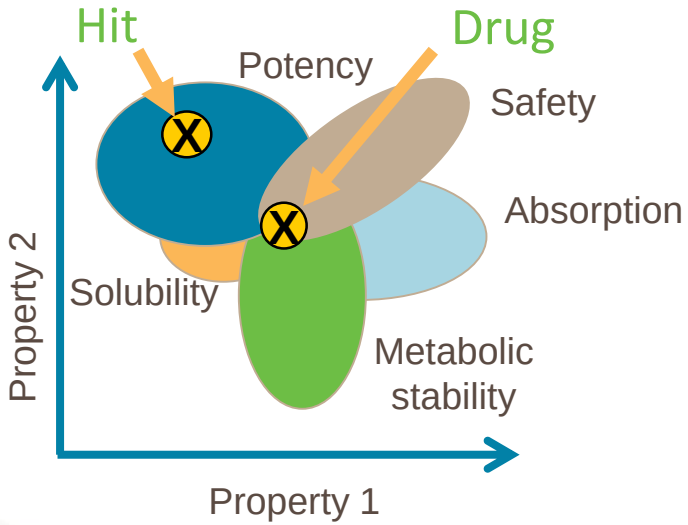
**Edmund Champness**

# Overview

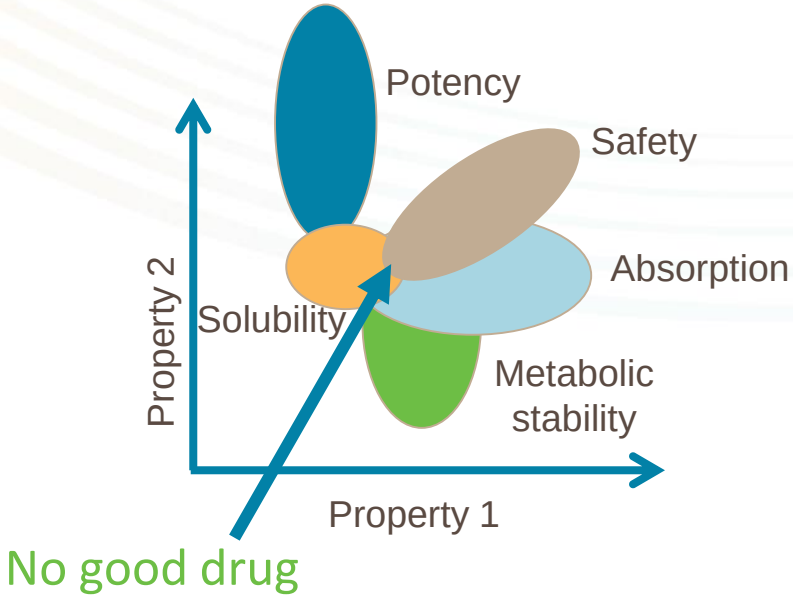
- Drug Discovery - what we'd like to happen…
- Challenges
  - Uncertain data
  - Missing data
- Putting it all together (MPO)
- Case Study
- Conclusions

# Drug Discovery:
## What we'd like to happen…

- 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**

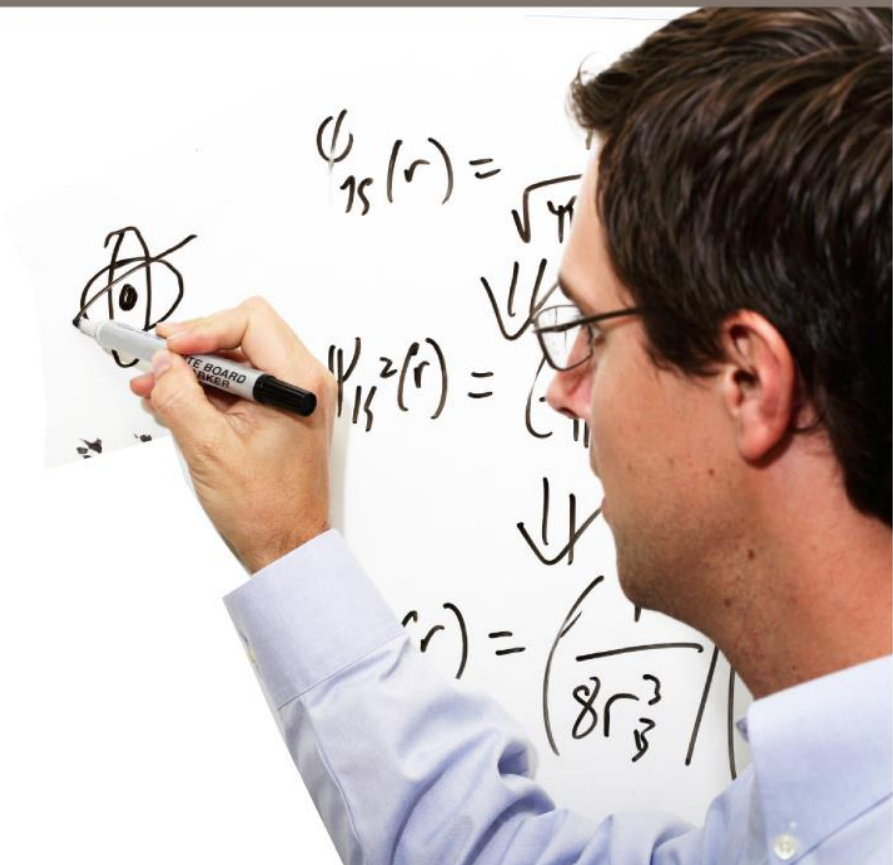




*M.D. Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

# The Challenges

optibrium



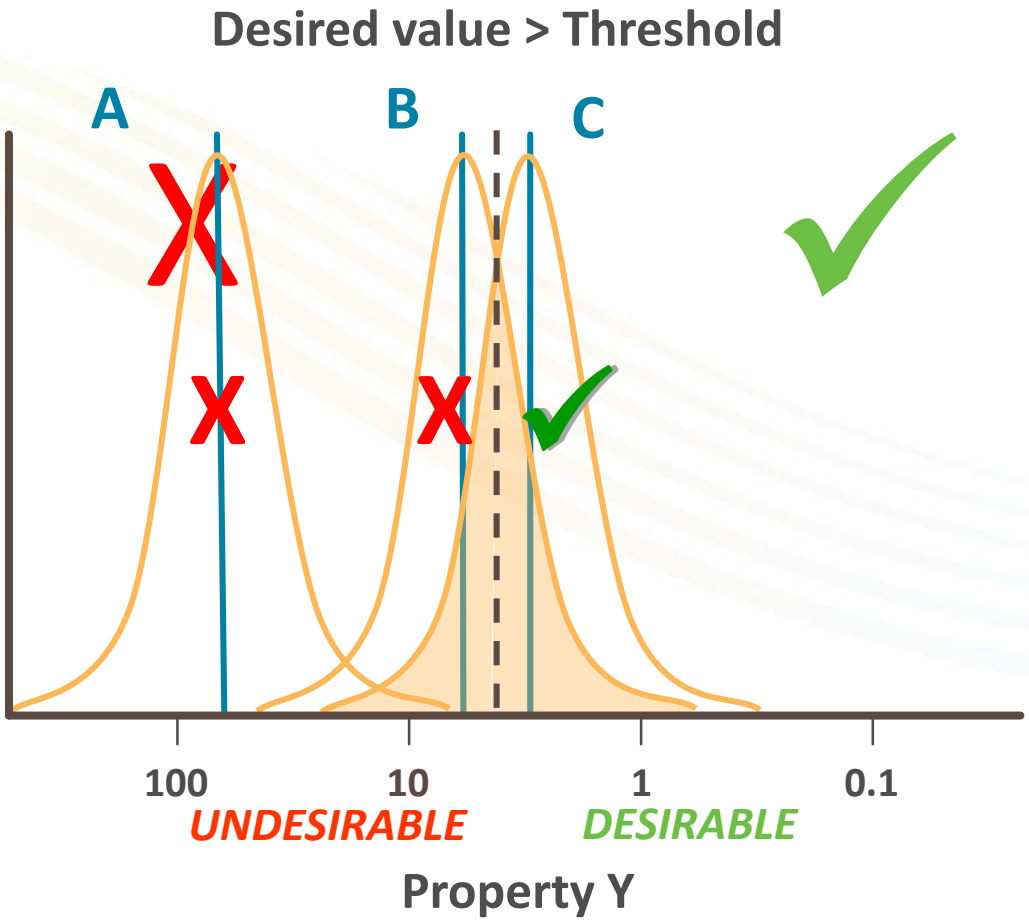
# The Challenges:
## Uncertain data

- What's certain?
  - We know some simple properties of our compounds
- What's not so certain?
  - *In vitro*/*In vivo* measurements
    - experimental variability
    - inference/translation (modelling…statistical error)
  - *In silico* predictions
    - statistical error

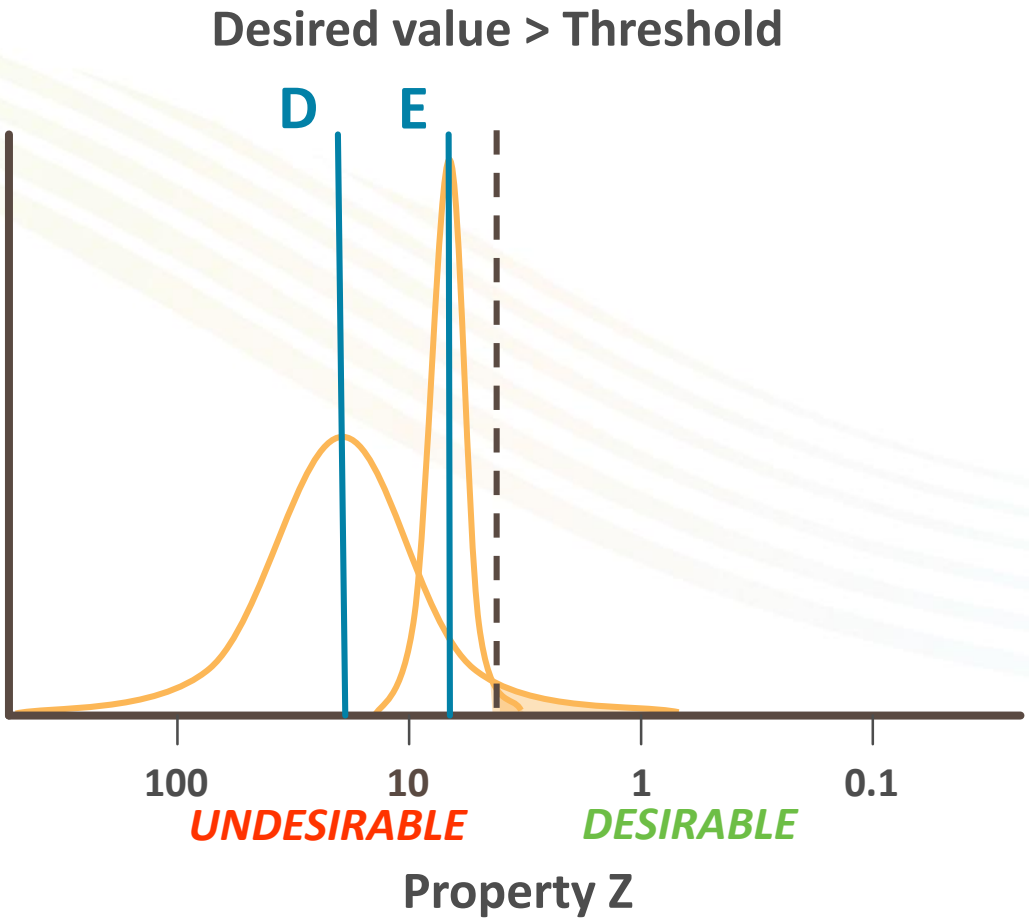
# The Challenges:
## Uncertain data

- So what does that mean…
- A good RMSE for logS (solubility) is 0.6
- Assuming normal distribution this means that when I have logS value of 2 (that's 100uM) then
  - 68% of the time this represents an actual value between 1.4 and 2.6 (25uM to 400uM)
  - 95% of the time this represents an actual value between 0.8 and 3.2 (6um to 1.6mM)
  - 99% of the time this represents an actual value between 0.2 and 3.8 (1.6uM to 6.3mM)

# Importance of Uncertainty



# Importance of Uncertainty



# The Challenges:
## Missing data

- What can I do when I don't know a property value?
  - Infer a value from other known values (*in silico* prediction), assuming:
    - The property is not too complex to model
    - We have enough data
    - The data we have are not too biased
    - …

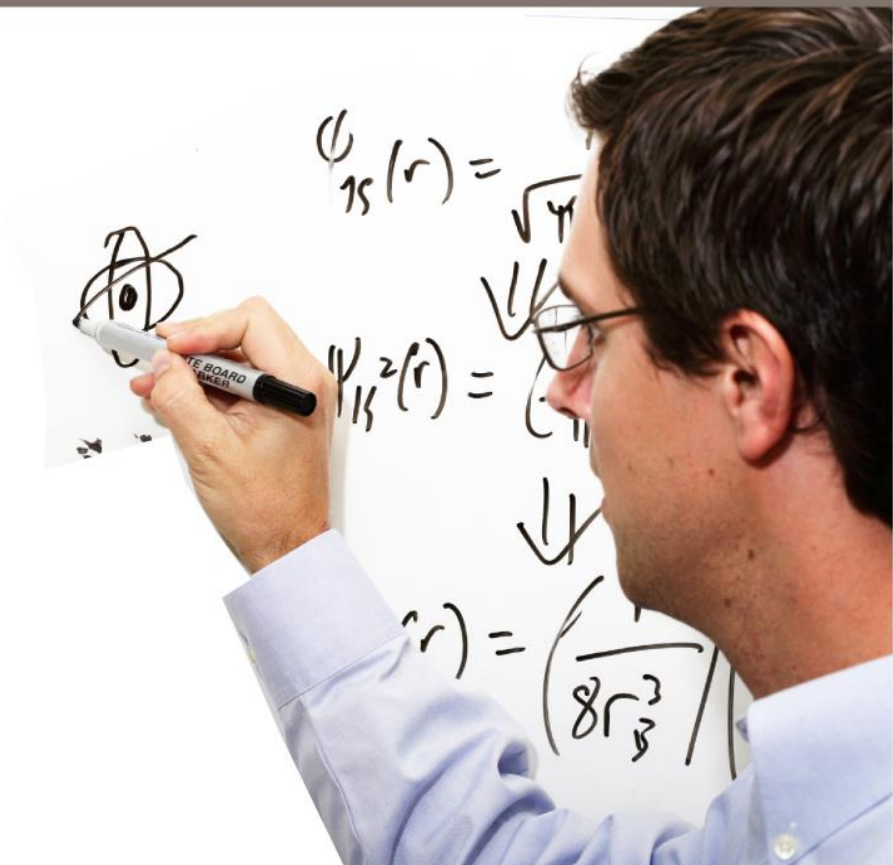
    However, if we can come up with model then we have to remember that this will have statistical error which we need to take into consideration
  - Treat it as a true unknown

# The Challenges:
## …and one more thing

- We probably have quite a few properties we need to optimise!
  - Each will have their own uncertainty or missing values
  - Each will have its own criteria we'd like to achieve
  - Each will have its own level of importance relative to the other properties

# Putting it all together (Multi-parameter Optimisation)

optibrium



# Putting it all together (MPO):
## A simple example

- 3 properties
  - logS – RMSE 0.6
  - Potency (Ki) – 2 fold
  - Selectivity – 2.6 fold
- What would I like to see?
  - logS > 2
  - Potency (Ki) < 100nM
  - Selectivity > 10

- 10 compounds

| | logS | Potency (Ki nM) | Selectivity |
|---|---|---|---|
| #1 | 1.8 | 0.1 | 4 |
| #2 | 3.7 | 50 | 5 |
| #3 | 1.5 | 60 | 1 |
| #4 | 2.0 | 100 | 10 |
| #5 | 1.0 | 120 | 12 |
| #6 | 1.7 | 900 | 20 |
| #7 | 2.4 | 1200 | 10.5 |
| #8 | 1.9 | 1500 | 40 |
| #9 | 3.9 | 10000 | 0.04 |
| #10 | 3.2 | ? | 9.8 |

So which is best?

# Putting it all together (MPO):
## Filtering

- 3 properties
  - logS – RMSE 0.6
  - Potency (Ki) – 2 fold
  - Selectivity – 2.6 fold
- What would I like to see?
  - logS > 2
  - Potency (Ki) < 100nM
  - Selectivity > 10

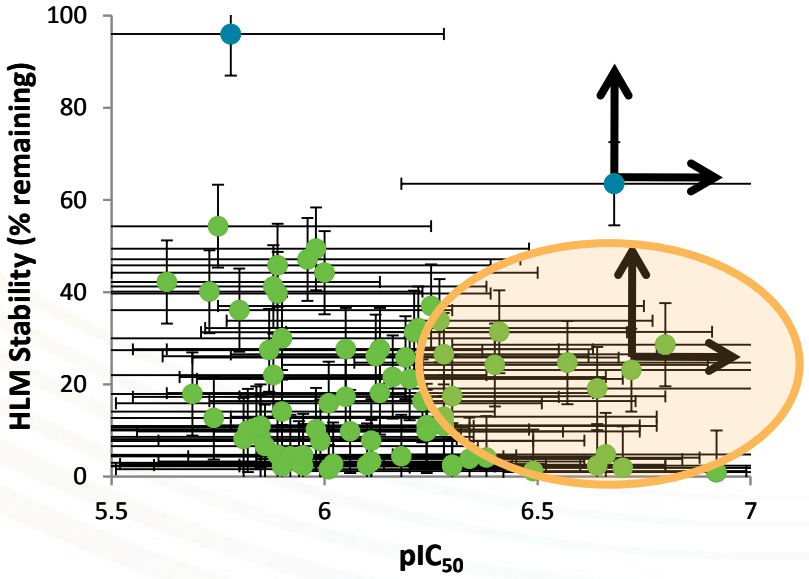
- 10 compounds

| | logS | Potency (Ki nM) | Selectivity |
|---|---|---|---|
| #1 | | | |
| #2 | | | |
| #3 | | | |
| #4 | | | |
| #5 | | | |
| #6 | | | |
| #7 | | | |
| #8 | | | |
| #9 | | | |
| #10 | | | |

None?

# Putting it all together (MPO):
## Pareto Optimisation

- Not one optimum, but many
  - Explore different balances
  - Pareto front
- Advantages
  - Very good if appropriate balance is unknown *a priori*
  - Flexible
  - Easy to interpret
- Disadvantages
  - Overwhelmed by large numbers of parameters (>5)
  - Uncertainty?



# Putting it all together (MPO):
## Pareto Optimisation

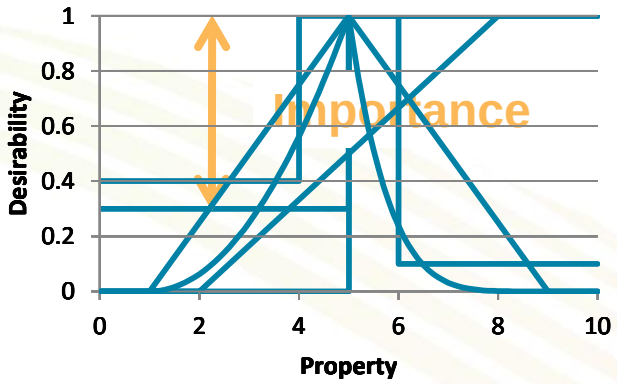
- 3 properties
  - logS – RMSE 0.6
  - Potency (Ki) – 2 fold
  - Selectivity – 2.6 fold
- What would I like to see?
  - logS > 2
  - Potency (Ki) < 100nM
  - Selectivity > 10

- 10 compounds

| | logS | Potency (Ki nM) | Selectivity |
|---|---|---|---|
| #1 | 1.8 | 0.1 | 4 |
| #2 | | | |
| #3 | | | |
| #4 | | | |
| #5 | | | |
| #6 | | | |
| #7 | | | |
| #8 | 1.9 | 1500 | 40 |
| #9 | 3.9 | 10000 | 0.04 |
| #10 | | | |

# Putting it all together (MPO):
## Desirability Functions*

- Relate property values to how 'desirable' the outcome




**(Derringer Function)**

- Combine multiple properties into 'desirability index'
  - Additive: $D = \frac{d_1(Y_1) + d_2(Y_2) + \cdots + d_n(Y_n)}{n}$
  - Multiplicative: $D = (d_1(Y_1) \times d_2(Y_2) \times \ldots \times d_n(Y_n))^{1/n}$
- Strengths
  - Very flexible; Explicitly weight properties; Easy to interpret
- Caveats
  - No explicit consideration of uncertainty; Need to know criteria *a priori*



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

# Putting it all together (MPO):
## Desirability Functions

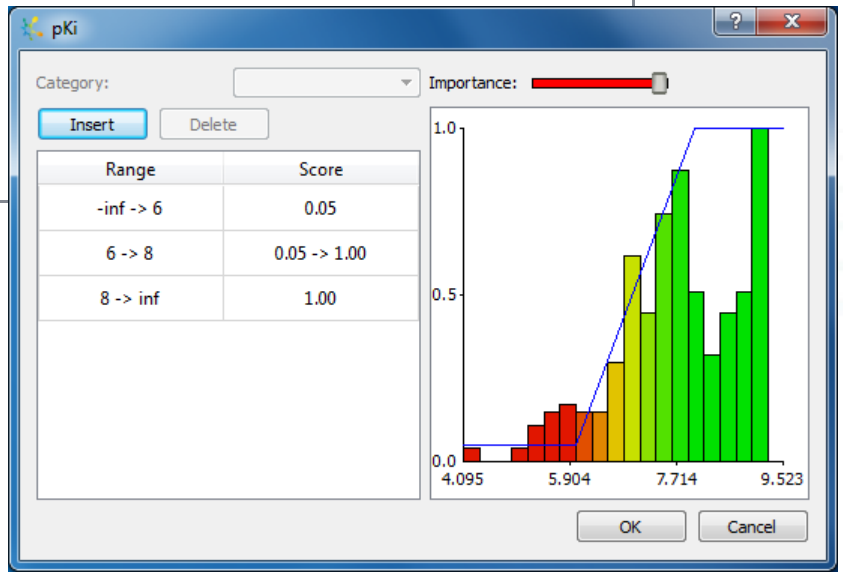
- 3 properties
  - logS – RMSE 0.6
  - Potency (Ki) – 2 fold
  - Selectivity – 2.6 fold
- What would I like to see?
  - logS > 2
  - Potency (Ki) < 100nM
  - Selectivity > 10

- 10 compounds

| | logS | Potency (Ki nM) | Selectivity |
|---|---|---|---|
| #1 | | | |
| #2 | 3.7 | 50 | 5 |
| #3 | | | |
| #4 | 2.0 | 100 | 10 |
| #5 | 1.0 | 120 | 12 |
| #6 | 1.7 | 900 | 20 |
| #7 | 2.4 | 1200 | 10.5 |
| #8 | | | |
| #9 | | | |
| #10 | | | |

# Putting it all together (MPO):
## Probabilistic Scoring* – Scoring Profile

| Property | Desired Value | Importance |
|---|---|---|
| pKi | > 8 | |
| logS | > 1 | |
| BBB log([brain]:[blood]) | -0.2 -> 1 | |
| Cytotox. (pLD50) | ⩽ 5 | |
| logP | 0 -> 3.5 | |
| P-gp category | no | |
| Ames | - | |
| hERG pIC50 | ⩽ 5 | |
| 2C9 pKi | ⩽ 6 | |
| 2D6 affinity category | low medium | |






* Segall *et al.* (2009) Chem. & Biodiv. **6** p. 2144

# Multi-parameter Optimisation
## Probabilistic Scoring*

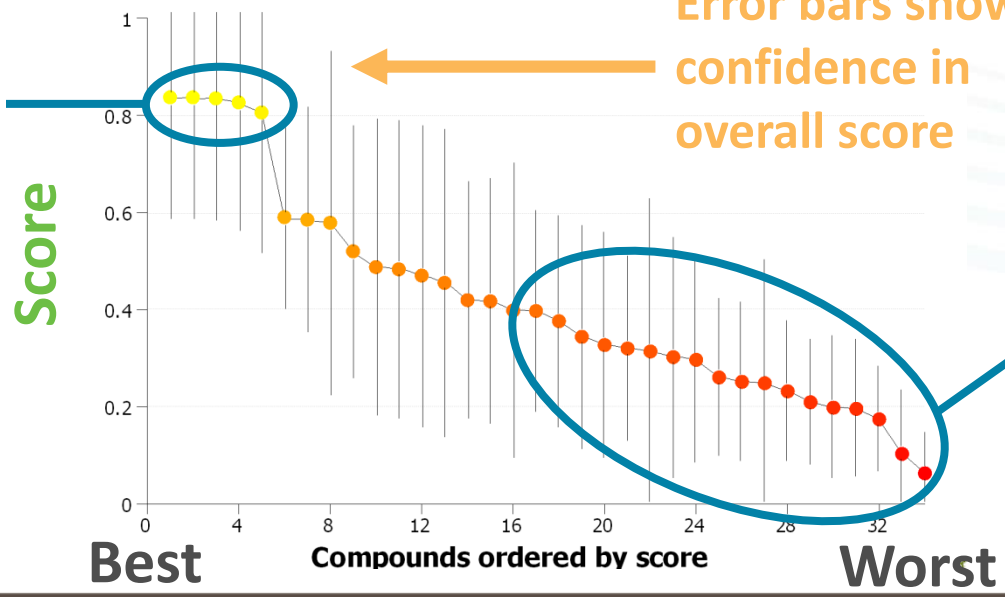
- **Property data**
  - Experimental or predicted
- **Criteria for success**
  - Relative importance
- **Uncertainties in data**
  - Experimental or statistical

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

**Data do not separate these as error bars overlap**

**Error bars show confidence in overall score**

**Bottom 50% may be rejected with confidence**






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

# Putting it all together (MPO):
## Probabilistic Scoring (equal weighting)

- 3 properties
  - logS – RMSE 0.6
  - Potency (Ki) – 2 fold
  - Selectivity – 2.6 fold
- What would I like to see?
  - logS > 2
  - Potency (Ki) < 100nM
  - Selectivity > 10

- 10 compounds

| | logS | Potency (Ki nM) | Selectivity |
|---|---|---|---|
| #1 | 1.8 | 0.1 | 4 |
| #2 | 3.7 | 50 | 5 |
| #3 | | | |
| #4 | 2.0 | 100 | 10 |
| #5 | | | |
| #6 | | | |
| #7 | 2.4 | 1200 | 10.5 |
| #8 | 1.9 | 1500 | 40 |
| #9 | 3.9 | 10000 | 0.04 |
| #10 | 3.2 | ? | 9.8 |

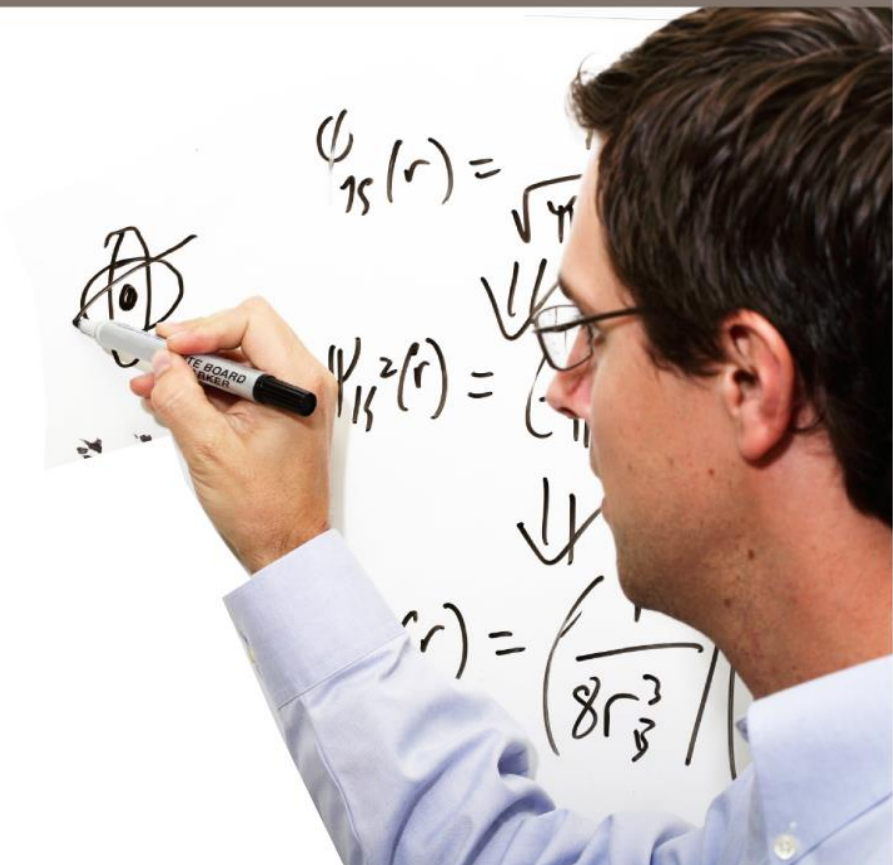
# Putting it all together (MPO): Probabilistic Scoring (equal weighting)

# Putting it all together (MPO):
# …the wrong approach could lead us astray!

| | logS | Potency (Ki nM) | Selectivity | Filter | Pareto | Desirabiity | Probabilistic Scoring |
|---|---|---|---|---|---|---|---|
| #1 | 1.8 | 0.1 | 4 | | Y | | 4 |
| #2 | 3.7 | 50 | 0.2 | | | Y | 1 |
| #3 | 1.5 | 60 | 1 | | | | 10 |
| #4 | 2.0 | 100 | 10 | ? | | Y | 3 |
| #5 | 1.0 | 120 | 12 | | | Y | 9 |
| #6 | 1.7 | 900 | 20 | | | Y | 8 |
| #7 | 2.4 | 1200 | 10.5 | | | Y | 6 |
| #8 | 1.9 | 1500 | 40 | | Y | | 7 |
| #9 | 0.7 | 10000 | 0.04 | | Y | | 5 |
| #10 | 3.2 | ? | 9.8 | | | | 2 |

# Case Study
## Balancing Properties in Lead Optimization

optibrium



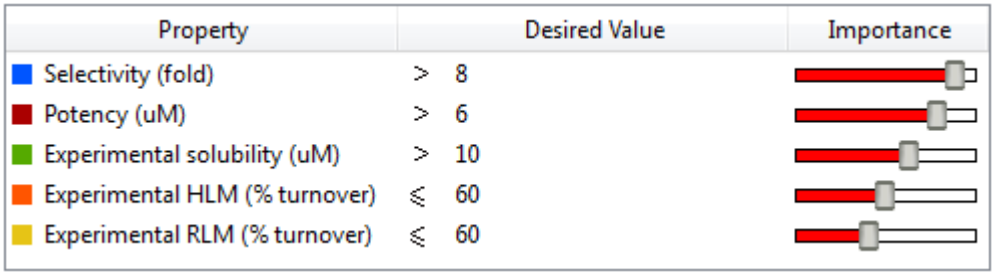
# Objective

- *In vitro* potency, selectivity, solubility and microsomal stability data had been generated for a set of 150 client compounds

- Compounds had previously been selected for *in vivo* study based on selectivity and potency, ignoring potential solubility and metabolic stability problems, resulting in poor bioavailability in rats

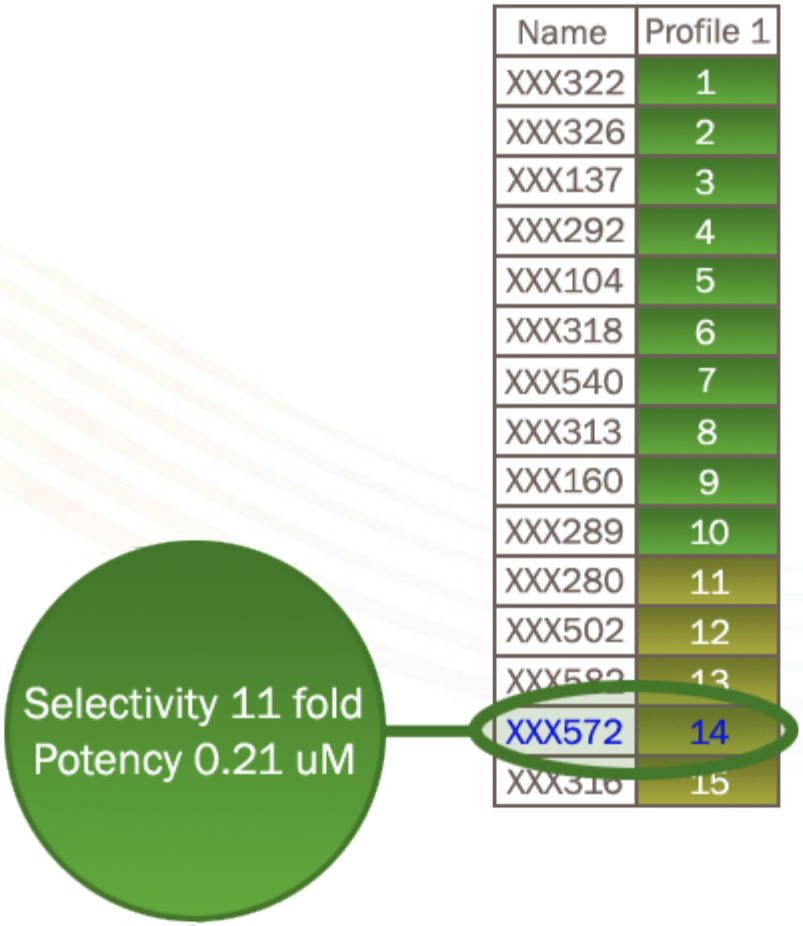
**Select compounds with a balanced set of properties for progression *in vivo***

Project Scoring Profile:



# Profile 1: Potency and Selectivity Only

- Historically, compounds were filtered and ranked on the basis of their selectivity and potency alone, with a bias towards selectivity
- This approach did not take into account the errors and uncertainties in the experiments
- The table on the right shows the top 15 compounds when ranked by this method
- Highlighted is compound XXX572, which was neither the most selective nor the most potent compound in the set (its relative position in the rankings will be followed throughout this case study)



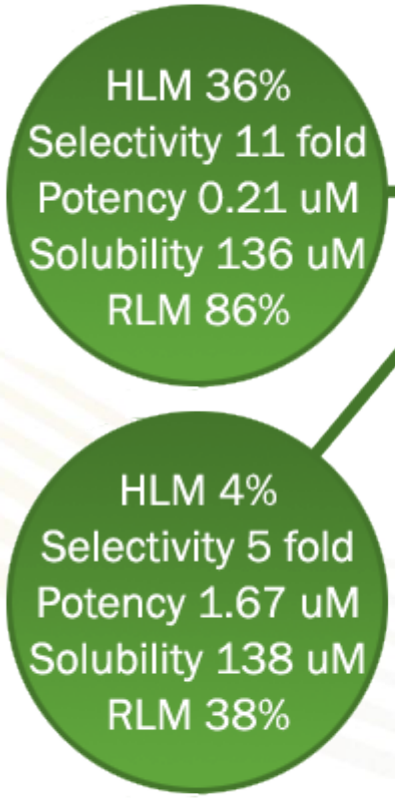
# Profile 2: Factoring in Uncertainty

- Estimates were made of the experimental uncertainties in the assays and the compounds rescored
- Some compounds, now ranked according to Profile 2, shifted significantly in rank
- Compound XXX561 jumped from 28th to 12th position (it was extremely potent but had previously “failed” the selectivity cut-off of 8-fold, despite the uncertainties in the selectivity measurement which meant there was a relatively high probability that its true selectivity was in excess of this!)


| Name | Profile 2 | Profile 1 |
|---|---|---|
| XXX326 | 1 | 2 |
| XXX137 | 2 | 3 |
| XXX322 | 3 | 1 |
| XXX313 | 4 | 8 |
| XXX540 | 5 | 7 |
| XXX160 | 6 | 9 |
| XXX572 | 7 | 14 |
| XXX104 | 8 | 5 |
| XXX292 | 9 | 4 |
| XXX541 | 10 | 22 |
| XXX561 | 12 | 28 |
| XXX537 | 14 | 17 |
| XXX502 | 15 | 12 |

# Profile 3: All Available *In Vitro* Data

- Finally, the compounds were scored taking into account all of the *in vitro* data along with accompanying statistics relating to experimental uncertainties
- This gave a considerable change in the compound order
- Compound XXX572 was now on top because it satisfied four out of the five criteria
- XXX518 came second, as the only compound to satisfy all three of the ADME criteria with potency and selectivity data that, based on assay statistics, were not significantly below the required levels




| Name | Profile 3 | Profile 2 | Profile 1 |
|---|---|---|---|
| XXX572 | 1 | 7 | 14 |
| XXX518 | 2 | 52 | 36 |
| XXX582 | 3 | 21 | 13 |
| XXX295 | 4 | 16 | 16 |
| XXX321 | 5 | 26 | 26 |
| XXX025 | 6 | 38 | 35 |
| XXX502 | 7 | 15 | 12 |
| XXX274 | 8 | 35 | 25 |
| XXX292 | 9 | 9 | 4 |
| XXX316 | 10 | 44 | 15 |
| XXX280 | 11 | 13 | 11 |
| XXX278 | 12 | 18 | 23 |
| XXX319 | 13 | 51 | 44 |
| XXX294 | 14 | 17 | 24 |
| XXX282 | 15 | 19 | 20 |

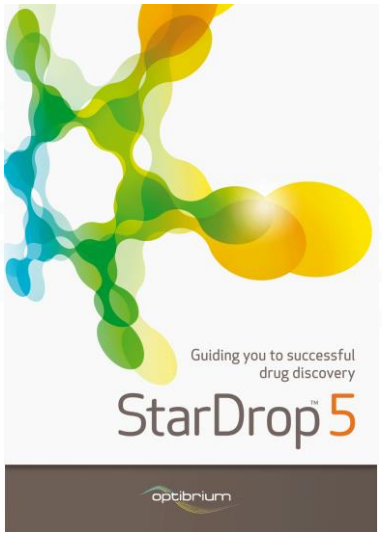
# Case Study - Summary

- MPO is essential to developing compounds with the correct balance of properties
- Identified four compounds that had been overlooked by traditional compound selection based on selectivity and potency cut-off values
- When tested *in vivo*, one of these compounds, XXX518, the only synthesised representative of a novel chemotype, was found to have a superior PK profile
- Project chemists have now expanded this series, investigating ways of improving selectivity and potency in what appears to be a "Good ADME" area of chemistry

**This new chemistry would not have been considered**

# Conclusion

- All the data we work with in drug discovery come from models, be they *in vivo*, *in vitro* or *in silico*, are subject to experimental variability or contain statistical errors
- We can use this information to enable us to highlight the compounds with the greatest potential and to help avoid missed opportunities
- …but…
- We need to make sure we use an appropriate method to account for this



# Acknowledgements


- StarDrop group past and present, including:
  - Matthew Segall
  - Chris Leeding
  - Iskander Yusof
  - James Chisholm
  - Olga Obrezanova
  - Alan Beresford
  - Dawn Yates
  - Dan Hawksley
  - Joelle Gola
  - Brett Saunders
  - Simon Lister
  - Mike Tarbit