optibrium

Making Priors a Priority

Dr Matthew Segall – matt.segall@optibrium.com

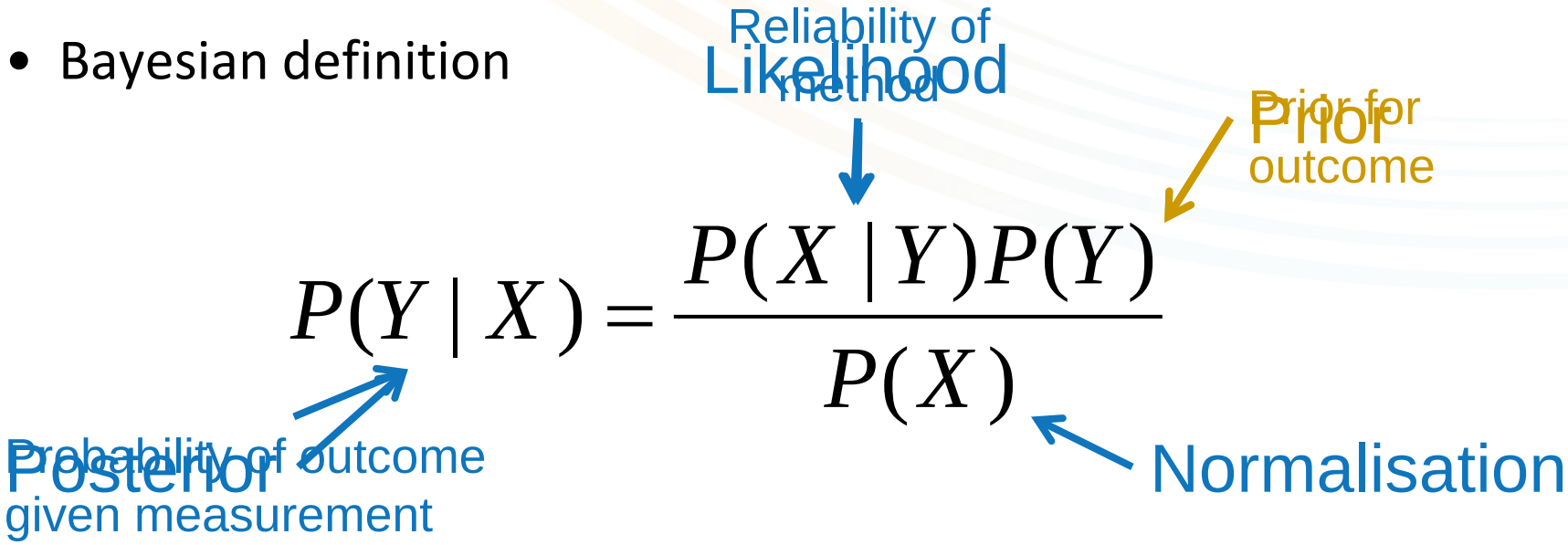
ACS Spring Meeting – March 2011

Overview

- What is a prior?
- Why are priors important?
  - Medical analogy
- Example applications in chemistry
  - Interpreting results
  - Planning screening strategies
  - Multi-parameter optimisation – importance of parameters
- Determining priors for key properties
  - An opportunity for open sharing of data
  - Challenges
- Conclusions

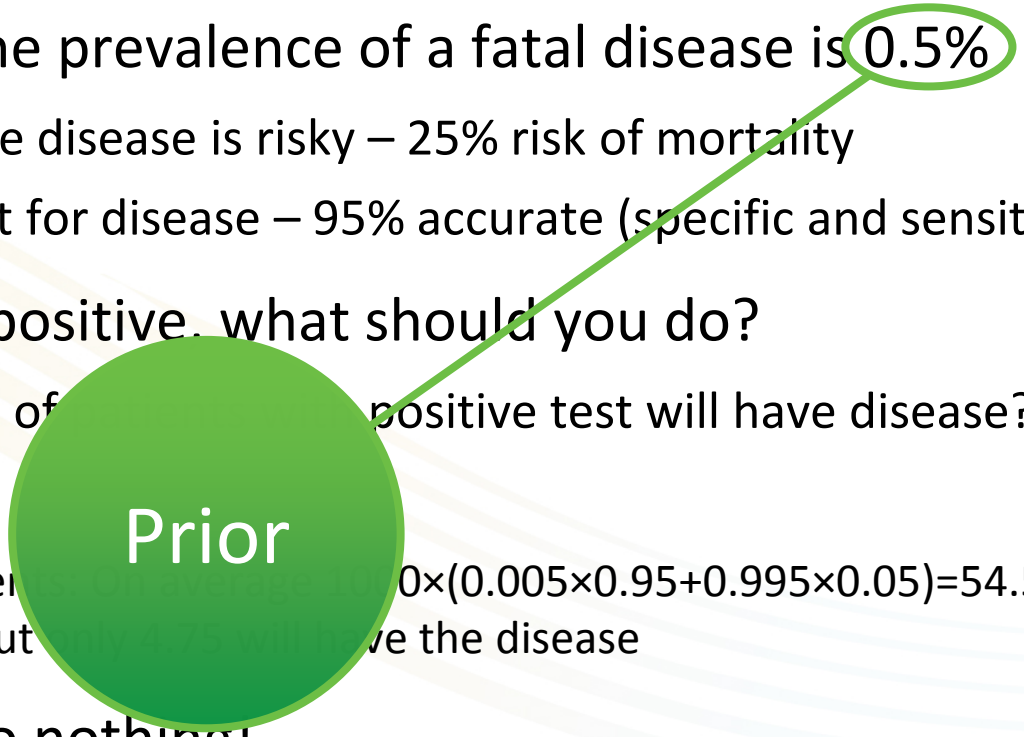
What is a Prior?

- A *prior* captures our understanding, or belief, of the likely outcomes of an event before the collection of new information (e.g. a measurement or prediction)
- More specifically, it is a probability distribution of an outcome, *P(Y)*, in the absence of the additional information
- Bayesian definition



Why are Priors Important?
Medical Analogy

- In a population the prevalence of a fatal disease is 0.5%
  - Treatment for the disease is risky – 25% risk of mortality
  - Simple blood test for disease – 95% accurate (specific and sensitive)
- If a patient tests positive, what should you do?
  - What proportion of [illegible] positive test will have disease?
  - Answer: 9%
    - Test 1000 patients: On average 1000×(0.005×0.95+0.995×0.05)=54.5 will test positive, but only 4.75 will have the disease
- Best decision – do nothing!
  - 9% of those that test positive will die due to disease
  - 23% will die unnecessarily due to treatment

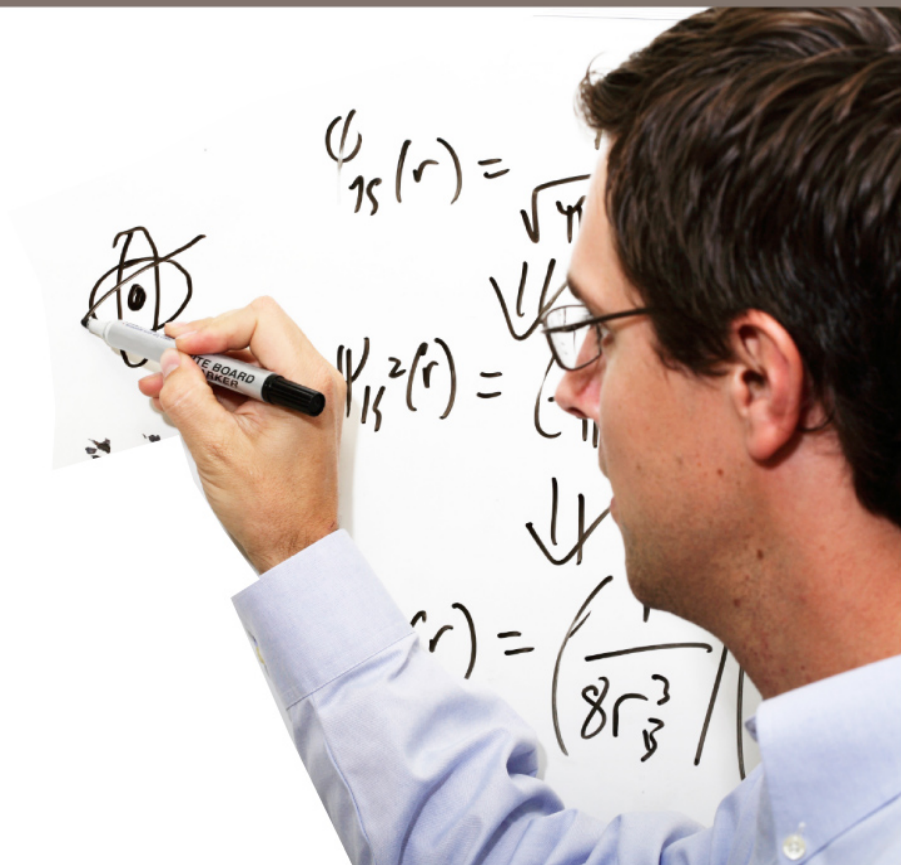


Why are Priors Important?
Medical Analogy

- What must prevalence of disease be before 95% accurate test is useful?
  - 1.3%
- How accurate must test be to be useful with a prevalence of 0.5%
  - 98%
- **Key point: Utility of test depends critically on prevalence of negative outcome being tested for, i.e. the prior**

Example Application in Chemistry

optibrium



Example Application

How well does this assay conserve your options?

- You have purchased a series of compounds:
  - You expect 1% of your compounds have a particular kind of toxicity
  - You apply a screening method to all the compounds that is 90% reliable (both 90% sensitive and 90% specific)
  - What percentage of the compounds that fail the screening 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 test, of which only 9 really are toxic.

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

*A Chadwick and M Segall, Drug Discov. Today, **15**(13/14), pp. 561-9 , July 2010

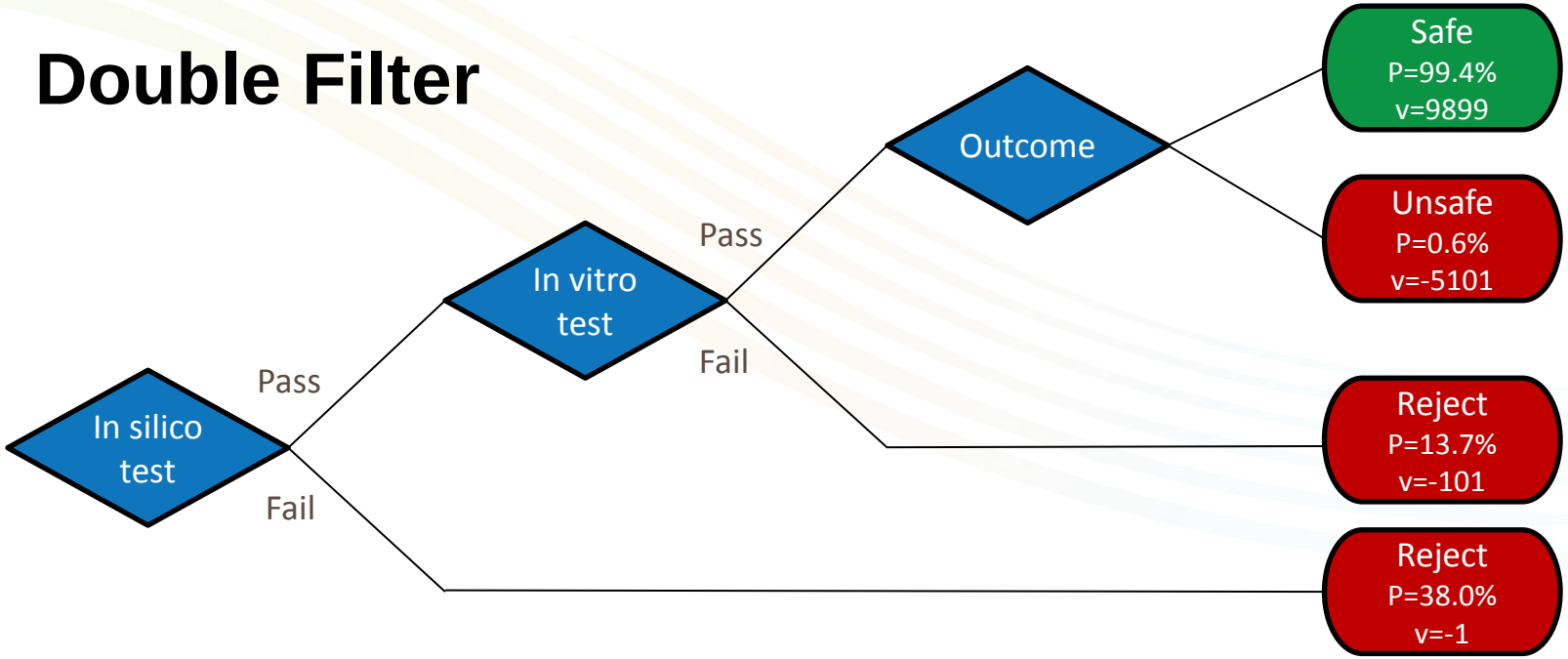
Example Application
Screening Strategy

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

Example Application
Screening Strategy

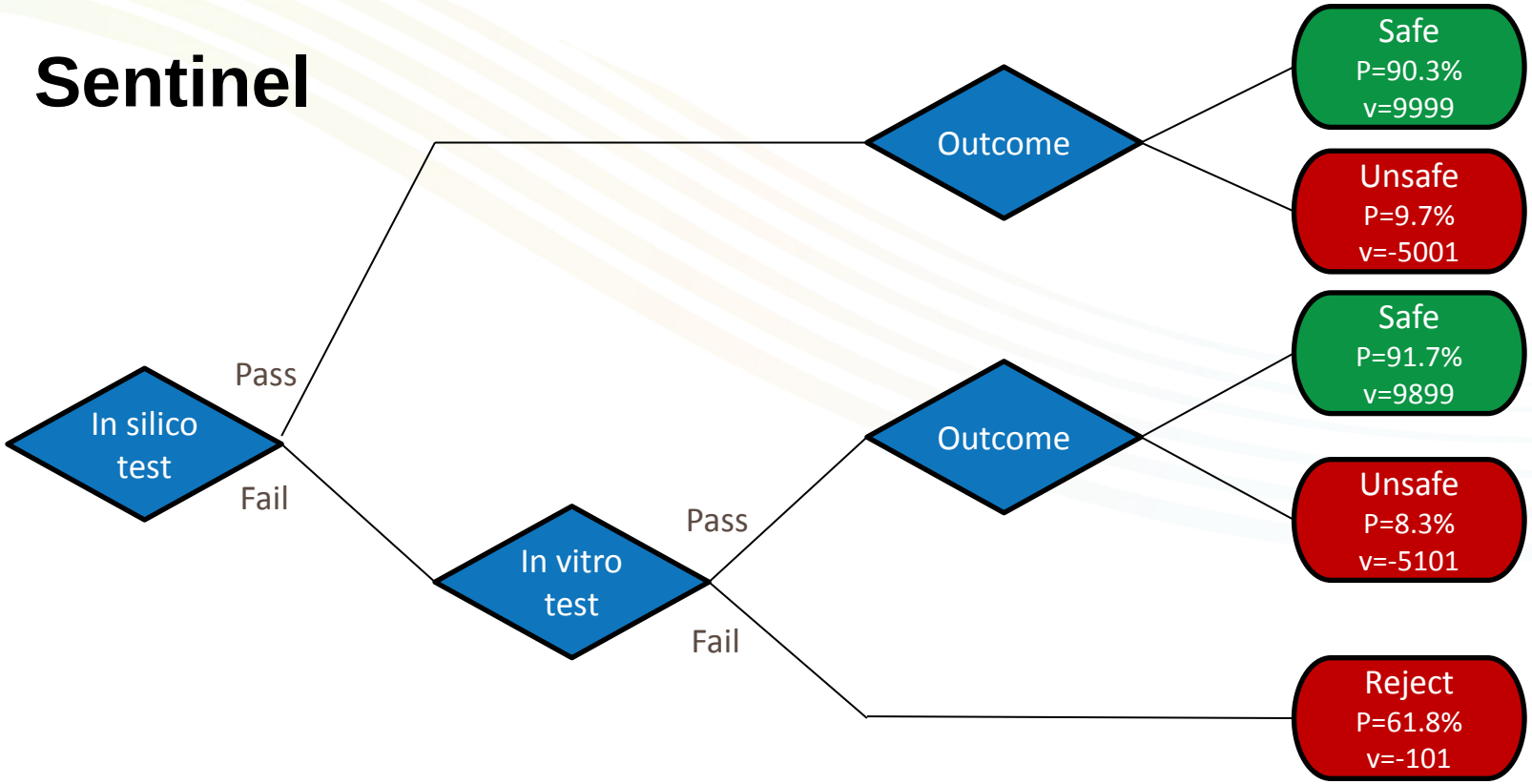
- Two screens for toxicity: *in silico* and *in vitro*
- 5 Possible screening strategies:

Double Filter



Example Application
Screening Strategy

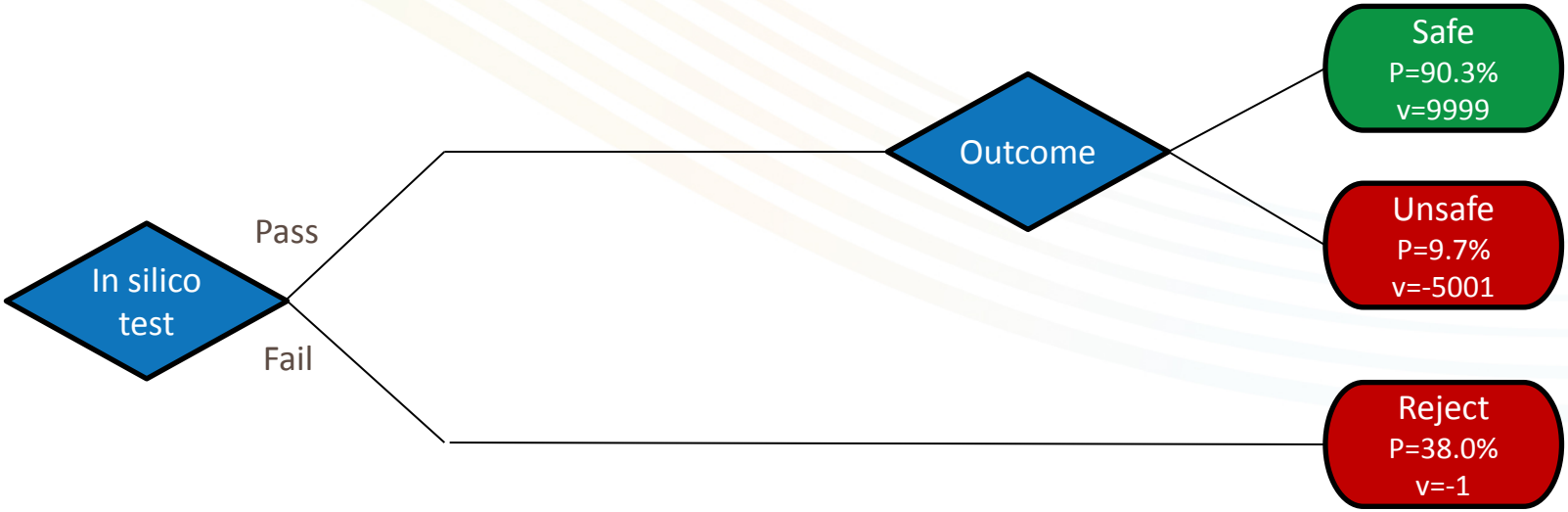
- Two screens for toxicity: *in silico* and *in vitro*
- 5 Possible screening strategies:



Example Application
Screening Strategy

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

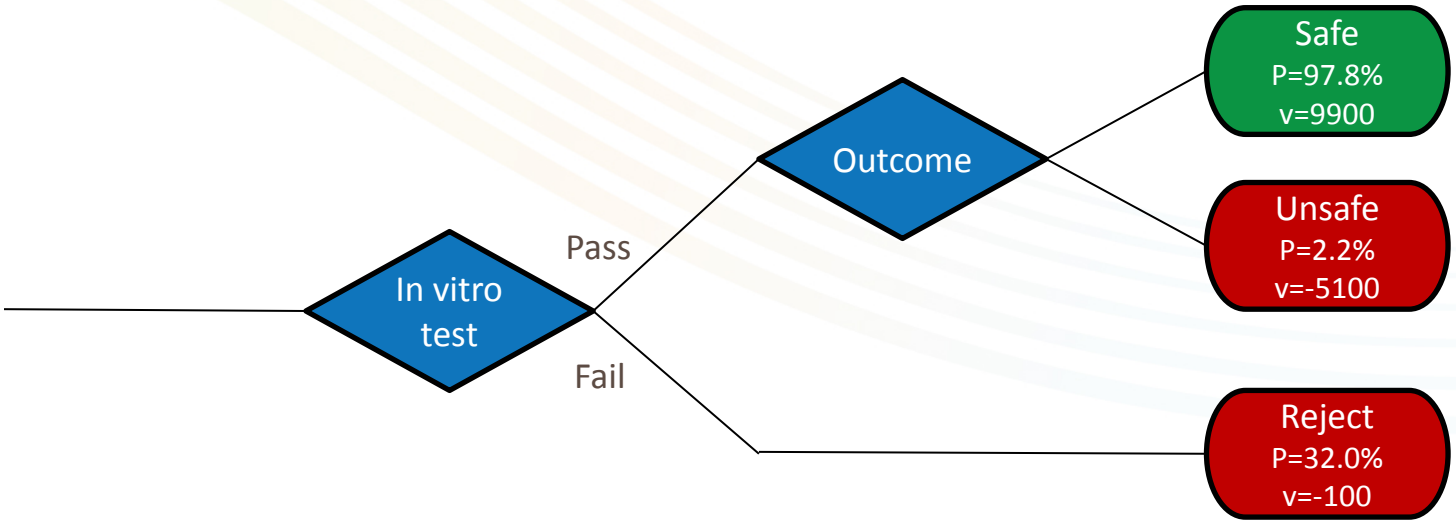
In Silico Only



Example Application
Screening Strategy

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

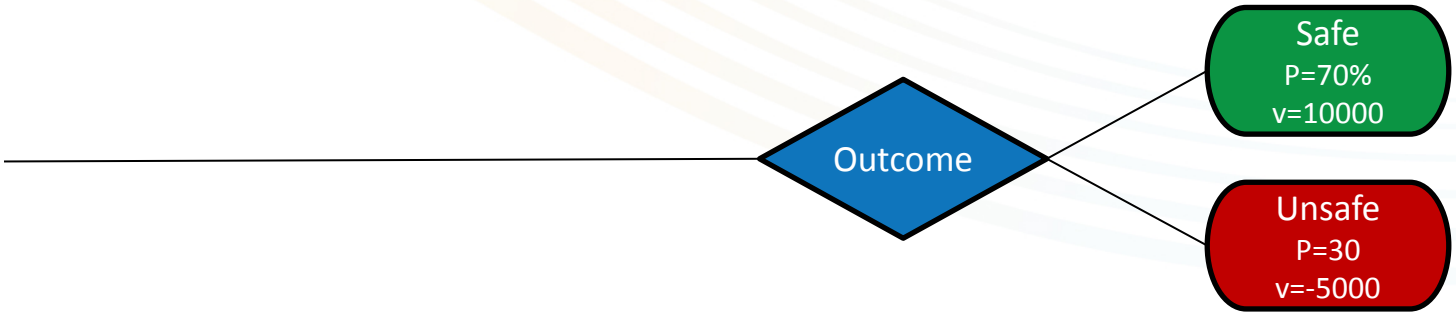
In Vitro Only



Example Application
Screening Strategy

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

No Screen



Example Application
Screening Strategy

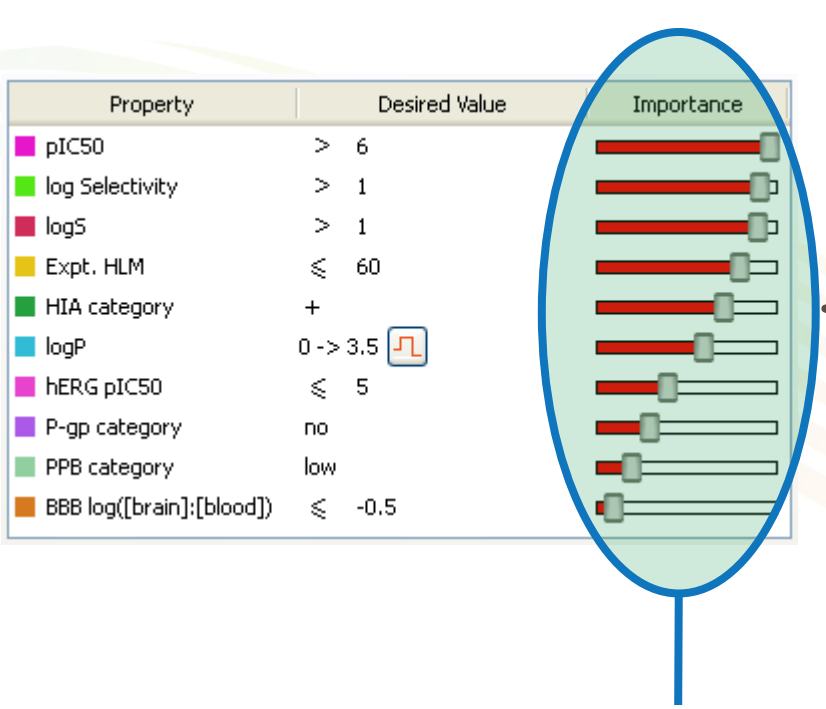
- Parameters:
  - *In silico*: cost 1, accuracy 80%
  - *In vitro*: cost 100, accuracy 95%
  - Cost to confirm safety 5,000; Net value of safe compound 10,000

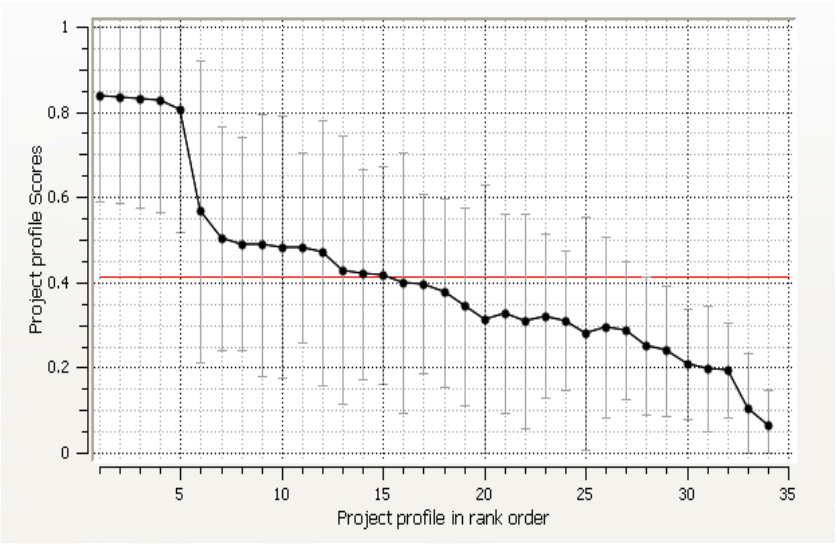
Strategy	Value (Prior for risk 30%)	Value (Prior for risk 40%)
Double filter	5242	4483
Sentinel	6531	5415
In silico only	5299	4399
In vitro only	6475	5500
No screen	5500	4000

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

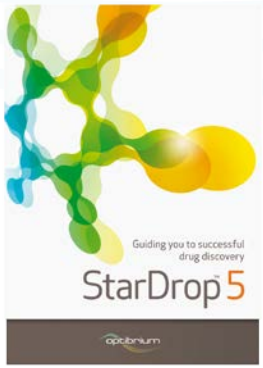
Example Application
Multi-parameter Optimisation

- E.g. Probabilistic scoring*





Importance values related to downstream risk due to negative result.

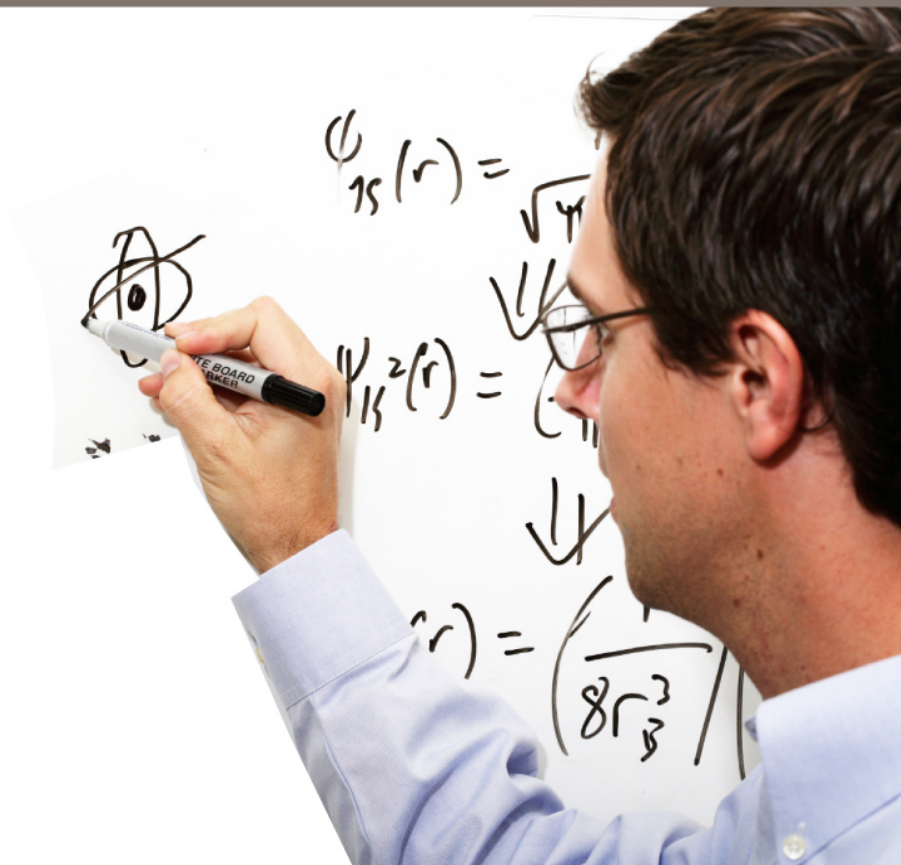


* Segall *et al.* Chemistry and Biodiversity **6**(11), p. 2144 (2009)

Determining priors for key properties

An opportunity for open data

optibrium



Determining Priors

- Outcomes for key endpoints for large numbers of compounds
  - E.g. physicochemical properties, ADME*, PK†, toxicity…
- Early and late stage endpoints (late most valuable)
  - E.g. hERG inhibition vs. Torsade de points in humans
- Ideal opportunity for sharing data
  - No compound structures required
  - Data is almost free of I.P.

* Absorption, Distribution, Metabolism, Elimination

† Pharmacokinetics

Determining Priors
Questions and challenges

- What is appropriate population to sample?
  - Even a chemist's intuition is a filter
  - Subdivided by field: Drug discovery, agrochem, cosmetics…
  - Perhaps subdivided into indication: anti-infectives, oncology, pesticide…
- Normalisation of data
  - Different assay protocols
- Insufficient data for late stage outcomes, e.g. clinical
  - Late stage compounds have been heavily filtered
  - Need to use early screening data to infer late-stage outcomes based on reliability
  - Need to share data on reliability

Conclusions

- Knowledge of priors is essential to good decision-making
  - How good does my assay/model need to be to be useful?
- Priors for critical endpoints are essentially unknown
- Challenges for analysis of data
- This is an ideal opportunity for an open data project
  - Outcome will benefit entire community
  - Pre-competitive
- To discuss:
  - www.optibrium.com/community
  - matt.segall@optibrium.com

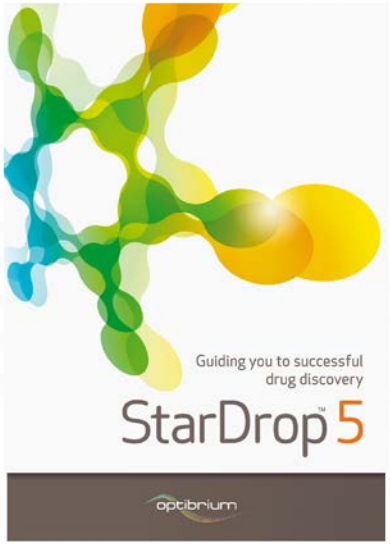
Stand #1121

* Segall and Chadwick, J. Comp. Aided Mol. Des. **24**(12), pp. 957-960 (2010)

Acknowledgements

- Andrew Chadwick – Tessella plc.

- StarDrop group past and present
  - Ed Champness
  - Chris Leeding
  - Iskander Yusof
  - James Chisholm
  - Olga Obrezanova
  - Alan Beresford
  - Dan Hawksley
  - Joelle Gola
  - Brett Saunders
  - Simon Lister
  - Mike Tarbit



Stand #1121