Simulation of Drug Development Programs & estimation of value

Tom Parke

The proposition

- By simulating the proposed sequence of trials in a drug development program, and assessing the expected value under different circumstances we gain a tool that enables us to make complex decisions in designing a drug development program in an open and quantified way.

Relative sizes of phase 2 & phase 3

EXAMPLE 1

Expectation of dose response

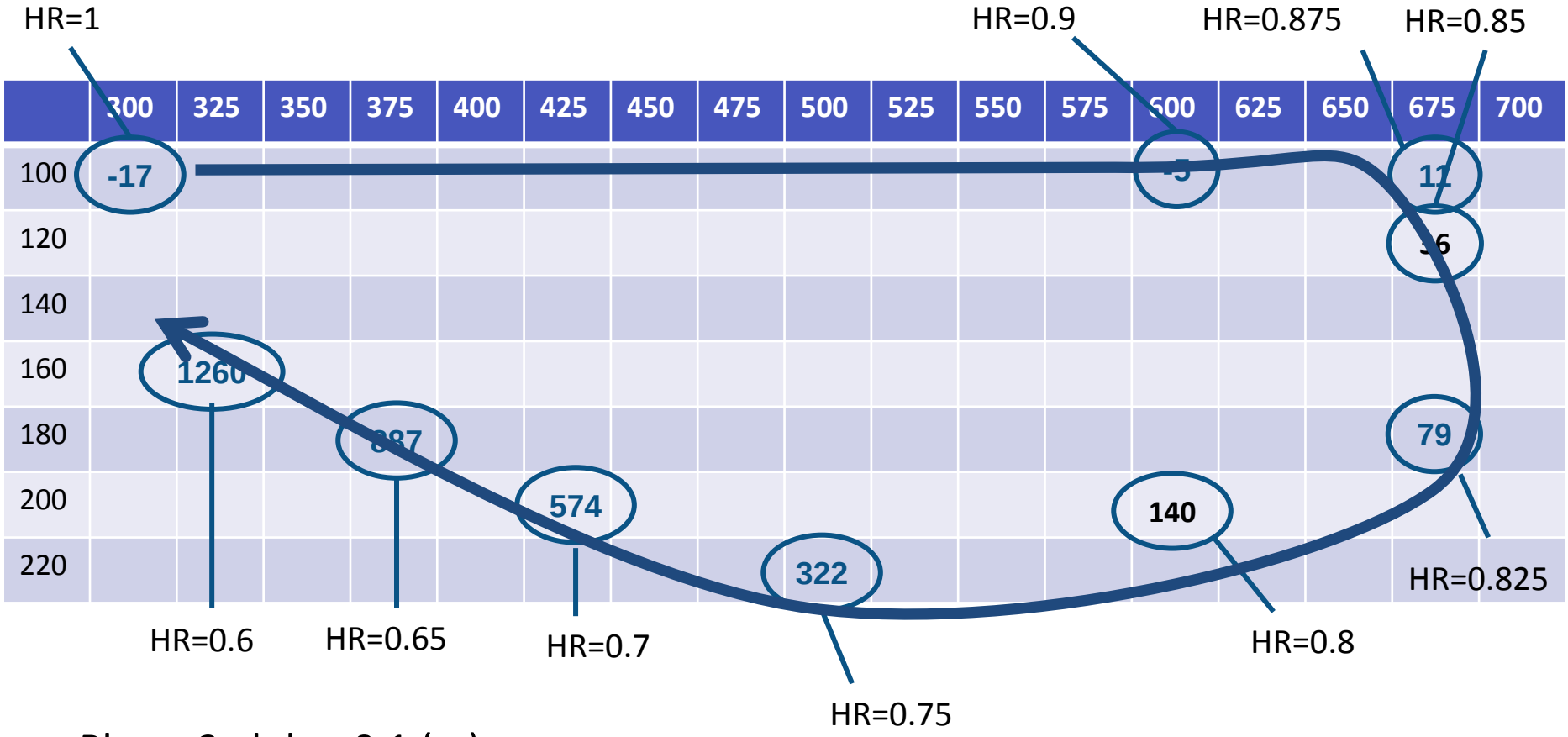
- Treatments to be compared to Gemcitabine
- Expected HR is in the range 1 – 0.6
- Use scenarios of 1, 0.9, 0.8, 0.7, 0.6
- Reflect uncertainty in how likely drug is to be good by looking at 3 ‘priors’:

	1	0.9	0.8	0.7	0.6
Optimistic	9%	9%	18%	46%	18%
Uniform	20%	20%	20%	20%	20%
Pessimistic	41.5%	16.5%	16.5%	16.5%	9%

If we had just one treatment, how big should Phase 2 & Phase 3 be?

- Calculate probability of success & expected NPV for all combinations of phase 2 & phase 3
- for all Hazard Ratios
- What combination of phase 2 & phase 3 size is optimal for a given HR?
- If we average over our prior expectation of HR for our drug which is the optimum combination?

Pancreatic Cancer Example – optimum eNPV ($M) phase 2 & phase 3 by expected HR



Phase 2 alpha: 0.1 (ss)

Phase 3 alpha: 0.025 (ss)

Berry Consultants
Statistical Innovation

Power

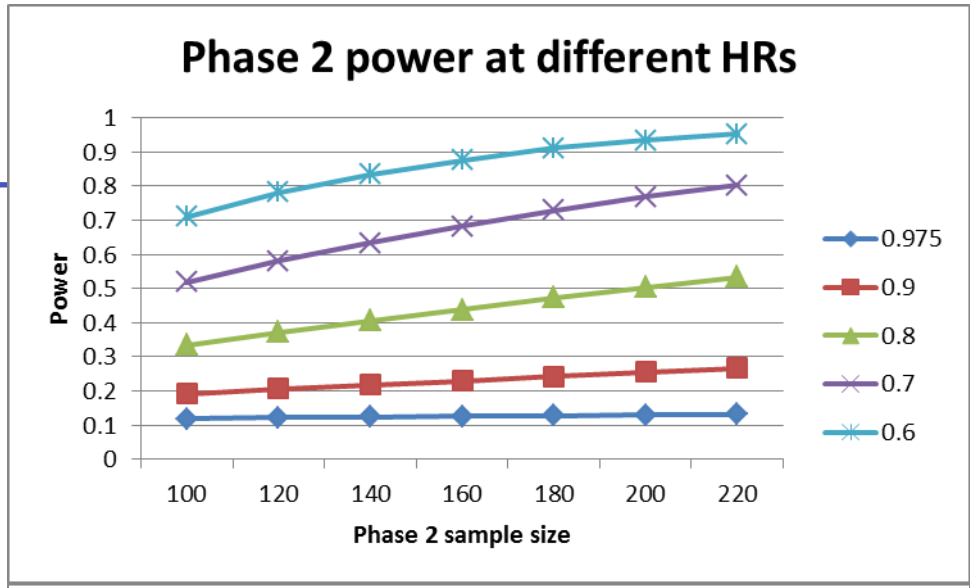
Power increases with sample size

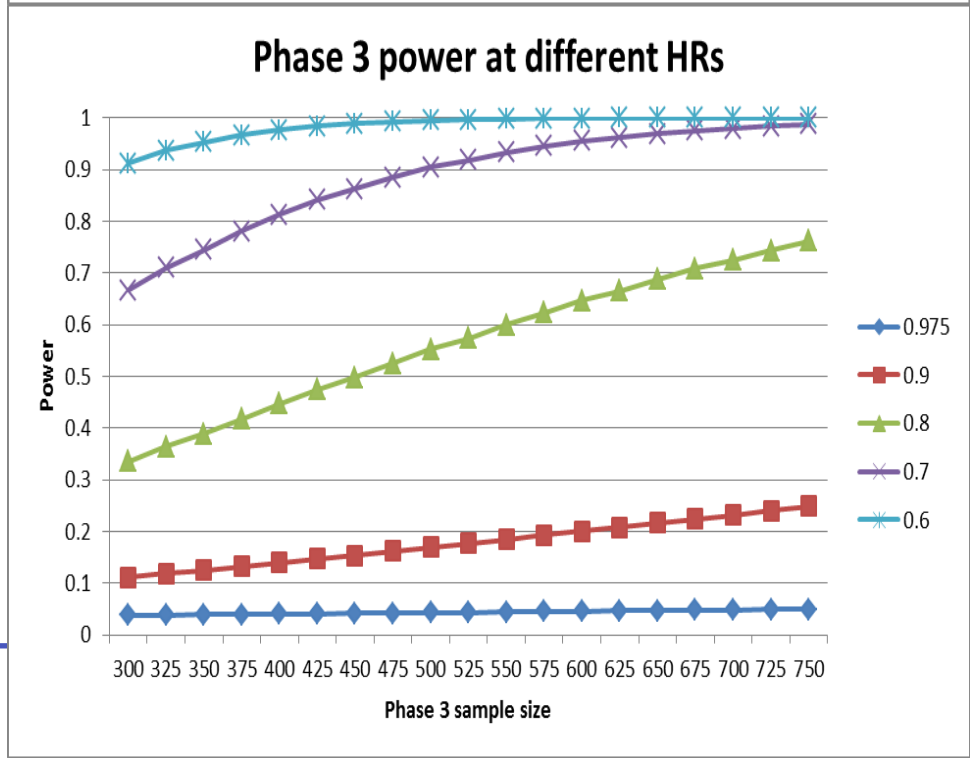
- Very little at poor HR
- Then ~linearly at moderate HR
- Then rapidly at high HR but tails off

Revenue decreases with sample size ~linearly (decreasing patent life)

Cost increases with sample size ~linearly

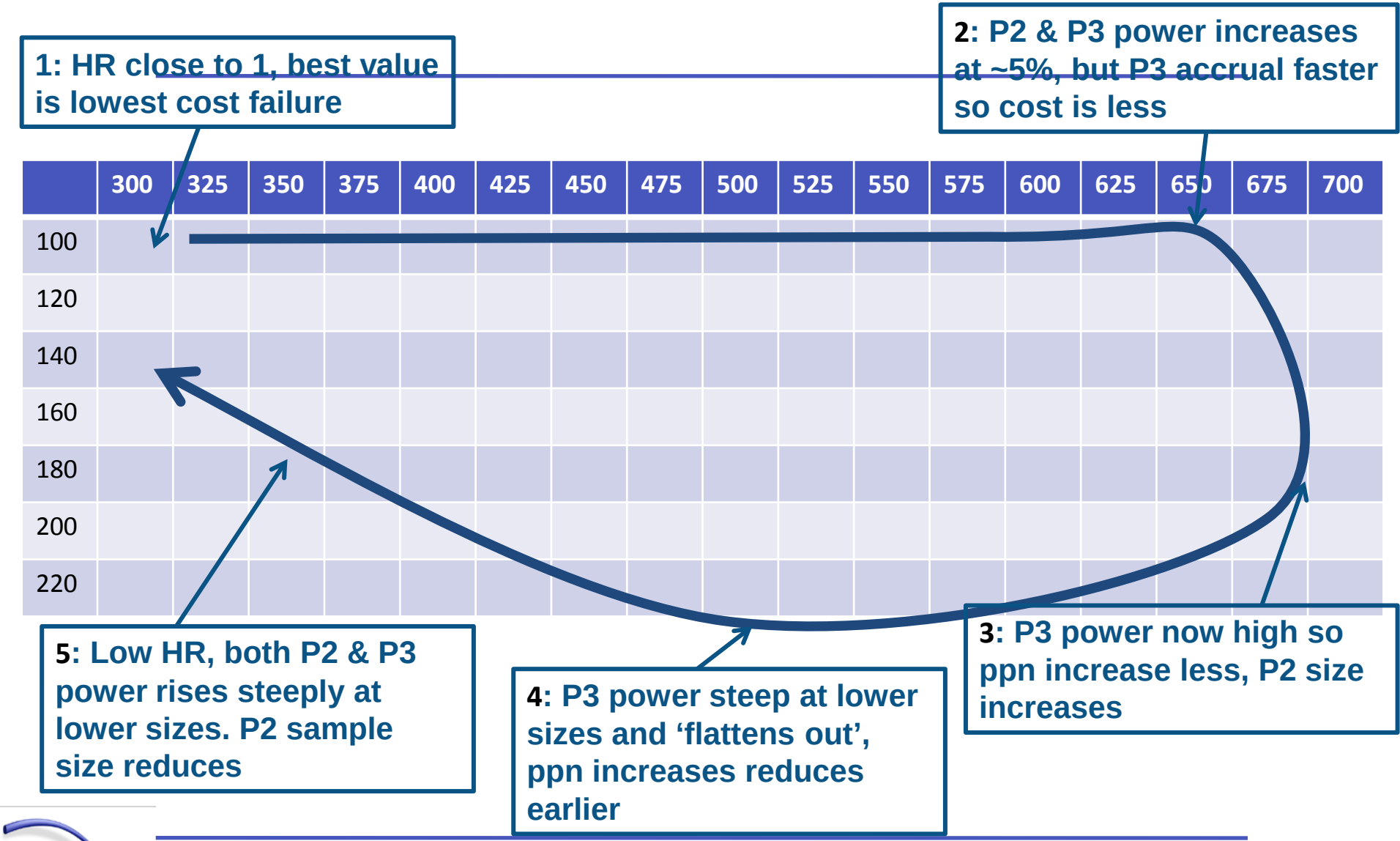
So increasing sample size while the proportional increase in power exceeds the proportional decrease in value => increases eNPV





Berry Consultants
Statistical Innovation

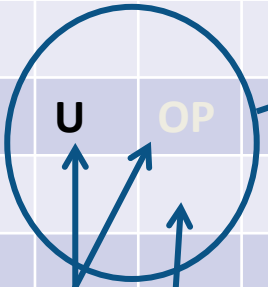
Optimum eNPV ($M) explained



Berry Consultants
Statistical Innovation

Average over different expectations

	300	325	350	375	400	425	450	475	500	525	550	575	600	625	650	675	700
100																	
120																	
140																	
160																	
180																	
200																	
220																	



P2 & P3 sample sizes that maximise expected value over different assumed probabilities of drug response (HR)

	1	0.9	0.8	0.7	0.6
Optimistic	9%	9%	18%	46%	18%
Uniform	20%	20%	20%	20%	20%
Pessimistic	41.5%	16.5%	16.5%	16.5%	9%

Schizophrenia – phase 2 dose finding

EXAMPLE 2

Scope of the prototype

- Phase 2 & Phase 3
  - A dose finding phase 2 simulated in FACTS (dose finding with a continuous endpoint)
  - Simple, fixed size 2 arm phase 3's run in parallel – use simple power calculation
- Simple, but flexible, cost & revenue models
- A range of scenarios
- Evaluate time, cost, probability of success and eNPV.
- Re-assess with varying parameters
  - Sensitivity analysis
  - Optimisation

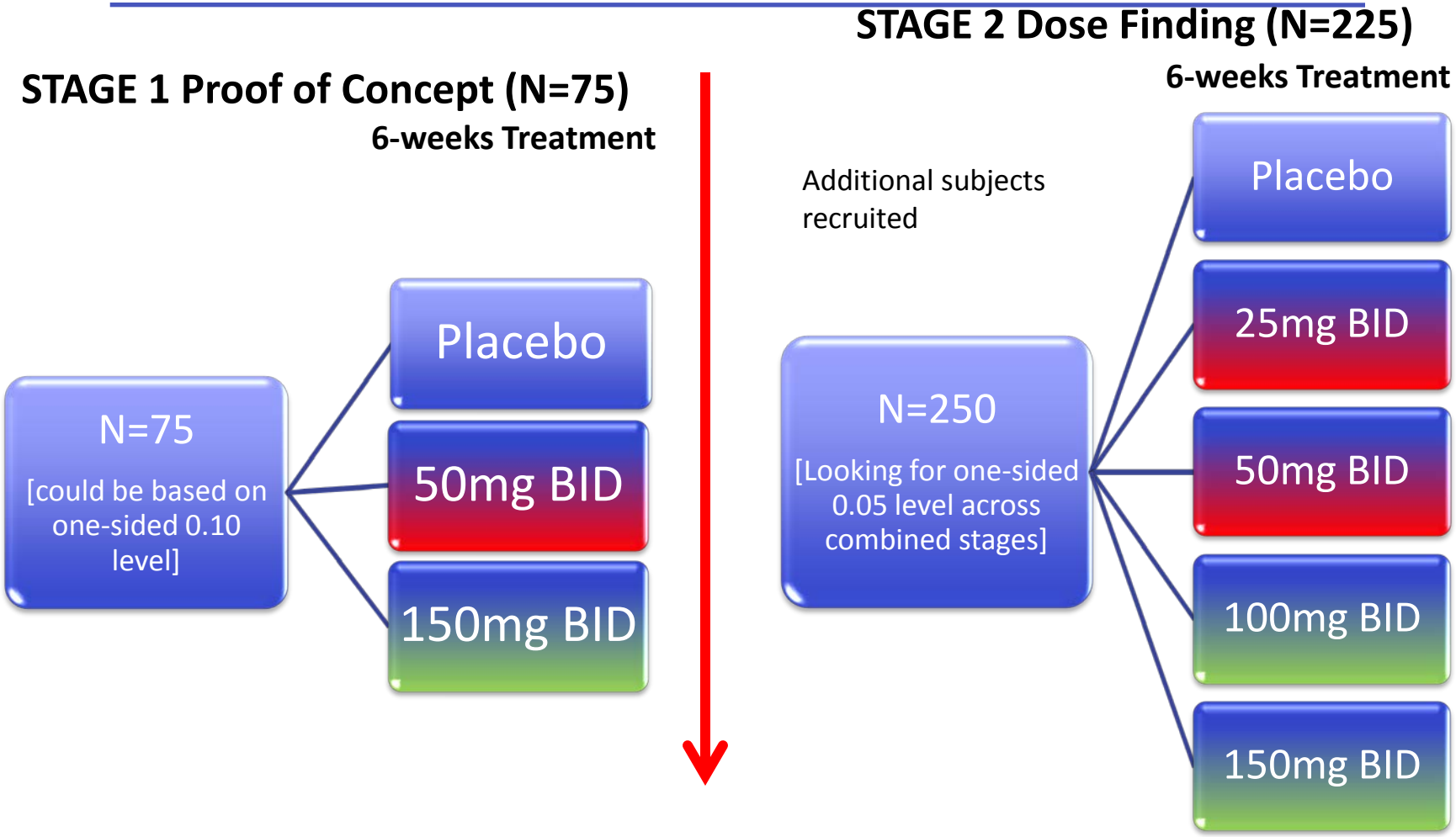
Case Study from ISCTM adaptive design group

- International Society for CNS Trial Methodology
- http://www.isctm.org
- Presented at the 2012 Autumn Conference
- Published in DIA Journal: “Therapeutic Innovation and regulatory Science” Jan 2014

Study Proposal

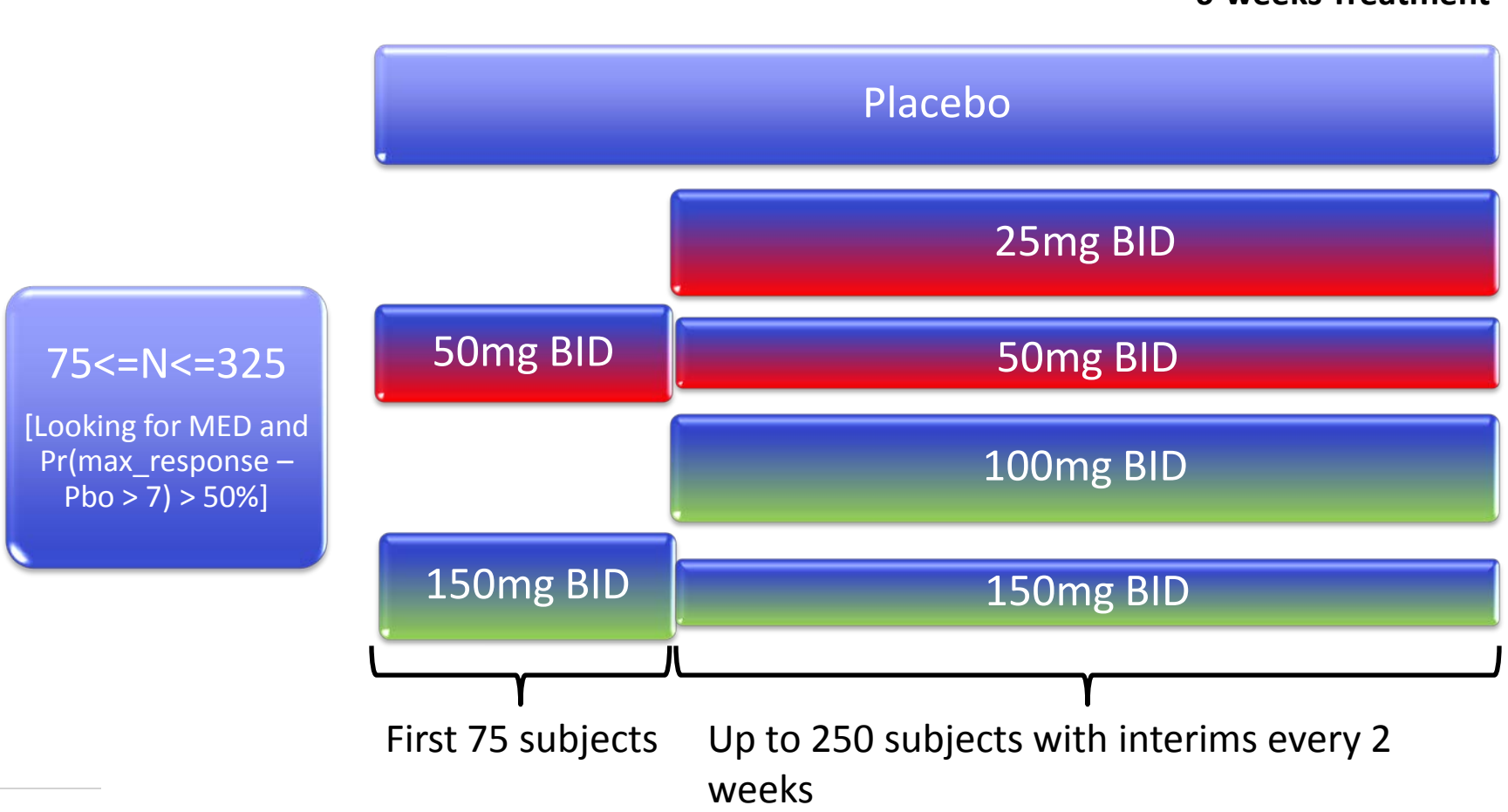
- Multicenter, double-blind, randomized, placebo-controlled, parallel-group, dose-response study in male and female subjects with schizophrenia
- The primary objective is to evaluate the efficacy via change from baseline in the total Positive and Negative Syndrome Scale (PANSS) total score of multiple fixed doses of Compound X as an adjunctive treatment to a D2-based antipsychotic compared with adjunctive placebo after 6 weeks treatment in subjects with schizophrenia

Conventional Phase II Study Design



Adaptive Seamless Phase II Study Design

Test 50mg & 150mg during 'burn-in' then open up other doses and adaptively allocate to find MED.



To compare

- 7 simple scenarios
  - SD of endpoint 20 points
  - Lower score is better
  - Usual placebo response: -16

Scenario	Pbo	25mg	50mg	100mg	150mg
Null	-16	-16	-16	-16	-16
High Dose	-16	-17	-18	-20	-28
Middle Dose	-16	-17	-22	-27	-28
Low Dose	-16	-21	-25	-27	-28
Weak	-16	-16	-17	-18	-21
Peak 50mg	-16	-21	-26	-22	-18
Peak 100mg	-16	-20	-24	-28	-24

Scenario	Design	PPn Fail PoC	Fail before Cap	Fail at Cap	Success	Mean Sample Size
Null	fixed	0.814	0.0	0.168	0.018	118 / 325
	adaptive	*0.379*	*0.563*	*0.028*	*0.007*	*125 / 273*
Weak	fixed	0.586	0.0	0.214	0.200	142 / 325
	adaptive	*0.143*	*0.448*	*0.069*	*0.197*	*169 / 288*
High Dose	fixed	0.183	0.0	0.194	0.623	204 / 325
	adaptive	*0.017*	*0.046*	*0.002*	*0.898*	*155 / 248*
Middle Dose	fixed	0.148	0.0	0.151	0.701	201 / 325
	adaptive	*0.006*	*0.012*	*0.001*	*0.968*	*111 / 241*
Low Dose	fixed	0.117	0.0	0.196	0.687	232 / 325
	adaptive	*0.006*	*0.009*	*0*	*0.974*	*124 / 247*
Peak 50 mg	fixed	0.259	0.0	0.187	0.554	180 / 325
	adaptive	*0.057*	*0.13*	*0.02*	*0.698*	*152 / 252*
Peak 100mg	fixed	0.245	0.0	0.215	0.540	192 / 325
	adaptive	*0.025*	*0.03*	*0*	*0.911*	*114 / 253*

Explanation of results

- Scenario – the dose response scenario being simulated. 1,000 simulations were run of each. For each scenario the mean change from baseline is as described on the 'to compare' slide, with SD of 20.
- Fail PoC: the proportion of simulations where the trial failed at the end of the "PoC" stage. In the fixed design this will be after the PoC trial is complete. In the adaptive design this will be the proportion that failed at the first interim.
- Fail before Cap: the proportion of simulations that stopped for futility after the first interim but before recruiting the full sample size (the 'cap'). Only the adaptive design checked for stopping for futility after the PoC interim.
- Fail at Cap: the proportion of simulations that reached the 'cap' (full accrual) and followed all subjects to their final endpoint and declared futility.

Explanation of results (cont'd)

- Success : the proportion of simulations that reached phase 2b and were successful. This is the probability of success across both stages.
  - In the fixed design futility and success are decided based on conventional statistical significance and hence all trials are either successful or futile.
  - In the adaptive design based on the posterior estimate that the best response is better than placebo by a 'clinically significant difference'. If the probability is > 0.55 the trial declared a success.
- Mean Sample Size: for that scenario the mean number of subjects recruited: if the result was futility / if the result was successful.
  - For the fixed design the number recruited is 75 in the PoC study and (if run) 250 in the phase 2 study.
  - For the adaptive design the minimum number of subjects recruited is 75 (time of the first interim) and the maximum number is 325. Interims are conducted every 2 weeks (approx every 14 subjects) that allows the randomization to be adjusted and for the trial to be stopped for futility or for success.

Scenario	Control	25mg	50mg	100mg	150mg
High Dose	-16	-17	-18	-20	-28
fix'd P(MED)	*0.000*	*0.005*	*0.005*	*0.071*	*0.542*
adpt P(MED)	*0.000*	*0.017*	*0.024*	*0.151*	*0.701*
Middle Dose	-16	*-17*	*-22*	*-27*	*-28*
fix'd P(MED)	*0.000*	*0.020*	*0.114*	*0.368*	*0.119*
adpt P(MED)	*0.000*	*0.038*	*0.293*	*0.577*	*0.058*
Low Dose	-16	*-21*	*-25*	*-27*	*-28*
fix'd P(MED)	*0.000*	*0.204*	*0.196*	*0.207*	*0.080*
adpt P(MED)	*0.000*	*0.206*	*0.525*	*0.201*	*0.039*
Peak 50 mg	-16	*-21*	*-26*	*-22*	*-18*
fix'd P(MED)	*0.000*	*0.037*	*0.502*	*0.004*	*0.012*
adpt P(MED)	*0.000*	*0.203*	*0.478*	*0.007*	*0.000*
Peak 100mg	-16	*-20*	*-24*	*-28*	*-24*
fix'd P(MED)	*0.000*	*0.054*	*0.143*	*0.154*	*0.190*
adpt P(MED)	*0.000*	*0.136*	*0.471*	*0.296*	*0.003*

Explanation of P(MED)

- P(MED): this is the probability of determining success **and** selecting the given dose as the MED.
- In the fixed design, the MED is the lowest dose tested in phase 2 (recall only 2 or 3 doses are tested depending on the performance of the 50mg dose in 2a) that has a treatment difference from the control arm of 5 points or better (the 'target' is 10 points, but to use this as the CSD results in very poor power to detect).
- In the adaptive design, the MED is the dose with the highest posterior probability of being the lowest dose with a treatment difference from the control arm of 7 points or better.

Comparison

- The Adaptive, seamless 2a/2b
  - In the Null scenario has lower type-1 error: 0.7% compared to 1.8%
  - And all positive scenarios has higher power: 89% on average compared to 69%
  - Is more likely to be successful and select the correct MED: 51% on average compared to 35%
    - In one scenario the fixed is marginally better, in the 'peak at 50mg scenario', fixed: 50.2% adapt: 47.8%
  - With only two exceptions the adaptive has a lower sample size
    - The exceptions are the expected mean sample size in the Null and Weak scenarios when the outcome is (correctly) futility .. In the Null scenario fixed : 118, adapt: 125, and Weak scenario fixed : 142, adapt:169.
    - The average expected sample size when successful in the successful scenarios is 254 compared to 325 for the fixed design.
  - Can we make the comparison clearer computing the eNPV?

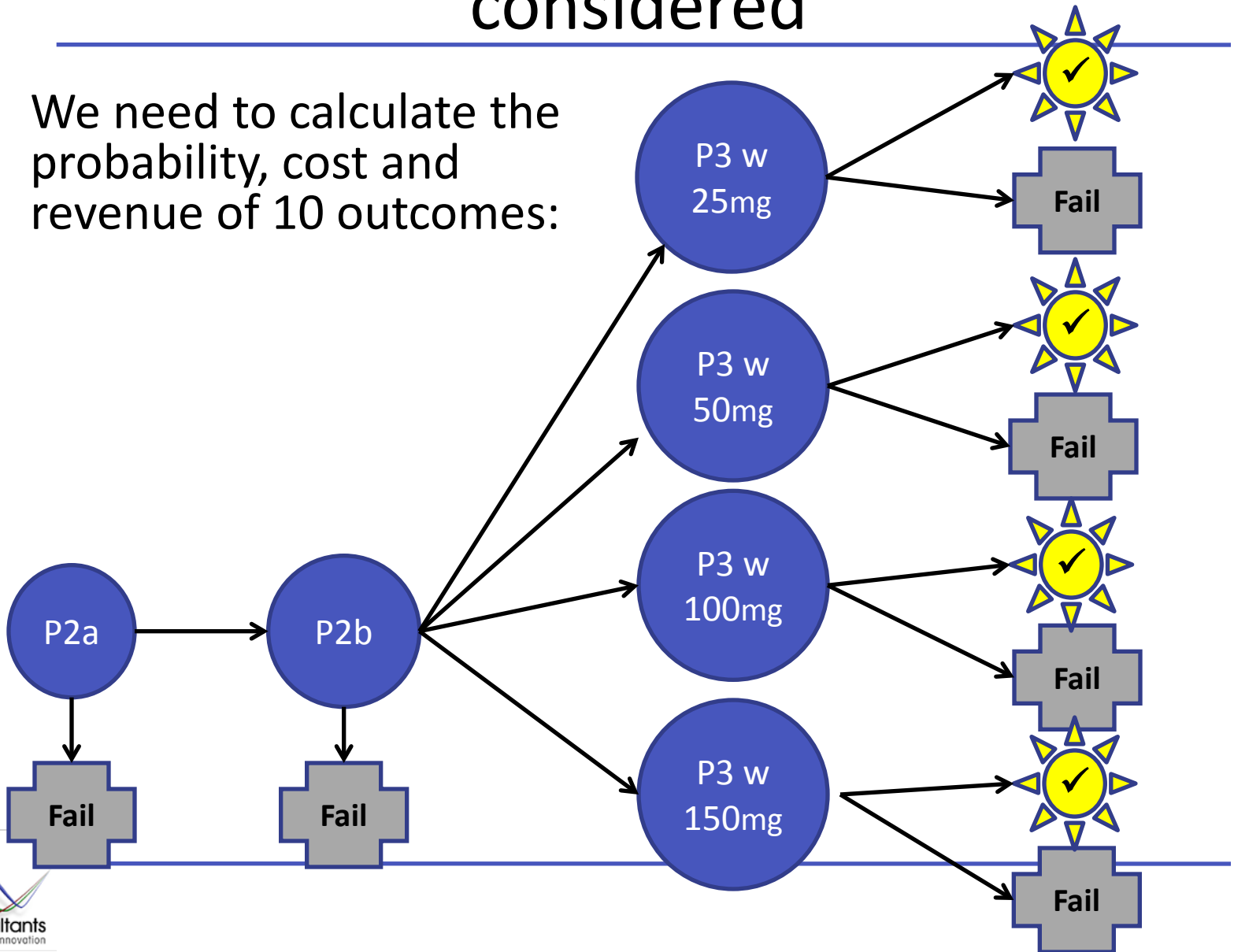
Compared over 7 dose response scenarios

- If drug doesn't work, fixed design has smaller expected sample size.
- If drug works, adaptive design has smaller expected sample size, better power and in all but 1 case better ability to select the MED.
- But is adaptive design worth it? Is it worth the risk, the upfront cost, the additional design time?

- So we performed an eNPV calculation
- Predicted eNPV value highly questionable due to huge uncertainty in commercial assumptions
- But relative eNPVs can be used to compare strategies

In each likely dose-response scenario considered

- We need to calculate the probability, cost and revenue of 10 outcomes:



Berry Consultants
Statistical Innovation

Fail phase 2 or 3

- Fail phase 2
  - Cost is cost of planning and running phase 2
  - Revenue is 0
  - Probability = P(P2 fail)
- Fail phase 3
  - Cost is cost of planning and running phase 2 + cost of planning and running 2 phase 3 trials
  - Revenue is 0
  - Probability =
    - $P(P2 \text{ success}) * \Sigma_i P(P3 \text{ fail} \mid D_i) * P(D_i)$
  - D_i = choice of dose for P3 at the end of P2

Success in P3

- Cost is cost of planning and running phase 2 + cost of planning and running phase 3 trials
- Revenue is Σ_i NPV(D_i)*P(D_i)
- Probability =
  - P(P2 success)* Σ_i P(P3 success | D_i)*P(D_i)
- D_i = choice of dose for P3 at the end of P2

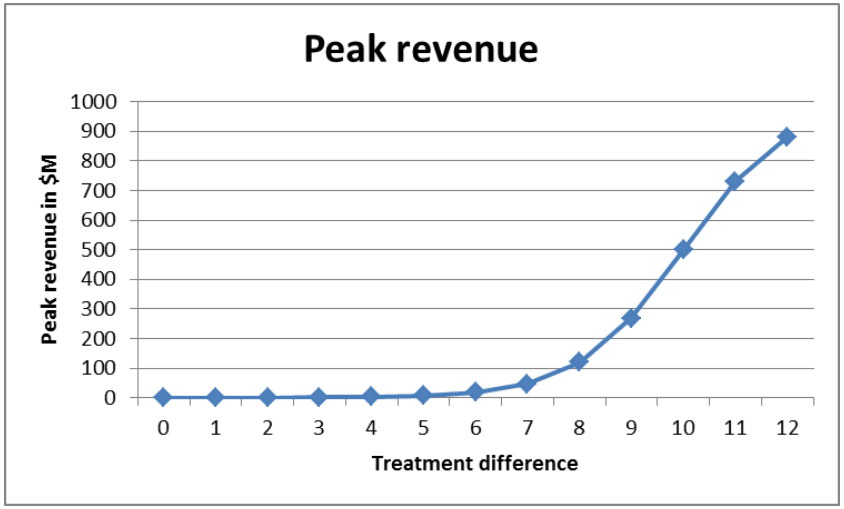
Phase 3 assumption

- To simplify we assumed P3 is fixed, pre-planned size (independent of result of phase 2)
- Specifically: 2 phase 3 trials, each one:
  - 2 arms
  - 2-sided alpha 0.05
  - Power 0.9 for assumed mean difference of 8 points
  - 132 per arm
  - Actual power of course depends on true effect size, which depends on the scenario and on the dose selected in phase 2.

NPV

- Total Revenue =
  - 0.5 * Peak revenue * remaining commercial life * discount
- Remaining commercial life = 12yrs – development time
- Discount = $(1 - 0.09)^{\text{development time}}$

- Expected peak annual revenue we model as based on the true mean treatment benefit relative to placebo:



Berry Consultants
Statistical Innovation

Time assumptions model

- Time elapsed already since IND
- Expected period of exclusivity
- Time to plan trials
- Time to recruit, follow up and analyse trials
- Time to register
- Period of exclusivity remaining

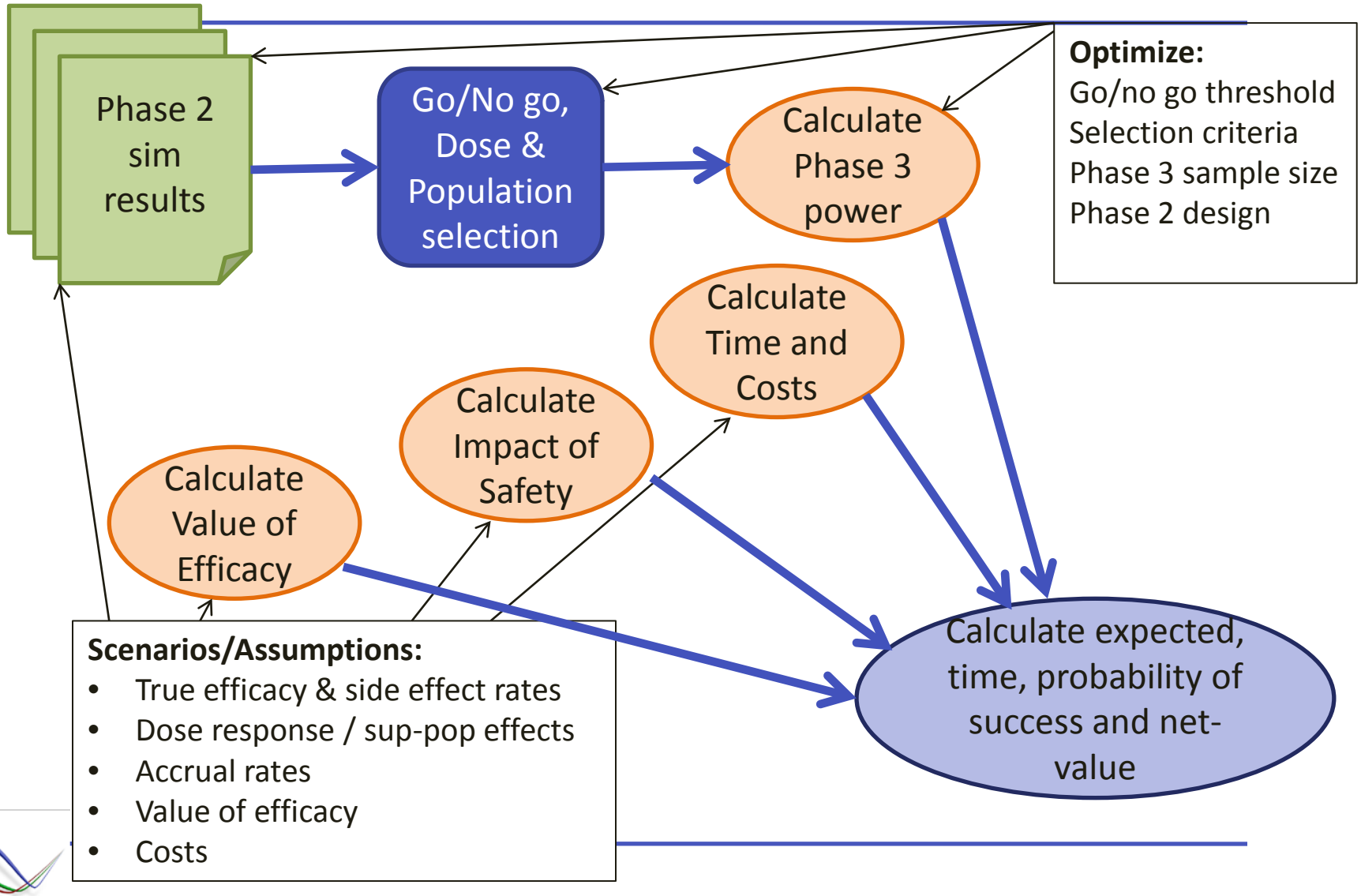
Cost model

- Development program annual overhead cost
- Per trial overhead cost
- Trial cost per subject

Development Program Prototype Simulator

- Specify dose response scenarios, and design phase 2 in FACTS
- Run simulations – e.g. 1,000 per scenario
- Load simulation results in DDPSim
- Enter phase 2 and phase 3 time and cost parameters
- DDPSim
  - Takes phase 2 results and final outcome defined in FACTS
  - Takes specified selected target dose (MED, ED or max)
  - Calculates phase 3 power given specified phase 3 design and true response of selected dose
  - Averages over all simulation results for scenario
  - Averages over all scenarios (using supplied weightings)
- Allows alternate phase 3 designs, sensitivity to parameters and some alternate phase 2 criteria to be explored

Program Simulator



Berry Consultants
Statistical Innovation

NPV per scenario ($M)

Scenario	Fixed Design	Adaptive seamless P2a/b
Null	-10.32	-11.68
Weak	-20.17	-22.06
High Dose	650.05	1024.96
Middle Dose	477.05	756.06
Low Dose	282.23	431.13
Peak at 50mg	314.47	363.04
Peak at 100mg	201.90	484.58

Summary

- Best depends on relative expected likelihood of different scenarios.
- Say we think
  - 40% Null
  - 40% Weak
  - 20% Positive (4% each of 5 scenarios)
- Probability weighted eNPV
  - Fixed: $65M
  - Adaptive: $109M

Summary & Conclusion

- The adaptive design looks superior across the board in this comparison
- It would require greater investment upfront, but it is likely that only if the expectation that the drug is no good is very high will the fixed phase 2a make sense.
- It may be that there are superior fixed designs that could be considered, but it is not immediately obvious what those would look like.
- Unless there is a clear endpoint that, tested at a single dose, would reduce the risk of developing a compound that lacks efficacy, it would seem that the phase 2a is a waste of time and resources that would be better invested in the phase 2b.

INTRODUCING ENPV ESTIMATION

Open and quantified

- To be trusted the simulation needs to be explorable – users need to be able to tweak bits of the model and see what changes
- Simulation of a process helps bring teams together and break down silos
- Providing concrete predicted results enables strategies to be optimized and ranked

Support for complex decisions

- Important, complex decisions in organizations are usually the subject of human cognitive biases
- Bias = a predictable error
- Quantified predictions “are only as good as the inputs” but allow:
  - the consequences of the inputs to be understood and the inputs modified
  - sensitivity analysis: which inputs matter?

Generating a tool

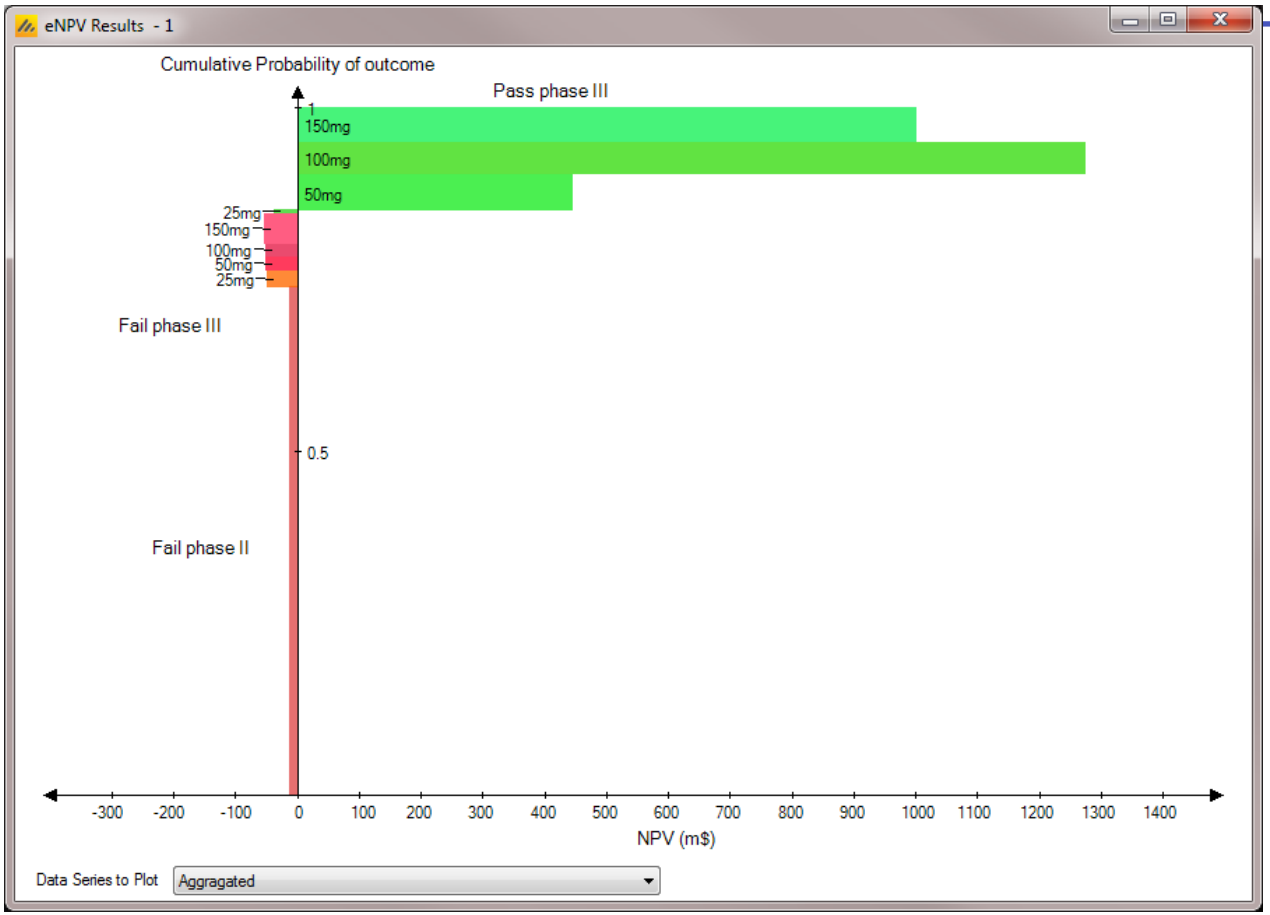
- A number of options:
  - Take a tool such as FACTS and extend it
  - Build bespoke scripts to tackle specific questions and learn
  - Take the output of trial simulations and make suitable for use in standard decision tools
  - Build a pluggable framework allowing arbitrarily complex development paths, and trial sim from different sources to be loaded

In decision support

- The most crucial thing to understand is the decisions that need to be supported
- Sounds trivial, but its surprising how often the initiatives become tool or technology driven and then not delivering any benefit in the end.

LOOKING AT ENPV RESULTS

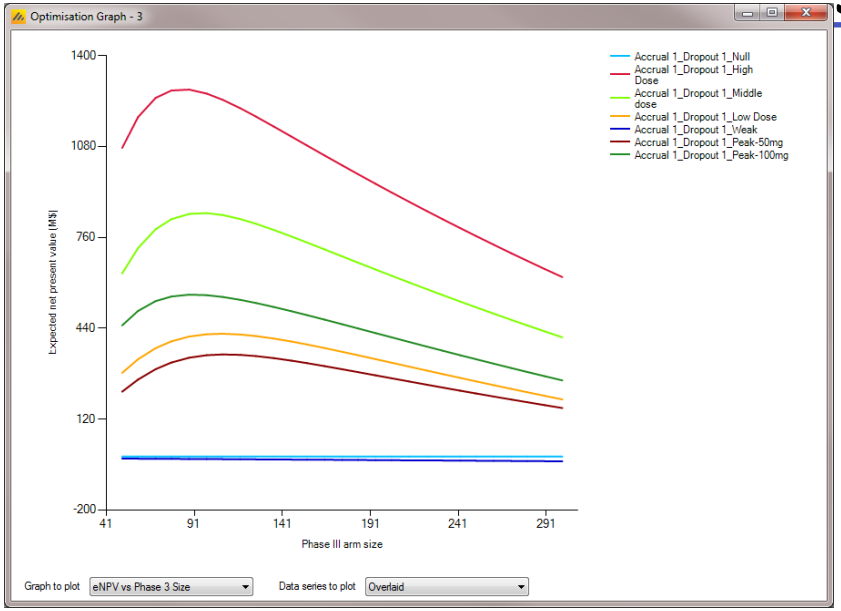
Results Graphs

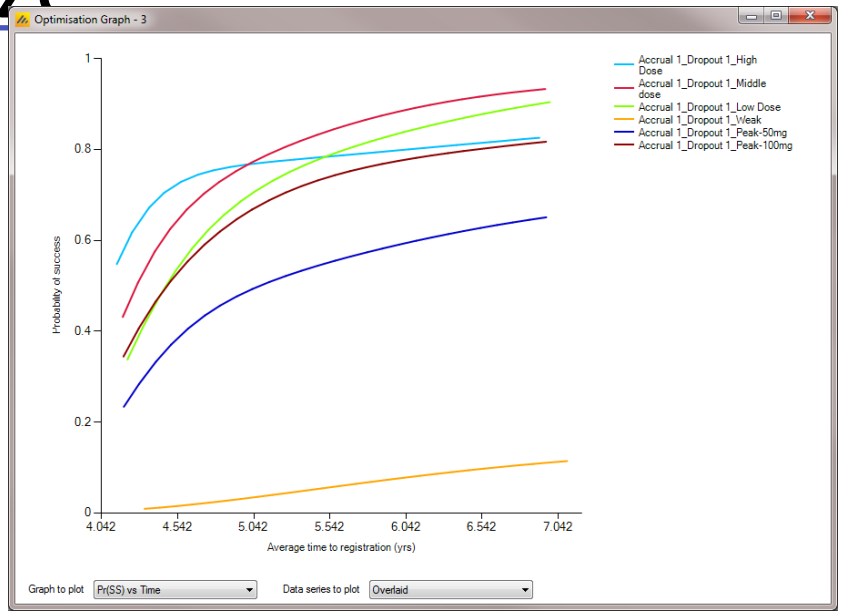


The graph plots bars for each outcome sub-divided by dose selected at phase 2

The bar's height is its probability (proportion of occurrences ion the scenarios) and direction, colour and length is the eNPV of the outcome

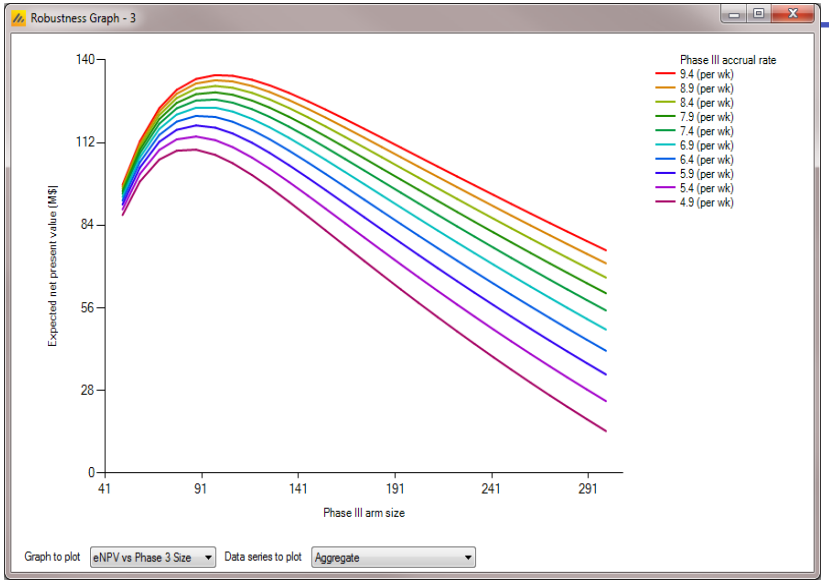
Illustration of optimisation – phase 3 size

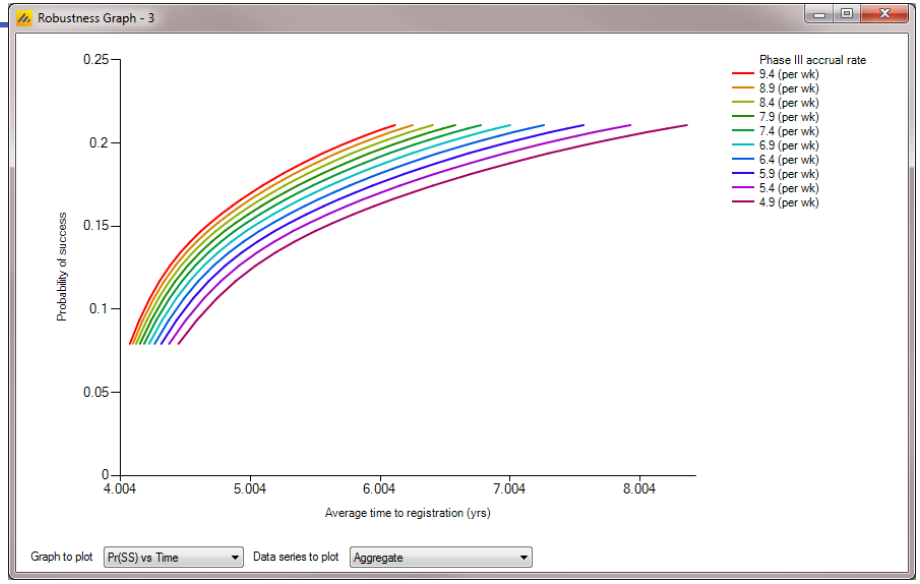




By re-calculating at different phase 3 sizes we can look how eNPV changes with phase 3 size or the probability of success changes with time taken or cost of development.
Here separate lines are being shown for each scenario.

Sensitivity Analysis

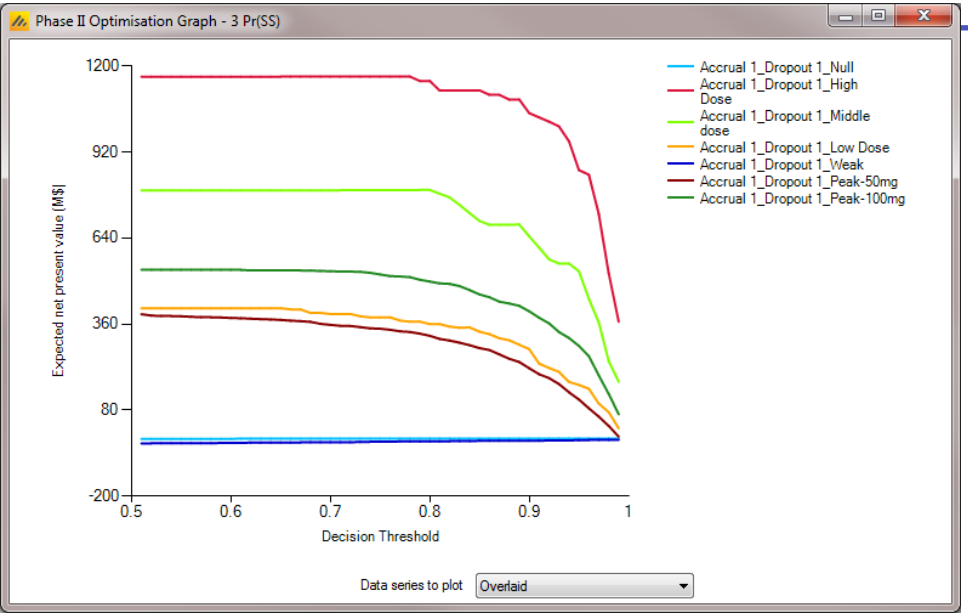


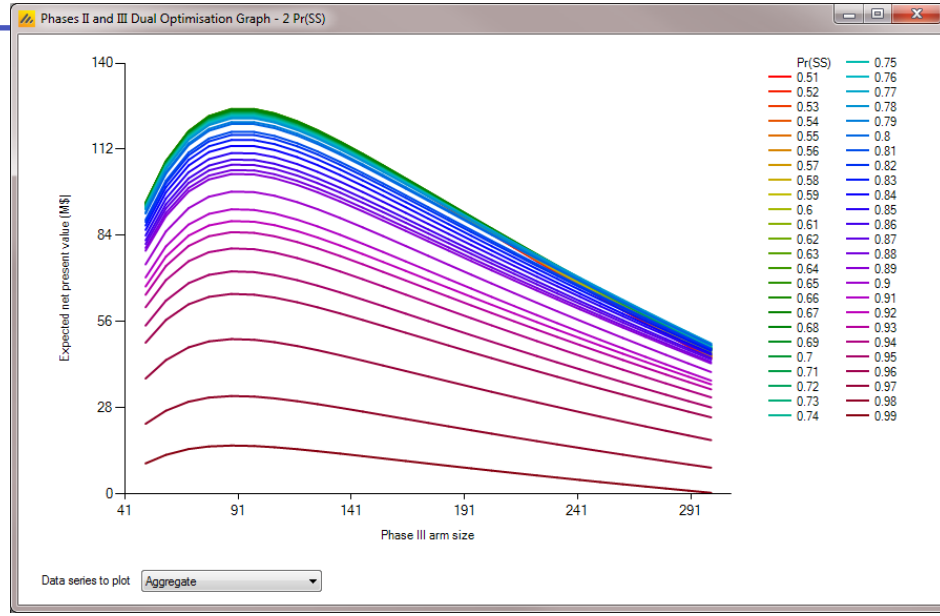


Re-calculating for different phase 3 sample sizes is here combined with varying the accrual rate, and we can look again at eNPV vs phase 3 sample size or probability of success against time.

Here aggregated results (summed over the scenarios) are being shown with different lines for each of the accrual rates being evaluated.

Optimising phase 2 decisions





Whilst many different phase 2 designs will require re-running simulations and separate analysis, some different options can be analysed statically (and thus very quickly).
Here the 'end of phase 2' dose selection and go/no-go decision criteria are explored and also combined with possible different phase 3 trial sizes.