

White Paper

An Example Adaptive Phase 2 Clinical Trial Design for HIV using FACTS

Tom Parke

Head of Clinical Trial Solutions, Tessella plc

An Example Adaptive Phase 2 Clinical Trial Design for HIV using FACTS

This is an example of a trial design for a phase 2 trial in HIV. The primary endpoint, which will be used for final decision-making and to drive the adaptation, will be no detectable viral load after 3 months and the treatment tolerated for the full three months.

The outcome is dichotomous, subjects starting with observable viral load at baseline and monitored after 1, 2 and 3 months. Effect and tolerability endpoints could have been analysed separately but here they are combined in a single score. A nice consequence of this is that subjects that drop out (for lack of tolerability) still contribute to the analysis.

The aim of the trial is to study 4 different doses of the study drug, comparing their effectiveness to a 'standard of care' control arm and determine if there is a good chance of being able to successfully run a subsequent phase 3 trial. As dose tolerability is included in the endpoint, the clinical team simply want to find the dose with the greatest response. However, because tolerability is included in the endpoint a monotonic response cannot be assumed, the highest dose may be less well tolerated and so score worse than a lower dose. Hence any dose response modelling cannot use a monotonic model.

Standard Study

The expected response rate in subjects on the standard of care is expected to be 60%, a conventional trial size calculation gives 80 subjects per arm to achieve a two-sided alpha of 0.05 and a power of 88.5% for a rate of 80% on the study drug. So a total study size of 400 subjects is reasonable.

Dose Response Modelling

Let R_i be the final outcome for subject i who has been assigned to treatment d_i . Let θ_d be the mean log odds response for R_i when $d_i = d$. i.e.:

$$\theta_d = \log\left(\frac{P_d}{1-P_d}\right)$$

where:

$$P_d = \Pr(R_i = 1 | d_i = d)$$

A first order NDLM model is used to model efficacy. The model is defined as follows:

Let doses $d=0, 1, 2, \dots, D$ be doses in the dose response model, and U_0, U_1, \dots, U_D the nominal dose strengths for modeling purposes of each dose. The NDLM is defined with the following assumptions:

$$\theta_0 \sim N\left(\log\left(\frac{0.6}{0.4}\right), 0.5^2\right)$$

Thus the 95% interval for θ_0 is (0.36-0.80) with a mean of 0.6.

$$\theta_d \sim N(\theta_{d-1}, \tau_{d-1}^2) \text{ for } d=1, \dots, D,$$

where:

$$\tau_d^2 = \tau^2 (v_{d+1} - v_d)$$

The prior distribution for the "drift" parameter in the NDLM is:

$$\tau^2 \sim IG\left(\frac{\tau_n}{2}, \frac{\tau_\mu^2 \tau_n}{2}\right)$$

where τ_μ is the prior mean for τ , and τ_n are the "number of observations" on which it is based.

(Here the gamma distribution is taken to be $f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}$)

Which we set to $\tau^2 = 0.64$ and $\tau_n = 1$ giving

$$\tau^2 \sim IG(0.5, 0.32)$$

Earlier Data

As well as the final 12th week score, subjects will be assessed at 4 and 8 weeks, and until a subject's final score is available, their earlier score will be used to impute their final score. If subjects drop out their final score can also be imputed in this manner. The imputed values are treated as random variables to be estimated and included in the Bayesian analysis with the appropriate amount of uncertainty (whereas an actually observed value has no uncertainty).

A simple Beta-Binomial model is used to model $Pr(R_i = 1 | r_{it} = 1)$ and $Pr(R_i = 1 | r_{it} = 0)$ or the visits $t=4$ and $t=8$. Separate models are built for control and for the treatment arms and for each visit.

With priors:

$$Pr(R_i = 1 | r_{it} = 1) \sim \text{Beta}(6, 1)$$

$$Pr(R_i = 1 | r_{it} = 0) \sim \text{Beta}(2, 2)$$

Estimated Probabilities

From the combined longitudinal and dose-response models, when fitted to the available response data, two key probabilities are assessed:

- For each dose, the probability that it is the dose with maximum effect [$Pr(d = d_{Max})$]
- For the dose with the maximum response, the predictive probability that a subsequent phase III trial will be successful given 250 subjects and a single-sided alpha of 0.025. [Calculating the 'predictive probability' means that the probability of success is not simply calculated at the point estimate of the response at d_{Max} , but integrated over the full uncertainty of that estimate.]

Decision Making

For the purposes of assessing this design (in particular to assess its type-1 error rate and power) the final evaluation of the trial data will be defined in terms of the probability of success in a subsequent phase 3 trial.

If $Pr(\text{Success in Phase III}) < 0.5$ then we will deem the compound to have failed, and no phase 3 will be run.

If $Pr(\text{Success in Phase III}) > 0.8$ then the compound's efficacy will have been sufficiently established that a phase 3 trial will be run. In practice, there will also be safety, regulatory and commercial considerations taken into account. This is no different from the conventional case where the power of a design is estimated based simply on the probability of achieving statistical significance on the efficacy endpoint.

Trial implementation

It is expected that the trial will recruit at a rate of about 5 subjects a week (20 a month) and that it will take the first 12 weeks for the recruitment to ramp up to reach this rate.

Viral load and drug tolerability data will be collected from the investigators quickly and processed by the adaptive algorithm every 4 weeks.

Early Stopping

When the current response data is processed by the algorithm every 4 weeks, after each evaluation of the model, if at least 150 subjects have been recruited, the trial can be terminated if its final outcome looks no longer in doubt. Specifically:

If 150 or more subjects have been recruited and $Pr(\text{Success in Phase III}) < 0.15$ then it is so unlikely that, the final result could be successful (>0.8) that the trial will be terminated early for futility.

If 150 or more subjects have been recruited and $Pr(\text{Success in Phase III}) > 0.95$ then the trial is sufficiently certain to be successful that it will be terminated early for success.

Adaptive Allocation

Initially subjects will be allocated to the study arms in fixed ratio, with 20 on the control arm and 10 on each of the arms of the study drug. These first 60 subjects are expected to take about 18 weeks to recruit by which time the first 8-9 subjects should have completed and final results for successive subjects starting to arrive regularly. In addition there should be around a further 14-15 subjects with both first and second month results, and nearly 20 with just first month results.

After this fixed allocation period and every 4 weeks after that, the randomisation will be adjusted. 2 subjects in every 6 will be allocated to control and the remaining 4 will be allocated between the study arms in proportion to the current assessed probability that that arm is using the dose with the maximum response.

Simulated Scenarios

The following dose response scenarios were used to initially evaluate the design:

Scenario	Dose				
	Control	Dose 1	Dose 2	Dose 3	Dose 4
Flat	0.6	0.6	0.6	0.6	0.6
Dose 3 Max	0.6	0.6	0.65	0.78	0.7
Dose 4 Max	0.6	0.65	0.7	0.75	0.8

The Flat scenario allows us to estimate the equivalent of the frequentist type-1 error, the Dose 3 Max and Dose 4 MED scenarios allow us to estimate the power and ability to determine the correct dose with the maximum response.

Longitudinal results were simulated based on probabilities of transition from no response to response of 0.6, 0.4 and 0.2 at visits 1, 2 and 3. An probabilities of remaining a responder if already a responder of 0.8 and 0.9 at visits 2 and 3.

These probabilities give an overall probability of response of 0.648. For each dose in each scenario these transition probabilities were modified by a fixed amount in the log-odds to give the overall probability of response required for that scenario. For example subtracting 0.145 in log odds from each of these gives probabilities of 0.565, 0.366 and 0.178 for transition and 0.776 and 0.886 for remaining resulting in an overall probability of response of 0.60.

For comparison two fixed trial designs were considered, both with a total sample size of 432. The first trial with the same 6 doses, 108 subjects allocated to control and 54 subjects to each study arm, and the second trial with 3 doses (doses 2, 4 & 6) and 108 subjects allocated to control and each arm. The analysis is a simple pair-wise comparison using a Bonferroni adjusted p-value, the lowest dose to be statistically significant being taken to be the MED and the trial to be successful if any one dose is statistically significant and futile otherwise.

Results

First looking at average sample size, type 1 error and power to detect a successful compound:

Scenario	3 Dose Fixed		6 Dose Fixed		6 Dose Adaptive		
	Mean sample size	% correct futile / success	Mean sample size	% correct futile / success	Mean sample size	% correct futile / success	% inconclusive
Flat	432	97.5%	432	97.5%	220	99.4%	0.6%
Weak	432	78.3%	432	85.0%	355	72.5%	22.2%
Dose 2	432	98.8%	432	96.5%	343	93.3%	6.4%
Dose 3	432	96.7%	432	93.3%	365	91.5%	7.7%
Dose 4	432	94.3%	432	86.6%	382	87.2%	11.8%
Dose 5	432	83.8%	432	73.4%	395	83.6%	13.2%
Dose 6	432	72.1%	432	46.9%	403	57.7%	31.0%

The results for the 6 Dose Adaptive design are collected over 1,000 simulations (which took ~90 minutes on a dual core laptop).

The design has better type-1 error control than the fixed designs along with the ability to, on average, save over 200 subjects and thus nearly 50% of the direct grant costs if the compound is not successful at all. If we take an inconclusive result to be futile, then the adaptive design has similar power to the 6 dose fixed design but better ability to reject the weak response and a 10-20% saving in sample size and direct grant costs. [Not stopping early for success increases the selection of the correct MED by between 0-5%, this difference is sufficiently small that a far greater number of simulations need to be run in order to be sure of it].

The following table shows, for each success scenario, the % of times the trial is successful and identifies the correct dose as the MED. For a phase 2 trial this is perhaps a more useful estimate of its 'power' than simply detecting significance.

Scenario	3 Dose Fixed		6 Dose Fixed		6 Dose Adaptive	
	% correct MED	% correct MED or MED+1	% correct MED	% correct MED or MED+1	% correct MED	% correct MED or MED+1
Dose 2	70.7%	70.7%	38.7%	67.1%	56.8%	75.7%
Dose 3	0.0%	65.7%	37.0%	64.0%	51.6%	70.7%
Dose 4	68.2%	68.2%	36.6%	63.5%	48.0%	69.6%
Dose 5	0.0%	64.6%	36.4%	63.1%	54.1%	70.7%
Dose 6	67.4%	67.4%	36.3%	36.3%	47.2%	47.2%

The adaptive design is markedly better at determining the exact MED out of 6 doses, the advantage is reduced if the selecting the MED or the next highest dose as the MED is acceptable. The only scenario in which the adaptive 6 dose design does not out perform the fixed 3 dose design is when Dose 6 is MED, which is also the scenario when there is only 1 successful dose and its effect size is not as large as the best doses in the other scenarios.

Note that in the 3 dose fixed trial design, in the 'dose 2' scenario there is a risk that the trial will have to be repeated due to there being no lower ineffective dose.

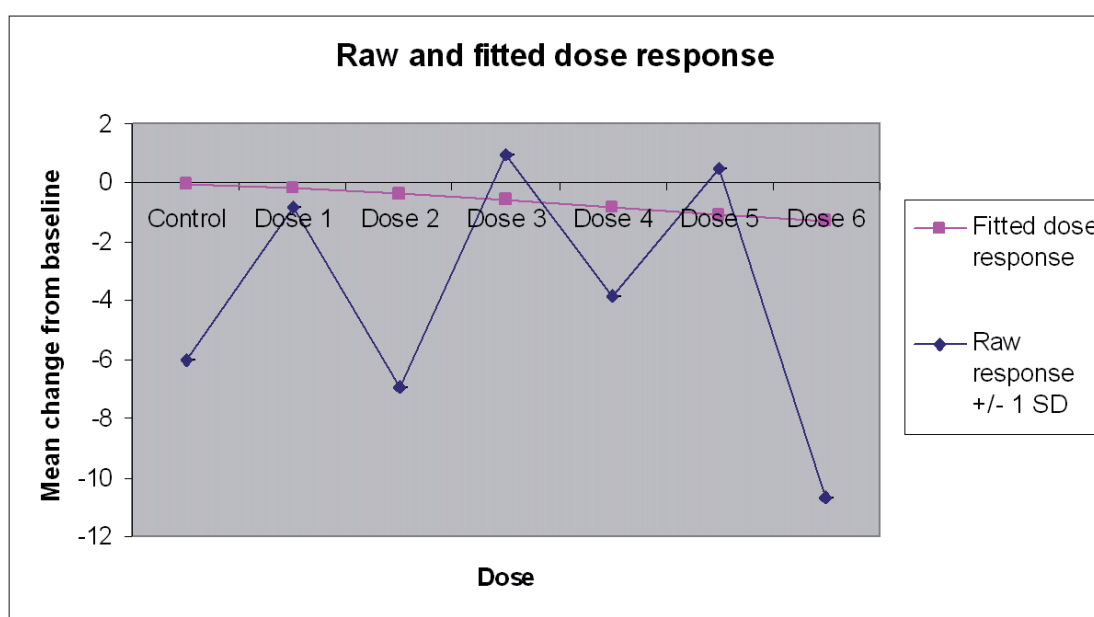
Example Adaptive Trial

The following is the first trial from the Dose 3 Scenario simulations which terminates early and selects the correct MED. Below are some details of the how the method performs at a few adaptive updates during the trial

At week 15 at the end of the fixed allocation with only one subject completed from each arm:

	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	20	10	10	10	10	10	10
Number of subjects completed	1	1	1	1	1	1	1
Probability of allocation	0.00	0.07	0.13	0.17	0.20	0.26	0.17
Fitted mean response	-0.04	-0.17	-0.35	-0.58	-0.85	-1.10	-1.26
Raw mean response	-6.00	-0.81	-6.95	0.94	-3.85	0.52	-10.67
SD of raw response	0.00	0.00	0.00	0.00	0.00	0.00	0.00

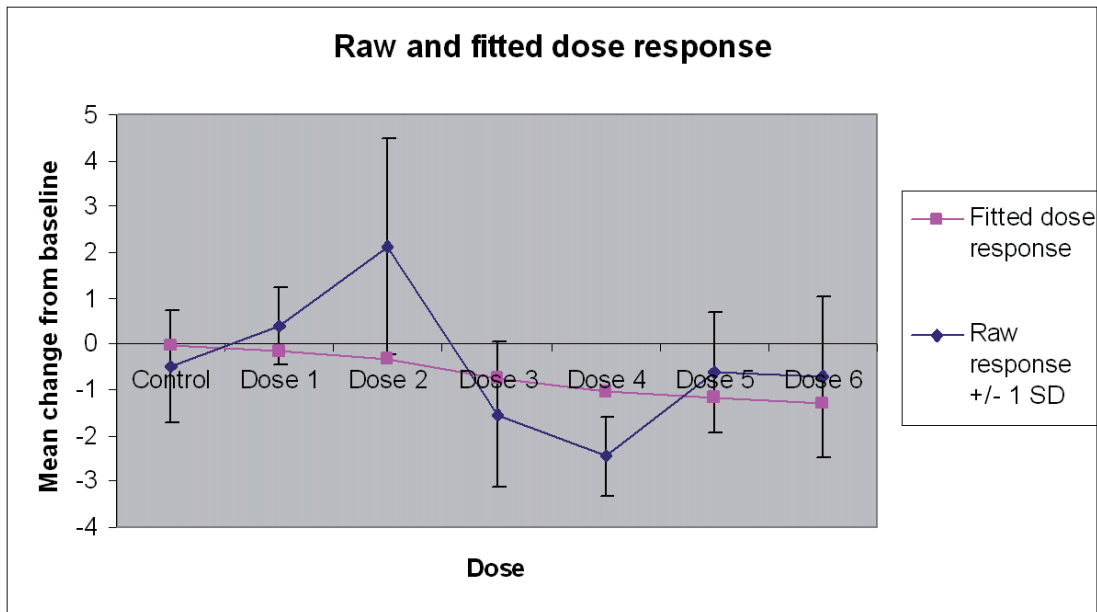
Note that the fitted response is also taking into account early responses from subjects who have not completed yet.



10 weeks later at week 25, 149 subjects recruited, 65 completed:

	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	38	12	12	17	22	23	25
Number of subjects completed	16	8	9	8	8	8	8
Probability of allocation	0.00	0.00	0.00	0.27	0.39	0.21	0.13
Fitted mean response	-0.02	-0.14	-0.32	-0.75	-1.04	-1.18	-1.28
Raw mean response	-0.50	0.39	2.12	-1.54	-2.44	-0.62	-0.72
SD of raw response	1.22	0.84	2.37	1.58	0.86	1.32	1.75

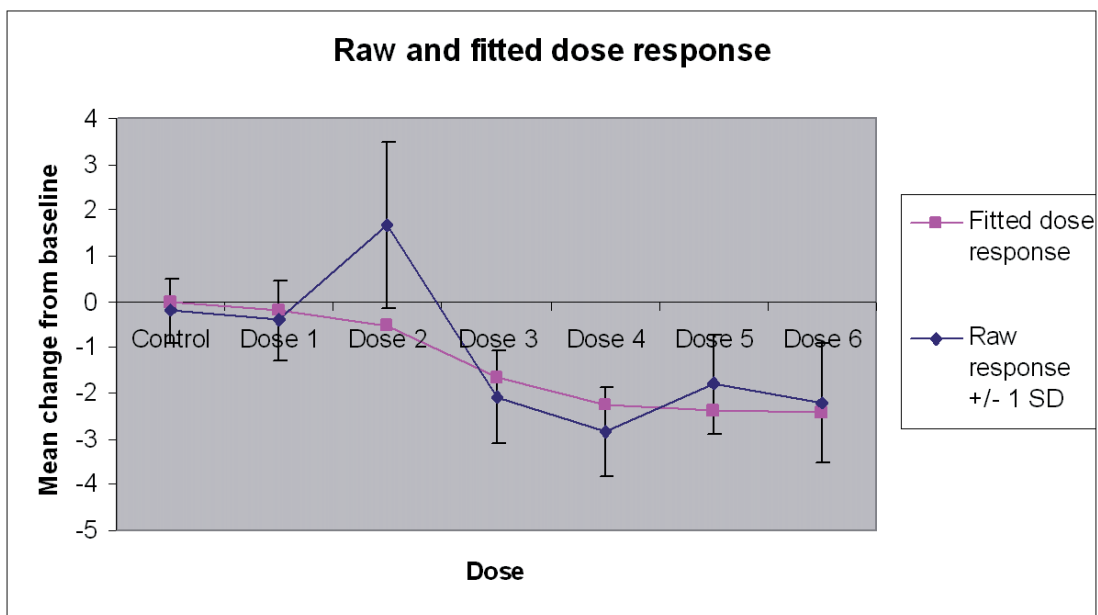
Note that there is already a useful bias in allocation to not allocate to doses 1 or 2 and to weight allocation to predominantly doses 3, 4 and 5. Note that although the probability of allocation to control is shown as 0, control is allocated to separately as 2 doses out of every block of 8, the probability of allocation shown here controls how the remaining 6 doses in the block are allocated.



10 weeks later at week 35, 219 subjects recruited, 135 completed:

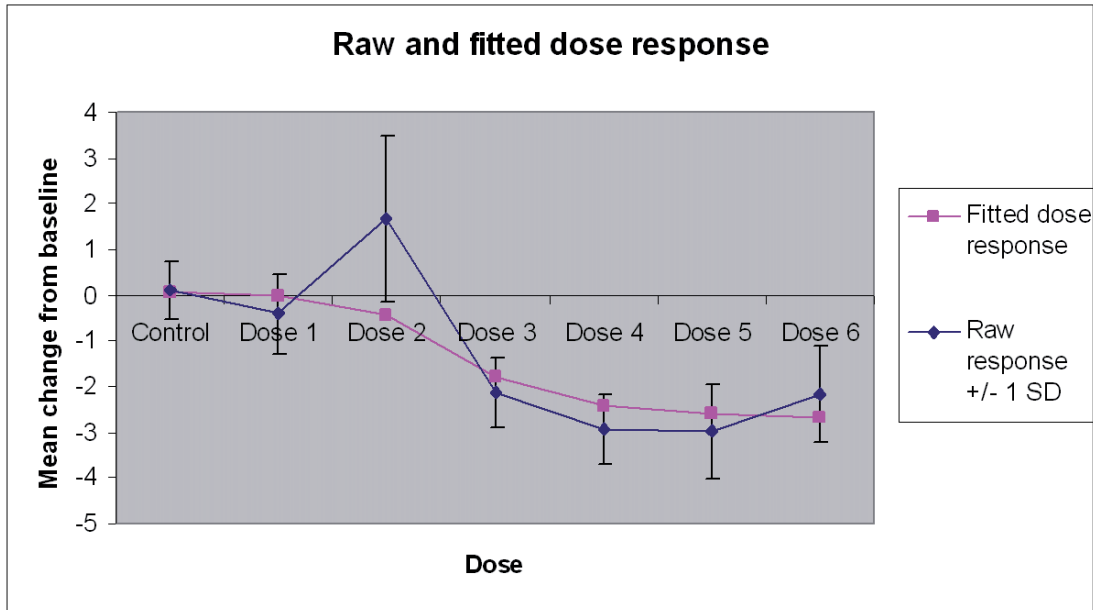
	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	55	12	12	44	41	28	27
Number of subjects completed	34	12	12	16	21	19	21
Probability of allocation	0.00	0.00	0.05	0.50	0.38	0.07	0.00
Fitted mean response	-0.03	-0.16	-0.53	-1.68	-2.24	-2.39	-2.44
Raw mean response	-0.19	-0.42	1.68	-2.07	-2.85	-1.81	-2.20
SD of raw response	0.69	0.88	1.83	1.02	0.96	1.06	1.32

Notice how out of the study arms, the allocation is predominantly to doses 3 & 4, and for the next period the probability of allocation to dose 3 is 50%.



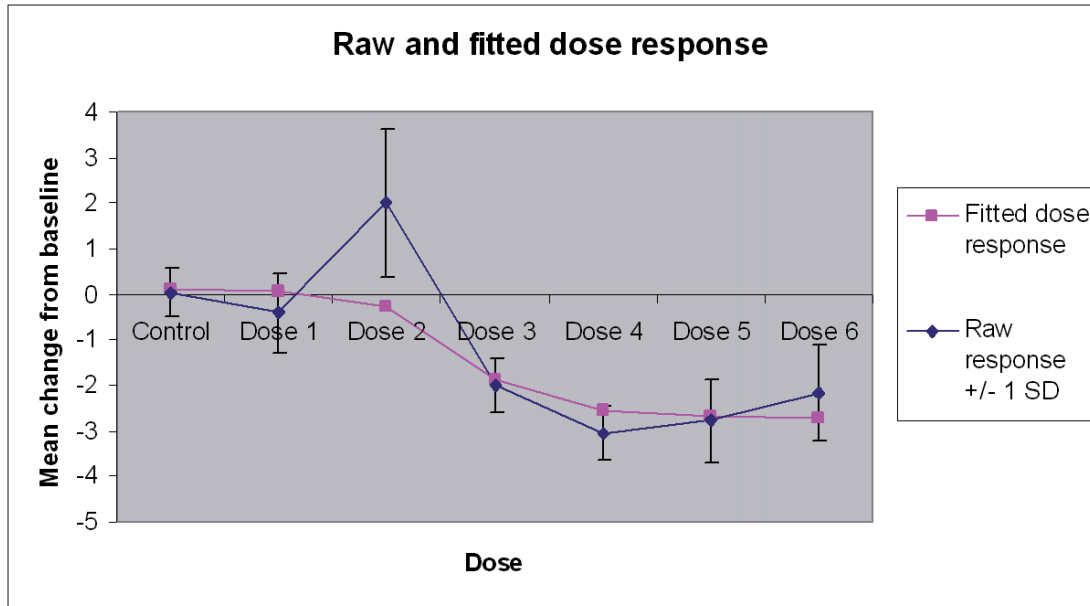
10 weeks further still, at week 45, 289 subjects recruited 205 completed, the design stops the trial for success:

	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	72	12	14	64	66	34	27
Number of subjects completed	52	12	12	39	35	28	27
Probability of allocation	0.00	0.00	0.06	0.54	0.34	0.06	0.00
Fitted mean response	0.09	-0.02	-0.44	-1.79	-2.42	-2.59	-2.66
Raw mean response	0.12	-0.42	1.68	-2.12	-2.93	-2.99	-2.17
SD of raw response	0.63	0.88	1.83	0.76	0.78	1.05	1.06



12 weeks later, at week 57, 289 subjects recruited and completed:

	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	72	12	14	64	66	34	27
Number of subjects completed	72	12	14	64	66	34	27
Probability of allocation	0.00	0.00	0.00	0.66	0.34	0.00	0.00
Fitted mean response	0.10	0.06	-0.26	-1.86	-2.56	-2.69	-2.73
Raw mean response	0.05	-0.42	2.00	-1.98	-3.06	-2.77	-2.17
SD of raw response	0.53	0.88	1.62	0.59	0.59	0.90	1.06



Tessella plc 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
T: +44 (0)1235 555511 | F: +44 (0)1235 553301 | E: info@tessella.com

Tessella Inc 233 Needham Street, Suite 300, Newton, MA 02464, USA
T: 1 617 454 1220 | E: info@tessella.com



Tessella – successfully delivering IT and consulting services to world leaders in R&D, science and engineering.

For decades, Tessella has been successfully delivering IT and consulting services to world leaders in R&D, science, and engineering. Through the application of scientific methods and rigorous quality procedures, we enable clients in life sciences, energy, the public sector, and consumer industries to achieve a wide range of objectives, including, forecasting floods, developing fusion power, enhancing military sensor capability, improving drug discovery and development efficiency, and reducing risk to health and the environment in the extraction and production of oil and gas. With offices in Europe and North America, global companies rely on Tessella for business critical assignments.

Copyright © Tessella plc 2010, all trademarks acknowledged. Issue: V1.R0.M0 | Apr-10

www.tessella.com

