

White Paper

An Example Adaptive Phase 2 Clinical Trial Design for Alzheimer's using FACTS

Tom Parke

Head of Clinical Trial Solutions, Tessella plc

An Example Adaptive Phase 2 Clinical Trial Design for Alzheimer's using FACTS

This is an example of a trial design for a phase 2 trial in Alzheimer's. The primary endpoint, which will be used for final decision-making and to drive the adaptation, will be the change from baseline in subjects' ADAS-cog score after 12 weeks.

It is determined that the minimum clinically significant difference (CSD) compared to placebo would be a 1.75 points greater reduction in ADAS-cog score. The clinical team believe that a credible good performance of the compound would be to achieve 2.2 points greater mean reduction than the control arm.

The aim of the trial is to study 6 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 a response that exceeds that on control by the CSD. In this trial there is not the time or number of subjects to study long term tolerability or safety, so, to inform our choice of dose (or doses) for phase 3, we wish to estimate the minimum efficacious dose (MED) – that is, the lowest dose that beats control by the CSD.

Standard Study

The ADAS-cog endpoint is expected to have a standard deviation of 5 points, a conventional trial size calculation to give a two-sided alpha of 0.05, a power of 80% to detect a difference of 2.2 points between a single study arm and the control gives a size of 81 per arm (before adjusting for multiplicity). Using a simply Bonferroni adjustment for multiplicity, to study 6 doses would require 125 per arm and a total study size of 875, or to study 3 doses would require 108 per arm and a total study size of 432.

For our adaptive design we will limit our study size to 432 subjects but study 6 doses, we believe that, through dose response modelling and adaptive allocation, we can study this increased number of doses whilst retaining the same alpha and power. By including more doses, we allow the study to include a dose strength that is almost certainly too weak, thus avoiding regulatory concern that the MED is not truly identified if our selected MED turns out to be the lowest dose included in the study. It will also enable us to more accurately characterise the MED.

Dose Response Model

It is strongly expected that the mean response – change in ADAS-cog - (if there is any) will increase with increasing dose, the mean response may plateau out, but it will not decrease at higher doses. Thus, we chose a 4-parameter sigmoid function to model the response θ_d at dose d :

$$\theta_d = a_1 + \frac{(a_2 - a_1)v_d^{a_4}}{a_3^{a_4} + v_d^{a_4}}$$

Where v_d is the effective dose strength of dose d .

Each parameter can be interpreted in terms of the dose response curve

a_1 is the response at the control arm

a_2 is the maximum response (possibly outside the range of doses being tested) so $(a_2 - a_1)$ is the maximum improvement over control.

a_3 is the dose strength at which half the maximum improvement is seen (and where the gradient is steepest)

a_4 determines the gradient of the response, the gradient when the dose strength = a_3 is :

$$(a_2 - a_1) \frac{a_4}{4a_3}$$

Given this interpretation we use the following priors:

$a_1 \propto N(0, 0.5^2)$ so the 2.5 and 97.5 percentiles of a_1 are: -0.98, 0.98

$a_2 \propto N(-1, 2^2)$ so the 2.5 and 97.5 percentiles of a_2 are: -4.92, 2.92

$a_3 \propto N^+(3, 3^2)$ so the 2.5 and 97.5 percentiles of a_3 are: 0, 8.88

$a_4 \propto N^+(6, 6^2)$ so the 2.5 and 97.5 percentiles of a_4 are: 0, 17.8

Whilst the means do not constitute a flat curve, a flat curve is so easily obtained from these distributions that they do not to inflation of the type-1 error, as can be seen from the simulation results below.

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).

Simple linear regression is used to create longitudinal models of the relationship between the week 4 scores and the final scores, and the week 8 scores and the final scores.

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 minimum efficacious dose (MED). [$\Pr(d = d_{MED})$]

For the dose with the maximum response, the probability that the mean response at that dose is better than that of the control arm by at least the clinically significant difference (CSD). [$\Pr(\theta_{d_{max}} - \theta_0 < -1.75)$]

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 that the response at the maximum dose is better than control by more than the CSD.

If $\Pr(\theta_{d_{max}} - \theta_0 < -1.75) < 0.1$ then we will deem the compound to have failed, and no phase 3 will be run. If $\Pr(\theta_{d_{max}} - \theta_0 < -1.75) > 0.4$ 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 7 subjects a week (28-30 a month) and that it will take the first 12 weeks for the recruitment to ramp up to reach this rate.

ADAS-cog data will be collected from the investigators quickly and processed by the adaptive algorithm every 2 weeks. In order to ensure the maximum amount of data is available, the data provided will comprise the cleaned data processed and stored in the EDC system supplemented where that data is not yet available by the initial responses supplied by the investigators by faxing in the completed ADAS-cog score sheets.

Early Stopping

When the current response data is processed by the algorithm every 2 weeks, after a required minimum number of subjects has been recruited, the trial can be terminated if its final outcome looks no longer in doubt. Specifically:

After 160 subjects have been recruited, if $\Pr(\theta_{d_{max}} - \theta_0 < -1.75) < 0.01$ then it will be so unlikely that, the final result could be successful that the trial will be terminated early for futility.

After 160 subjects have been recruited, if $\Pr(\theta_{d_{max}} - \theta_0 < -1.75) > 0.8$ **and** there is a dose sufficiently likely to be the MED: $\exists d \in \{D\}: \Pr(d = d_{MED}) > 0.5$ then the trial is sufficiently certain to be successful and the MED found 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 6 arms of the study drug. These first 80 subjects are expected to take about 18 weeks to recruit by which time the first 10 or so subjects should have completed and final results for successive subjects starting to arrive regularly. In addition there should be around a further 20 with first and second month results and nearly 30 with first month results.

After this fixed allocation period and every 2 weeks after that, the randomisation will be adjusted. 2 subjects in every 8 will be allocated to control and the remaining 6 will be allocated between the study arms in proportion to the current assessed probability that that arm is the MED.

Simulated Scenarios

The following dose response scenarios were used:

Scenario	Dose						
	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Flat	0	0	0	0	0	0	0
Weak	0	-0.05	-0.1	-0.2	-0.4	-0.8	-1.0
Dose 2 MED	0	-0.8	-2.0	-2.2	-2.2	-2.2	-2.2
Dose 3 MED	0	-0.4	-1.0	-2.0	-2.2	-2.2	-2.2
Dose 4 MED	0	-0.2	-0.4	-1.0	-2.0	-2.2	-2.2
Dose 5 MED	0	-0.1	-0.2	-0.4	-1.0	-2.0	-2.2
Dose 6 MED	0	0	-0.1	-0.2	-0.4	-1.0	-2.0

The Flat scenario allows us to estimate the equivalent of the frequentist type-1 error, the Dose 2-6 MED scenarios allow us to estimate the power and ability to determine the correct MED. The Weak scenario allows us to estimate the ability to determine that though there is a response, that it is insufficient to justify further development of the compound.

Longitudinal results were simulated with 50% and 80% of the final response being seen at the first and second months visits, with 75% and 90% correlation respectively with the subject's final result.

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.

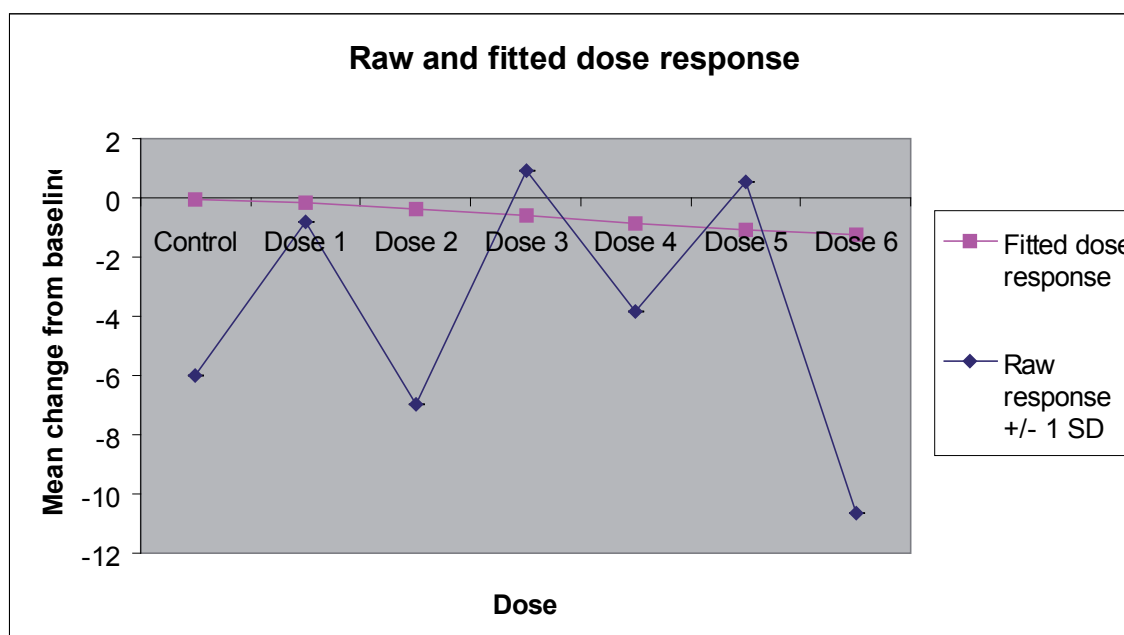
Example Adaptive Trial

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

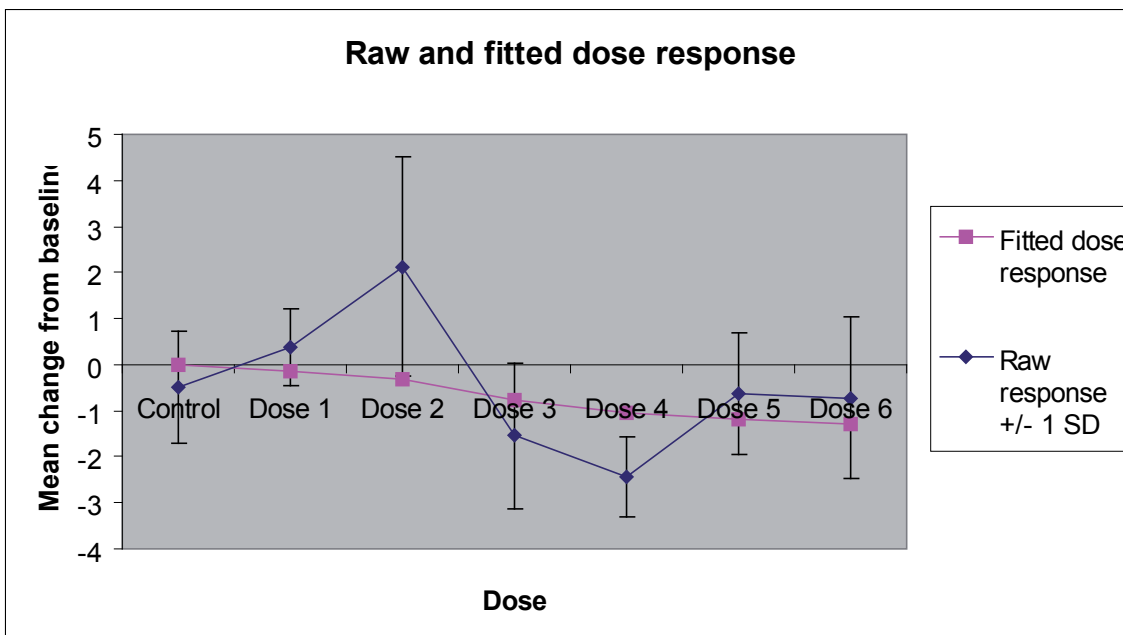
Note that the fitted response is also taking into account early responses from subjects who have not completed yet.



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

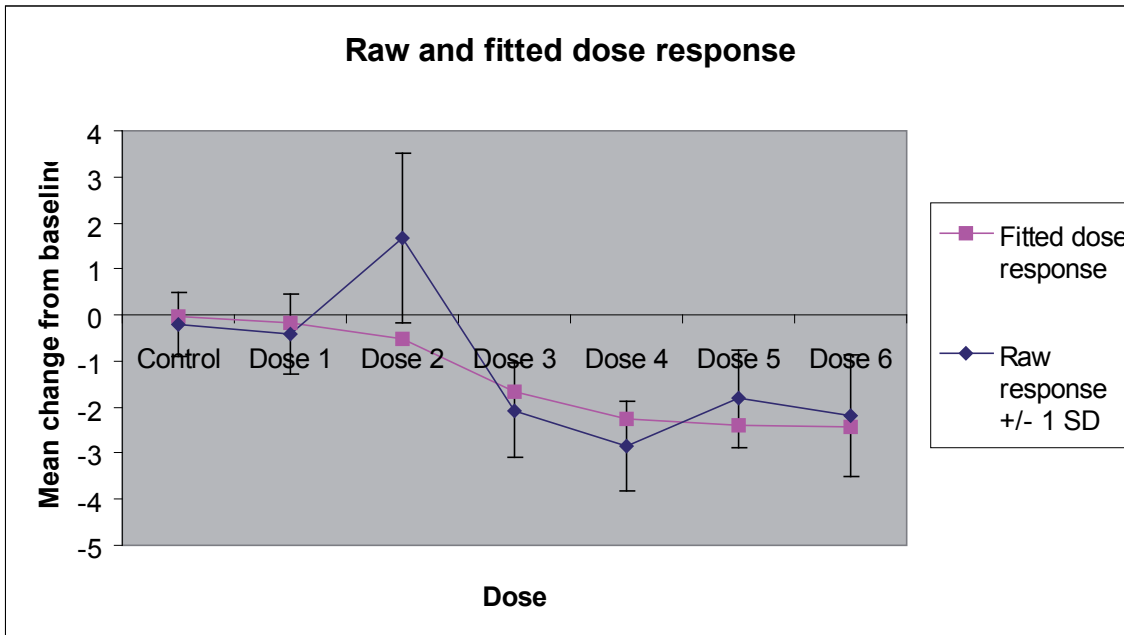
Note that there is already a useful bias in allocation to not allocate to doses 1 or 2 and to weight allocation to predominantly doses 3, 4 and 5. Note that although the probability of allocation to control is shown as 0, control is allocated to separately as 2 doses out of every block of 8, the probability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

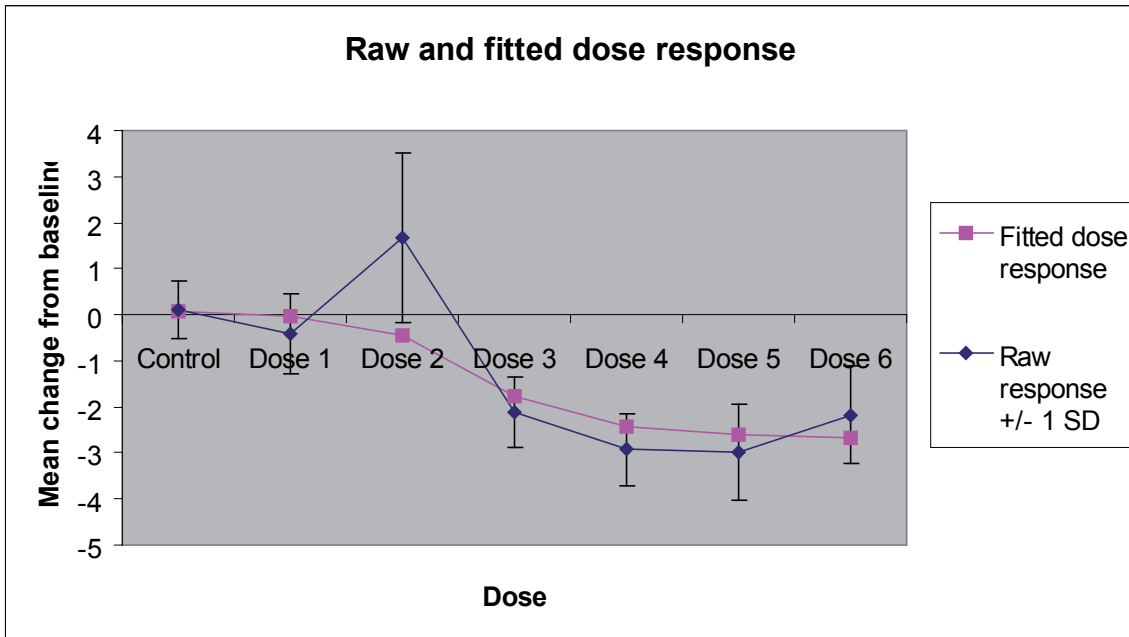
	Control	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Number of subjects allocated	55	12	12	44	41	28	27
Number of subjects completed	34	12	12	16	21	19	21
Probability of allocation	0.00	0.00	0.05	0.50	0.38	0.07	0.00
Fitted mean response	-0.03	-0.16	-0.53	-1.68	-2.24	-2.39	-2.44
Raw mean response	-0.19	-0.42	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%.



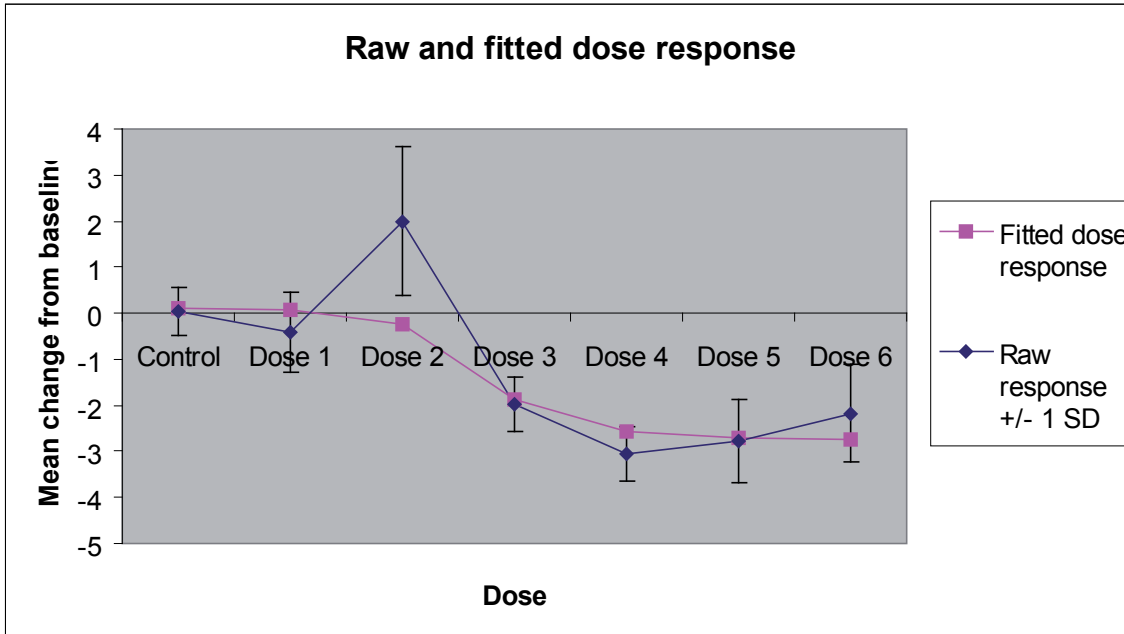
10 weeks further still, at week 45, 289 subjects recruited 205 completed, the design stops the trial for success:

	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 2009, all trademarks acknowledged. Issue: V1.R0.M0 | Jan-10

www.tessella.com

