Trial Sample Size Estimation and Assessment of Study Design using Monte Carlo Sampling

Claus Andersen and Göran Westerberg, Siena Biotech SpA, Strada del Petriccio e Belriguardo 35, 53100 Siena (SI), Italy

I-8

Highlights

Objective:

➢ Construct a simulated clinical trial as close to clinical reality as possible benefitting from available clinical information.

Investigations:

➢Trial design
- ✓ Inclusion criteria
- ✓ Balanced *vs.* unbalanced patient groups (treated, placebo)
- ✓ Statistical assessment criteria

➢Sample size and statistical power

Results:

➢Change in inclusion criteria (excluding patients with TFC=13)

➢Balanced design (same number of placebo as treated patients) increases power with respect to unbalanced designs.

➢ANCOVA with baseline TFC as covariate was optimal

➢Sample size estimates showed that a 92% power could be obtained with 444 enrolled patients assuming a 10% drop out rate and a 40% treatment effect.

➢ These investigations were possible with a Monte Carlo setup in contrast to *e.g.* an analytical trial setup.

Primary end-point

Total Functional Capacity

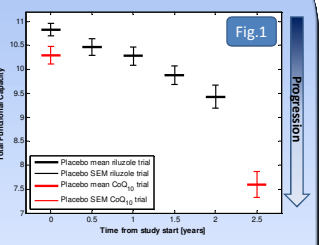
TFC = Σ function scores

Domain	Level of functioning	Score
Occupation	Normal Reduced capacity for usual job Marginal work only Unable	3 2 1 0
Finances	Normal Slight assistance Major assistance Unable	3 2 1 0
Domestic chores	Normal Impaired Unable	2 1 0
Activities of daily living	Normal Minimal impairment Gross tasks only Total care	3 2 1 0
Care level	Home Home with chronic care Full-time skilled nursing	2 1 0

TFC is assessed by a physician after patient/caretaker interview. Table adapted from Shoulson [1].

Disease progression

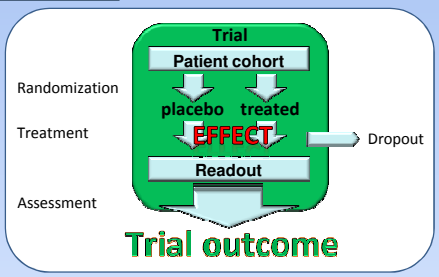
HD is a slowly progressing disease decreasing approximately 1 TFC point per year in placebo treated patients (see Fig. 1)[2,3], but the rate of decay depends on disease state. In particular, the highest TFC score = 13 covering manifest HD and slightly symptomatic subjects has been found to show a *ceiling effect*, where subjects may remain unchanged for several years [4]. This prompted a change in the inclusion criteria to exclude these subjects ensuring selection of a relevant population to conduct clinical trials.



The TFC measure is considered one of the best overall measures to assess disease progression[4], but only changes in integer steps, which alongside with the subjective assessment gives rise to significant variability. To ascertain a difference between treated and placebo groups, large patient groups are therefore required.

Monte Carlo setup

Trial design:



Patient Population:

➢ Initial inclusion criteria were: CAG repeat length ≥ 36, TMS ≥ 5, ambulatory, manifest HD, TFC ≥ 9, age ≥ 18. These criteria are similar to the riluzole trial and somewhat more stringent than *e.g.* the remacemide/CoQ_{10} trial.

➢Based on the disease progression statistic (see text box above) it was decided to exclude patients with TFC=13. Compared to the riluzole and CoQ10 trials, the start TFC distribution reads:

TFC_{start} simulation				
Mean	SD	Median	Max	Min
10.6	1.12	11	12	9
Riluzole trial n=180				
Mean	SD	Median	Max	Min
10.8	1.77	11	13	5
CoQ_{10} trial n=87				
Mean	SD	Median	Max	Max
10.3	1.7	-	13	7

Randomization:

➢ Standard block randomization, *i.e.* no stratification.

Treatment Effect:

➢ The placebo treated group's disease progression was simulated first by sampling the baseline patient distribution returning a $TFC_{start} \in \{9,10,11,12\}$ for each patient; second, the TFC_{end} was sampled at the end of the study for each subject (TFC_{start} *vs.* TFC_{end} distribution, Fig.1) resulting in:

$$\Delta TFC_{placebo} = TFC_{end} - TFC_{start} \quad \text{(DECLINE)}$$

➢ The treatment effect (TE) was introduced as an attenuation of decline:

$$\Delta TFC_{treated} = \Delta TFC_{placebo}\,(1\text{-}TE) \quad \text{(ATTENUATED DECLINE)}$$

Treatment Effect (TE):
TE = 25% attenuation → minimal effect
....
TE = 100% attenuation → no decline

Trial Assessment:

➢ The primary end-point of the trial can be assessed using various statistical tests. The following tests were compared:
- ✓ t-test (one-tailed, α=0.05, equal var.)
- ✓ t-test (two-tailed, α=0.05, equal var.)
- ✓ ANCOVA (α=0.05, covariate: TFC_{start})
- ✓ Mann-Whitney unit-test (α=0.05)

Power Calculation:

➢ The power was calculated in the standard manner as:

Power is the probability of getting the right result (given an effect has been imposed).

➢ To calculate the power, 5000 trial simulations were generated.

Results

Riluzole trial:

➢ The riluzole trial design was asymmetric with a 1-to-2 ratio between the placebo and treated groups, and included also TFC=13 patients. We found that a balanced design improved trial power with 1-2%, while excluding TFC=13 improved power 1% or more, depending upon how investigators interpreted the "manifest HD" inclusion criterion in TFC terms.

Primary end-point:

➢ The primary end-point result can be assessed by one of many statistical tests. The positive trial outcome should only cover positive treatment effects: $\Delta TFC_{treated} < \Delta TFC_{placebo}$ but the two-tailed t-test covers both positive and negative outcomes if they are statistically significant. The power of the one-tailed t-test is higher than the two-tailed test (although the FDA recommends a two-tailed t-test [5]).

➢We found that the ANCOVA test provided the highest power and the Mann-Whitney U-test the lowest power (excluding the one-tailed t-test), with an 18-23 percentage point difference.

Sample size:

➢ The sample size and corresponding power was:

Trial outcome: ANCOVA (α=0.05, covariate:TFC_{start})

Sample size with/without dropout	Treatment Effect attenuation of decline	#patients placebo/treated	Power	95% CI
333/300	25%	100/200	56.62	[55.24 58.00]
〃	30%	〃	60.26	[58.90 61.62]
333/300	25%	150/150	58.32	[56.96 59.68]
〃	30%	〃	61.42	[60.06 62.76]
444/400	25%	200/200	69.88	[68.60 71.14]
〃	30%	〃	75.12	[73.92 76.32]
〃	40%	〃	92.08	[91.32 92.82]

Missing data:

➢Patient drop-outs of 10% were assumed. Only observed data was used, *i.e.* no imputations or LOCF.

References: [1] Shoulson and Fahn; Neurology. 1979 Jan;29(1):1-3. **[2]** Landwehrmeyer *et al.*; AnnNeur-2007,62,p262 **[3]** HD-study group, Neurology 2001, 57, p397 **[4]** Marder *et al.* Neurology-2000,54,p452 **[5]** FDA Guideline: ucm129456