Multi-PART: Work package 6

Philip Bath – Nottingham

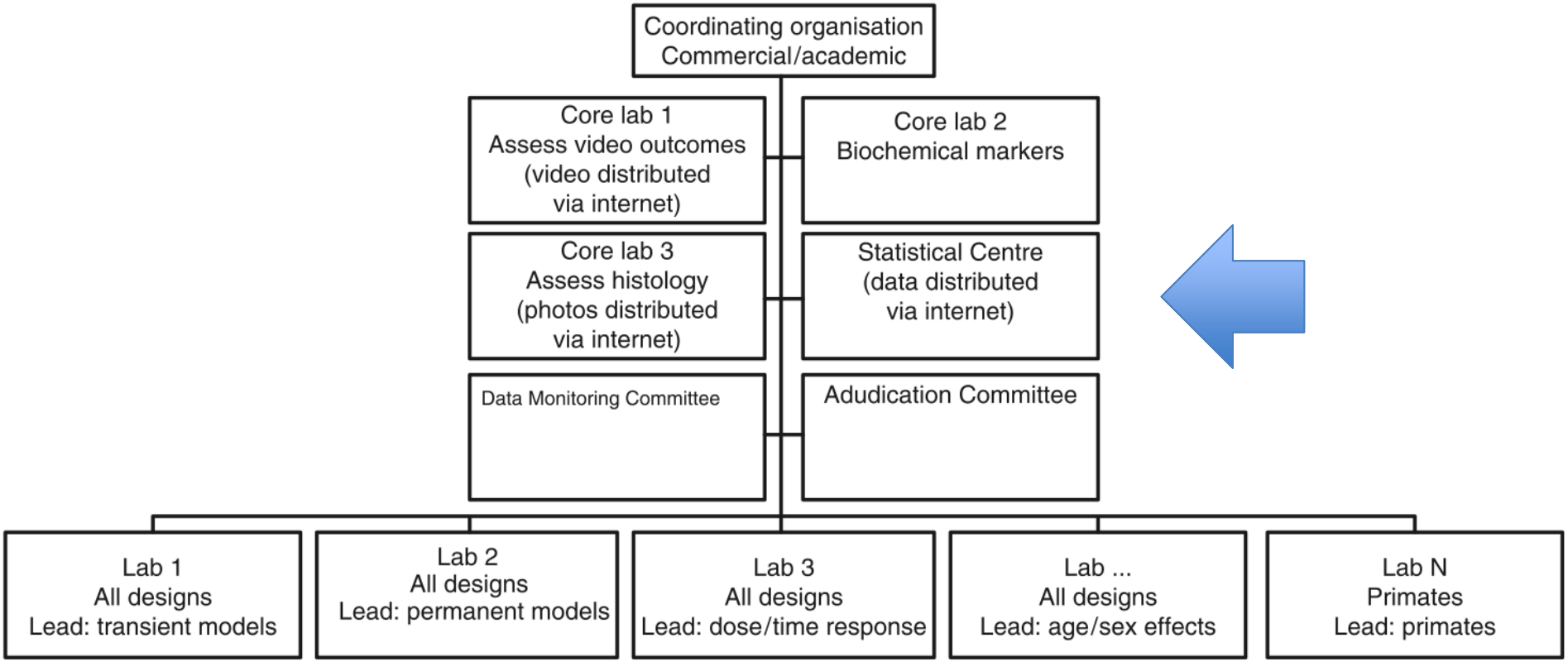
Joan Montaner - Barcelona

The Grant says we (WP6) will …

Develop statistical approaches for:

- Sample size calculations for various outcomes
- Blocked randomisation
- Interim analyses for efficacy or futility (including hazard)
- Primary & secondary analyses
- Adjustment of outcomes for baseline covariates
- Central statistical monitoring of protocol delivery

Multi-centre preclinical study organisation



Bath *et al*. Int J Stroke 2009;4:471-9

Sample size calculation, 1

Table 1 Comparison of preclinical experimental stroke studies and clinical stroke trials

	Animal/preclinical studies	Clinical (phase IIb/III) trials
Centres	Single or few centres	Multicentre
Staff	Industry and academic	Mostly clinical, some academic
Dose response studies	Variable	Common
Time response studies	Variable – most studies assess early administration	Uncommon
Sample size calculation*	Uncommon	Common (37)
Randomisation*	Variable	Almost all
Allocation concealment*	Variable	Common
Placebo control*	Variable	Common
Blinded surgery*	Variable	Not relevant
Blinded outcome assessment*	Uncommon (outcome often measured by surgeon)	Common
Functional outcome assessment	Uncommon	Common
Death (postrandomisation)	Animals discarded so death not counted	Counted within functional outcome scores (mRS = 6) and as a serious adverse event
Plasma drug analysis	Uncommon	Common
Publication bias	Common (18)	Variable
Publication quality	Poor	Moderate (23)
Study registration	Nil	Common (e.g. through controlled trials ISRCTN)
Data sharing	Limited [e.g. (20)]	Moderately common [e.g. VISTA, OAST (48, 49)
Systematic reviews based on summary (group) data	Uncommon (except by the applicants)	Very common (see Cochrane Library)
Systematic reviews based on individual subject data	Only 1 to date (20)	Moderately common (e.g. (50, 51))

*These procedures may be done but not reported. mRS, modified Rankin scale.

Bath *et al*. Int J Stroke 2009;4:471-9

Sample size calculation, 2

Statistics:

- n depends on measure (p, μ, …), variation (σ), effect size (p1-p2, μ2-μ1, OR, …), power (zβ), significance (zα)
- Binary

$$n = \frac{(z_\alpha + z_\beta)^2 (p_1(1-p_1) + p_2(1-p_2))}{(p_1 - p_2)^2}$$

- Ordinal

$$n = \frac{6\left[(z_\alpha + z_\beta)^2 / (Log\mathrm{OR})^2\right]}{\left[1 - \sum_{i=1}^{k} \overline{\pi}^3\right]}$$

- Continuous

$$n = \frac{2\sigma^2 (z_\alpha + z_\beta)^2}{(\mu_2 - \mu_1)^2}$$

OAST Collaboration. Int J Stroke 2008;3:78-84

Sample size calculation, 3

Parameters:

- Measure: from previous successful developments
- Variation: from previous successful developments
- Effect size: worthwhile, achievable (gaming)
- Power: 80%, 85%, 90% (higher means more n)
- Significance: 5%

Blocked randomisation

- Need to randomise with concealment of allocation
  - Reduce bias, choice, predictability
- Need to ‘block’ site, i.e. remove effect of site (and its staff, techniques, animal source, …)
- All sites need to do all (most) experimental parameters (model, dose, time, outcomes, …)
- Block-treatment interaction can (potentially) be assessed
- Approach needs to be programmed into randomisation website

Outcomes

Data types:

- Binary data: death, …
- Time to event data: time to death, …
- Ordinal data: neuro score, …
- Continuous data: lesion volume, foot faults, weight, time to, …

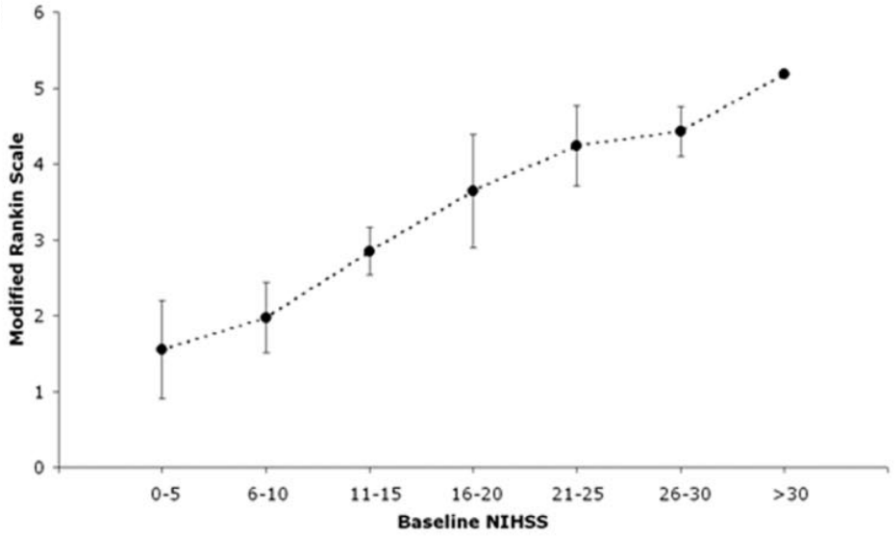
- All other things being equal, more statistical power: continuous > ordinal > binary

Primary and secondary analyses

- Only have one primary outcome/analysis
  - ‘put your money where your mouth is’
- Avoids multiplicity
  - Multiple testing may find something to be positive just by chance
- Varies in clinical stroke trials
  - Phase I: Concentration, biomarkers, …
  - Phase II: Safety, imaging
  - Phase III: functional (modified Rankin Scale)
- So we need to choose a primary outcome
  - Will vary by where we are in development cycle

Adjustment for covariates

- Deal with baseline imbalances
- Increase statistical power or reduce sample size (by ~25%)
- Which variables?
  - Imbalanced
  - Related to 1ry outcome



- Example model
  - mRS ~ treatment, age, sex, severity, … time, weight, model, comorbidity, …

OAST Collaboration. Stroke 2009;40:888-894

Interim analysis, 1

Stop study early if:

- Strongly positive ($p<0.001$)
  - > get on with next stage or development (or test in clinic, or introduce treatment to routine care)
- Neutral and probability of success low
  - > study something else / stop wasting time
- Negative ($p<0.01$)
  - > dangerous to participants, study something else

Otherwise carry on with protocol

Interim analysis, 2

But:

- If positive: ‘spends’ statistical significance, if overall $p<0.05$
  - Assessments: 1, $p<0.025$; 2, $p<0.017$; 3, $p<0.0125$
  - Avoid by only stopping if $p<0.001$
  - Only relevant if testing for benefit
- If neutral: might miss a benefit
  - So, make sure probability of success low
- Rules in animal experiments probably need to be less sensitive than those for clinical trials

Central statistical monitoring

- Monitoring: Site visits, central monitoring
- Database analysis by site (laboratory)
  - Errors/fraud: Extreme data/outliers, lack of variability, digit preference, …
  - Performance: speed of data entry, accuracy (proportion of incorrect data)
  - Delivered by regular database analyses with programming early on in project

Other statistical issues

- Death of animals
  - Death must be included
  - Need to exclude ‘kill or cure’
  - Progesterone SMD [1]:
    - Death -0.322 (p=0.16), no death -0.585 (p=0.004)
- Repeated measures (or gradient)
  - Relevant to functional/impairment measures

1. Wong *et al*. JCBFM 2013;in press