# **Multi**centre **P**reclinical **A**nimal **R**esearch **T**eam

Multi-PART

**Work Package 3 – Experimental Design**

Hanno Würbel, Denis Vivien

Presented by Philip Bath

# WP 3

## Aims and deliverables – Experimental design

- Describe the most appropriate design for future preclinical multicentre studies
  - Randomisation, blinding, sample size calculations, systematic variation
- Develop strategies to maximise both the internal and external validity of the studies performed within Multi-PART or other multi-centre research consortia

# WP 3

**Design of Multi-PART studies**

- Surgery — surgery, sham surgery
- Treatment — drug, vehicle
- Time — e.g. early, late
- Dose — e.g. high, low
- Stroke model — e.g. intraluminal filament, distal MCAO, …
- Co-morbidities — e.g. hypertension, …
- Animal Species — e.g. rat, mouse, macaque, …
- Centre — e.g. Macrae, Dirnagl, Planas, …

# WP 3

## Experimental design – Principles

"[Experiments] should be unbiased, be powerful, have a good range of applicability, not be excessively complex, and be statistically analysable to show the range of uncertainty in the conclusions."

(Festing, 2006)

# WP 3

## External validity – Avoiding spurious results

- Small scale, standardised, single-centre studies are prone to produce spurious results
- Multiple small scale studies may not help – sufficient power is important
- Power alone is not enough – heterogeneity is crucial

→ Multi-centre studies

# WP 3

## Design of multicenter trials

- Use mixed model stats, treat lab as random factor
- Sample of labs should represent population of labs
- Number of labs should allow estimation of between-lab variation
- Number of subjects per lab should allow estimation of within-lab variation

# WP 3

## Design of multicenter trials – Sample sizes

| Lab | n/Lab | External | Internal validity | |
|---|---|---|---|---|
| 1 | 100 | **Poor** | **Excellent** | |
| | | no estimate of between-lab variation | | |
| 10 | 10 | **Reasonable** | **Reasonable** | i.e. compromise |
| | | gives estimates of within & between-lab variation | | |
| 100 | 1 | **Excellent** | **Poor** | |
| | | no estimate of within-lab variation | | |

small sample sizes associated with low quality

# WP 3

## Internal validity – Avoiding bias

- Relevant sources of bias have long been identified
  (randomisation, blinding, sample size calculation, primary outcome, etc.)
- Methods for randomisation, blinding and sample size calculation are well established (e.g. CONSORT)
- Guidelines for preclinical (animal) studies have been developed (e.g. ARRIVE), but need to be implemented

# WP 3

## Internal validity – Randomisation

- Randomisation protocols for controlled multi-centre studies are readily available (e.g. sealed envelope)

**Recommendations:**

- Centralised treatment allocation and randomisation
- Distance randomisation (via www, phone, sms, etc.) instead of envelopes
- Stratification as needed, e.g. by blocked randomisation with random block sizes (Efird, 2011)

# WP 3

## Internal validity – Blind as much as possible

Level 0: Allocation concealment
→ *Centralised treatment allocation*

Level 1: Blinding persons interacting with animals

Level 2: Blinding persons collecting data
→ *Random number lists (animals, cages, groups, etc.)*

Level 3: Blinding persons assessing outcome
→ *Centralised, blinded outcome assessment*

Level 4: Blinding persons analysing data
→ *Centralised, blinded data analysis*

# WP 3

## Statistical power – Sample size calculation

- Software for simple sample size calculations with adjustments to experimental design are readily available
  e.g. PASS (NCSS), G*Power
- In the simplest form, the statistical design includes treatment as a fixed factor and centre as a random factor
- If between-centre variation is unknown, sample size calculation may be based on intraclass corr. coeff. (ICC) = 0.1
  ICC reflects proportion of variance explained by centre

# WP 3

## Future perspectives

- Use existing data to analyse different experimental designs
- Model trade-offs between variation of different factors
- Model trade-offs between number of labs and subjects/lab
- Model trade-offs between sample size and external validity

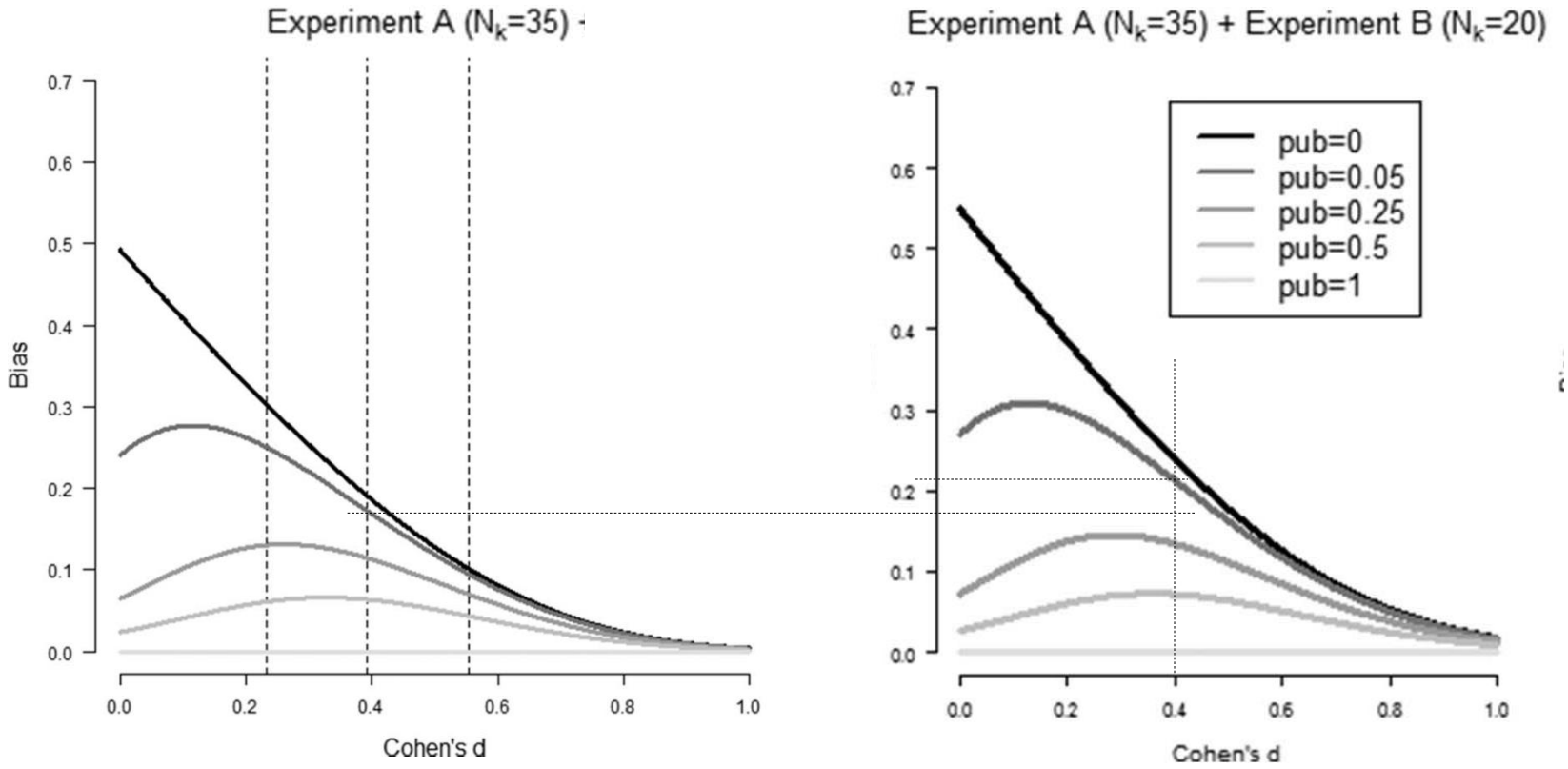
→ WP6

# **Multi**centre **P**reclinical **A**nimal **R**esearch **T**eam

Thanks

# WP 3

## Power vs. Replication – the Replication Paradox




Nuijten et al. 2015

# WP 3

## External validity – Avoiding spurious results

- Small scale, standardised, single-centre studies are prone to produce spurious results
- Multiple small scale studies do not help – sufficient power is important
- Power alone is not enough – heterogeneity is crucial
  → Multi-centre studies (or heterogenized single-centre studies)

# WP 3

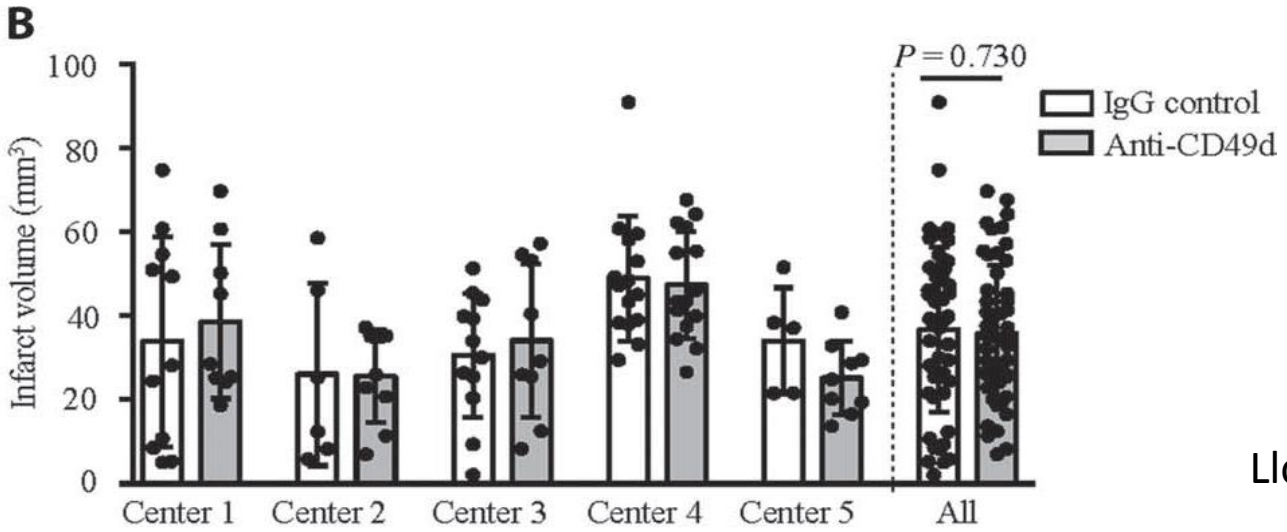
## ARRIVE guidelines – Endorsed but not implemented

- Survey (Reichlin et al., unpublished)
- Of 250 scientists, 79 had published their last paper in a journal that had endorsed ARRIVE
- Of these 79 scientists, 13 knew ARRIVE well; 9 had read them, 17 had heard of them
- **The remaining 40 (>50%) scientists had never even heard of the ARRIVE guidelines….**

# WP 3

## Design of multicenter trials

- Define criteria for animal models and outcomes (SOPs)
  (face, construct, and predictive; convergent and discriminant validity)
- Define standards for accreditation of centers




Llovera et al. 2015

# WP 3

## Design of multicenter trials – Balancing groups

**Table 3: Design effects calculations for three different group distributions among centers.**

| Group distribution among centers | Quite homogeneous | | | Heterogeneous | | | Cluster design | | |
|---|---|---|---|---|---|---|---|---|---|
| **Group size per center** | $m_{1j}$ | $m_{2j}$ | %* | $m_{1j}$ | $m_{2j}$ | %* | $m_{1j}$ | $m_{2j}$ | %* |
| Center 1 (n = 57) | 16 | 41 | **28** | 11 | 46 | **19** | 0 | 57 | **0** |
| Center 2 (n = 38) | 10 | 28 | **26** | 24 | 14 | **63** | 38 | 0 | **100** |
| Center 3 (n = 44) | 11 | 33 | **25** | 7 | 37 | **16** | 0 | 44 | **0** |
| Center 4 (n = 15) | 3 | 12 | **20** | 1 | 14 | **7** | 0 | 15 | **0** |
| Center 5 (n = 41) | 9 | 32 | **22** | 8 | 33 | **20** | 0 | 41 | **0** |
| Center 6 (n = 19) | 5 | 14 | **26** | 10 | 9 | **53** | 19 | 0 | **100** |
| Center 7 (n = 37) | 8 | 29 | **22** | 9 | 28 | **24** | 0 | 37 | **0** |
| Center 8 (n = 52) | 12 | 40 | **23** | 4 | 48 | **8** | 0 | 52 | **0** |
| Center 9 (n = 12) | 3 | 9 | **25** | 1 | 11 | **8** | 0 | 12 | **0** |
| Center 10 (n = 28) | 8 | 20 | **29** | 10 | 18 | **36** | 28 | 0 | **100** |
| S | | 0.14 | | | 5.79 | | | 33.77 | |
| ***Deff* ($\rho$ = 0.10)** | | **0.91** | | | **1.48** | | | **4.28** | |

*group 1 proportion in each center
The global proportion of subjects in group 1 is 25%, for each group distribution, and the Intraclass Correlation Coefficient is equal to 0.10.

# WP 3

## ARRIVE guidelines – Endorsed but not implemented

- Survey (Reichlin et al., unpublished)
- Of 250 scientists, 79 had published their last paper in a journal that had endorsed ARRIVE
- Of these 79 scientists, 13 knew ARRIVE well; 9 had read them, 17 had heard of them