Improving the Analysis and Interpretation of Vascular Prevention Trials

Philip M W Bath

Stroke, Division of Clinical Neuroscience

University of Nottingham

Work with Stuart Pocock, Chamila Geeganage, Lisa Woodhouse

Presentation

Prevention trials

Hang methods on Acute trials and Optimising Analysis of Stroke Trials Collaboration

Ordinalising prevention trials – experimental concept

- Examples
- Issues

Stroke trials

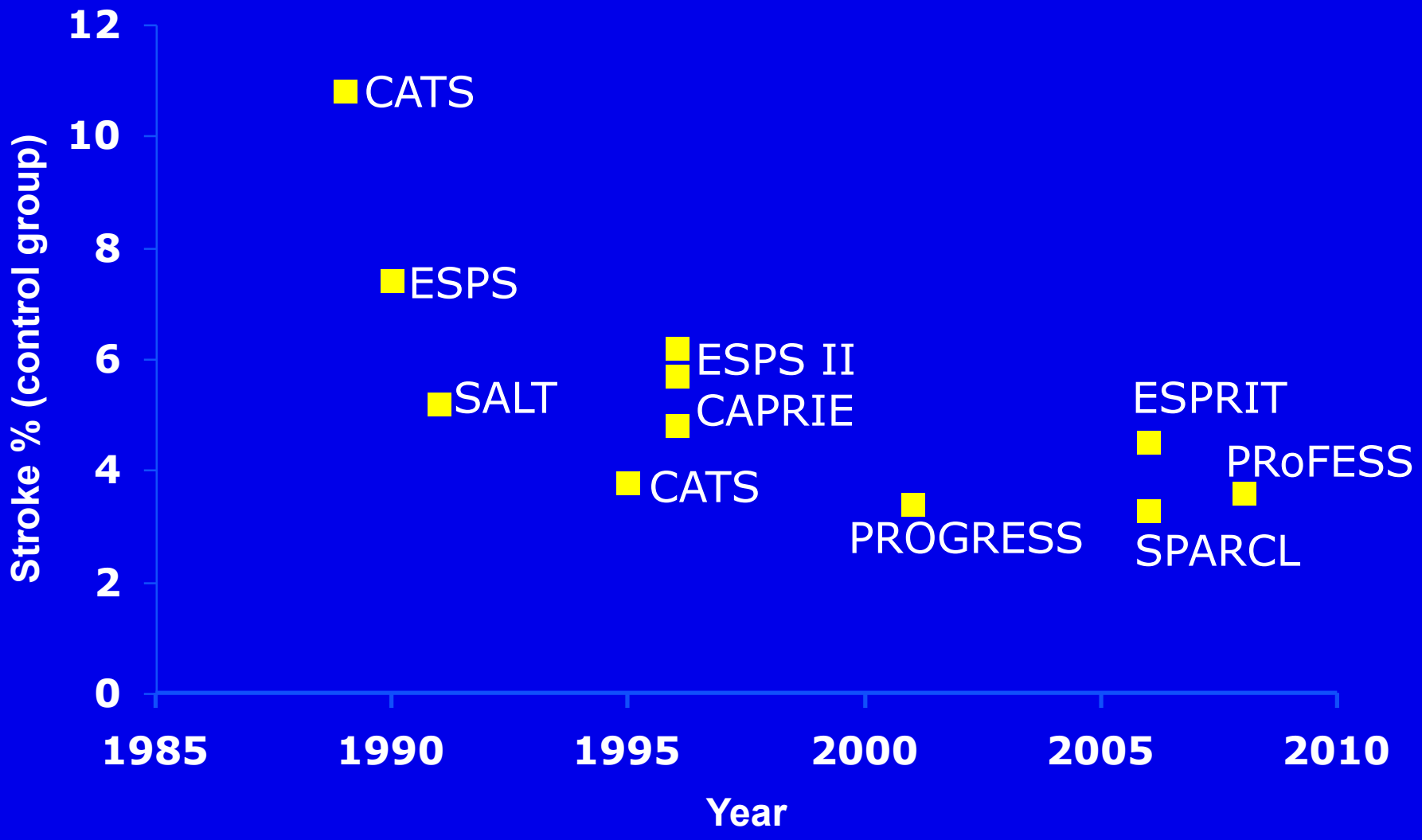
Acute trials:

- Few effective interventions: alteplase, aspirin, hemicraniectomy
- Suboptimal design: interventions inadequately worked up, trials too small, wrong outcomes, poorly analysed, …
- 7-level modified Rankin Scale (mRS) analysed as 2 levels
  - Several examples where neutral becomes significant

Prevention trials:

- Many effective interventions: antiplatelets, anticoagulation, BP lowering, lipid lowering, carotid endarterectomy, …
- Remaining reduction limited: mega-trials now the norm but expensive in £$€ and time, site availability, site quality
- Vascular events/stroke recurrence analysed as binary events with no account of severity

Stroke recurrence rates by time

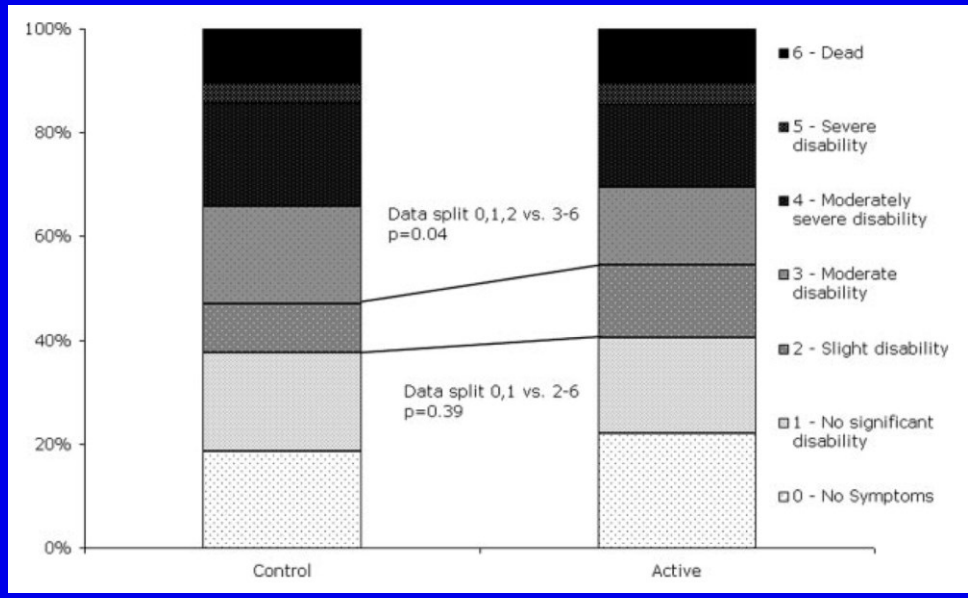


Optimising Analysis of Stroke Trials

Drivers for OAST:

ECASS-2 trial

- Results vary by position of dichotomy [1]
  - Primary – neutral
  - Post hoc - positive
- Significant with ordinal analysis:
  - Bootstrap [2]
  - Mann-Whitney U [3]



1. Hacke *et al*. *Lancet* 1998;352:1245-51.
2. Stingele *et al*. *Cerebrovasc Dis* 2001;11:30-3.
3. OAST. *Stroke* 2007;38:1911-15.

OAST: Acute stroke trials – Analysis

Aim:

- To identify optimal ways of analysing acute trials

Methods:

- Empirical
- Trial individual patient data
- Analyse each trial using each approach
- Rank analysis results

55 datasets, 16 tests

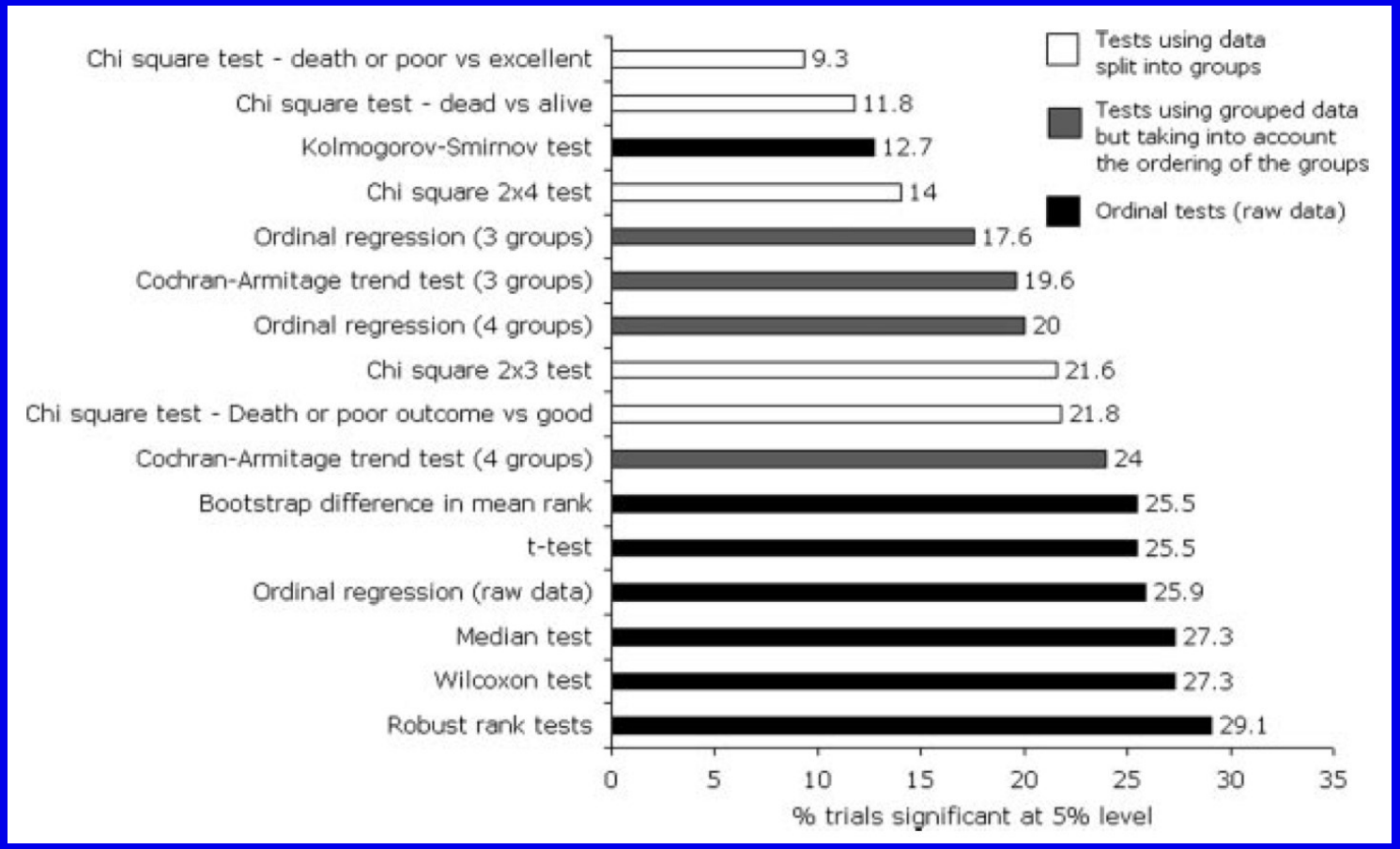
Test	Mean Rank	No. of Datasets	Banding
Ordinal logistic regression	6.11	54	
t test	6.51	55	
RRT	6.53	55	
Bootstrap difference in mean rank	6.85	55	
Wilcoxon test	7.31	55	
Cochran-Armitage trend test (4 groups)	7.36	50	
Ordinal logistic regression (4 groups)	7.50	50	
Ordinal logistic regression (3 groups)	7.92	51	
Cochran-Armitage trend test (3 groups)	8.27	51	
χ^2 – death or poor outcome vs good	8.87	55	
χ^2 – death or poor outcome vs excellent	9.24	54	
Median test	9.47	55	
χ^2 – 2x3 test	9.96	51	
χ^2 – death vs alive	9.98	51	
χ^2 – 2x4 test	10.02	50	
Kolmogorov-Smirnov test	11.29	55	

Comparison of rank scores for 16 statistical tests; lower ranks imply the test is more efficient. Analysis by two-way ANOVA and Duncan's multiple comparison procedure; tests joined by the same band are not significantly different from each other at $P<0.05$.

OAST. *Stroke* 2007;38:1911-15.

OAST: Acute stroke trials – Analysis

Number of trials that are ‘significant’ increases with optimal analysis



OAST. *Stroke* 2007;38:1911-15.

OAST: Acute stroke trials – Sample size

Sample size estimations:

- Binary outcomes

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

- Ordinal outcomes (Whitehead)

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

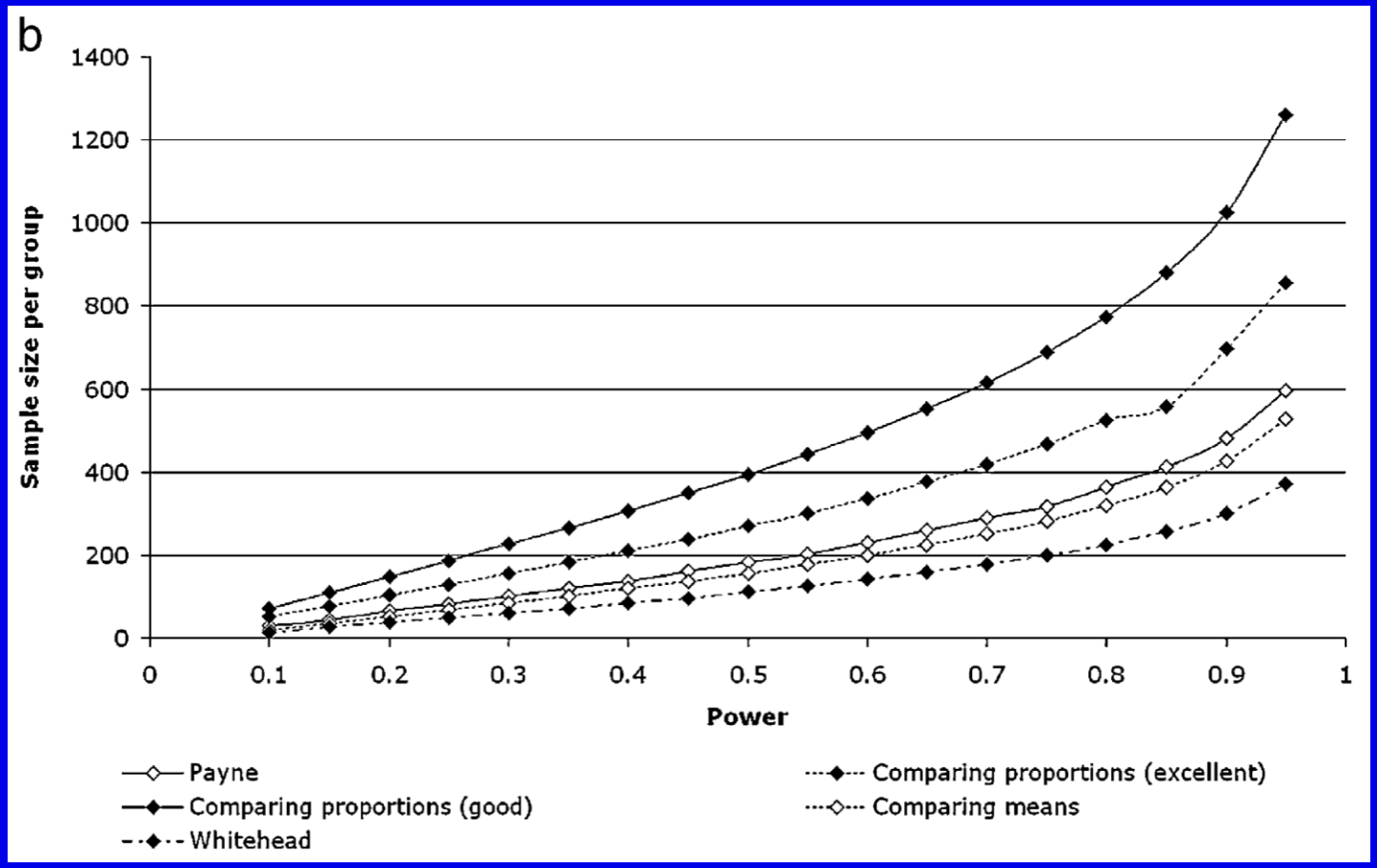
- Continuous outcomes

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

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

OAST: Acute stroke trials – Sample size

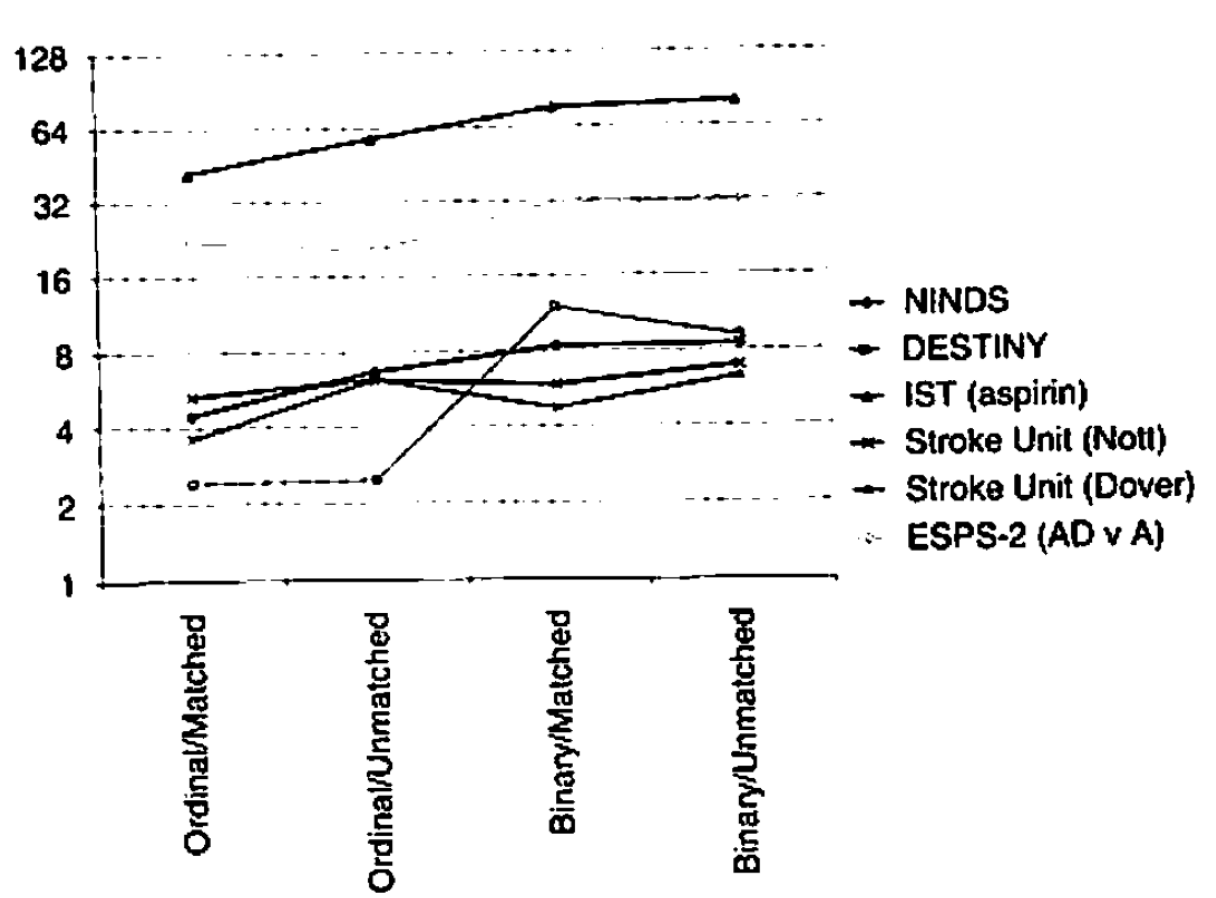
Comparison of sample sizes estimated using binary, ordinal and continuous approaches



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

OAST: Acute stroke trials - NNT

Ordinal Numbers Needed to Treat (NNT) can be calculated and are smaller, i.e. better, than binary



Bath *et al*. *Int J Stroke* 2011;6:472-9.

OAST: Acute stroke trials - Covariates

Sample size can be reduced by ~20% (or power increased) in large trials if analyses are adjusted for baseline covariates, including any used in minimisation.

Typical covariates:

Age, sex, severity, premorbid mRS, time to randomisation, …

Should use covariate analysis of minimisation

OAST. *Stroke* 2009;40:888-94.

Results consistency for non-binary analyses

Projects:

- OAST stroke
- Lees *et al* stroke
- Saver *et al* stroke
- Murray *et al* TBI

Methods:

- Statistical lore: continuous ~ ordinal > binary
- Empirical
- Modeling with trial data/artificial treatment effects
- Modeling with artificial data and treatment effects

European Stroke Organisation

Contemporary Outcome Measures in Acute Stroke Research

Choice of Primary Outcome Measure

Kennedy R. Lees, MD, FESO; Philip M.W. Bath, MD, FESO; Peter D. Schellinger, MD, FESO; Daniel M. Kerr, BSc; Rachael Fulton, MSc; Werner Hacke, MD, FESO; David Matchar, MD; Ruchir Sehra, MD; Danilo Toni, MD, FESO; for the European Stroke Organization Outcomes Working Group

Background and Purpose—The diversity of available outcome measures for acute stroke trials is challenging and implies that the scales may be imperfect. To assist researchers planning trials and to aid interpretation, this article reviews and makes recommendations on the available choices of scales. The aim is to identify an approach that will be universally accepted and that should be included in most acute trials, without seeking to restrict options for special circumstances.

Methods—The article considers outcome measures that have been widely used or are currently advised. It examines desirable properties for outcome measures such as validity, relevance, responsiveness, statistical properties, availability of training, cultural and language issues, resistance to comorbidity, as well as potential weaknesses. Tracking and agreement among outcomes are covered.

Results—Typical ranges of scores for the common scales are described, along with their statistical properties, which in turn influence optimal analytic techniques. The timing of recovery on scores and usual practice in trial design are considered.

Conclusions—The preferred outcome measure for acute trials is the modified Rankin Scale, assessed at 3 months after stroke onset or later. The interview should be conducted by a certified rater and should involve both the patient and any relevant caregiver. Incremental benefits at any level of the modified Rankin Scale may be acceptable. The modified Rankin Scale is imperfect but should be retained in its present form for comparability with existing treatment comparisons. No second measure should be required, but correlations with supporting scales may be used to confirm consistency in direction of effects on other measures. **(*Stroke*. 2012;43:1163-1170.)**

Assessment of additional endpoints for trials in acute stroke – what, when, where, in whom

Peter D. Schellinger[1]*, Philip M. W. Bath[2], Kennedy R. Lees[3], Natan M. Bornstein[4], Eitan Uriel[4], Wolfgang Eisert[5], and Didier Leys[6] for the European Stroke Organisation Outcomes Working Group†

Statistical Analysis of the Primary Outcome in Acute Stroke Trials

Philip M.W. Bath, FRCP, FESO; Kennedy R. Lees, FRCP, FESO; Peter D. Schellinger, MD, FESO; Hernan Altman, BSc, MBA; Martin Bland, PhD; Cheryl Hogg, MSc; George Howard, PhD; Jeffrey L. Saver, MD, FAHA; on behalf of the European Stroke Organisation Outcomes Working Group†

Abstract—Common outcome scales in acute stroke trials are ordered categorical or pseudocontinuous in structure but most have been analyzed as binary measures. The use of fixed dichotomous analysis of ordered categorical outcomes after stroke (such as the modified Rankin Scale) is rarely the most statistically efficient approach and usually requires a larger sample size to demonstrate efficacy than other approaches. Preferred statistical approaches include sliding dichotomous, ordinal, or continuous analyses. Because there is no best approach that will work for all acute stroke trials, it is vital that studies are designed with a full understanding of the type of patients to be enrolled (in particular their case mix, which will be critically dependent on their age and severity), the potential mechanism by which the intervention works (ie, will it tend to move all patients somewhat, or some patients a lot, and is a common hazard present), a realistic assessment of the likely effect size, and therefore the necessary sample size, and an understanding of what the intervention will cost if implemented in clinical practice. If these approaches are followed, then the risk of missing useful treatment effects for acute stroke will diminish. **(*Stroke*. 2012;43:1171-1178.)**

Can we optimise prevention trials?

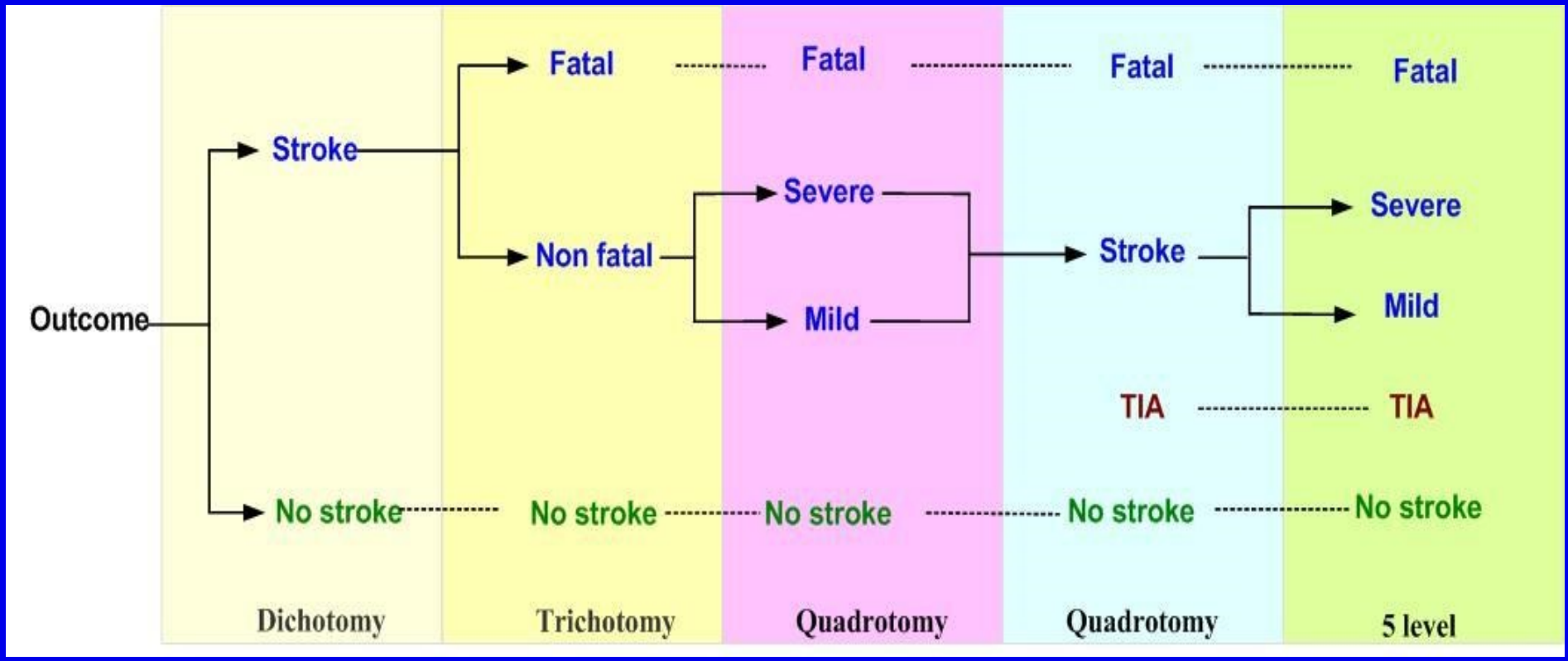
As compared with dichotomous or time-to-event analyses, ordering vascular outcomes or stroke recurrence might allow:

- Superior power / smaller sample size
- Demonstration that an intervention reduces both events AND their severity

Prevention trials

Ordering of vascular outcome/stroke events:

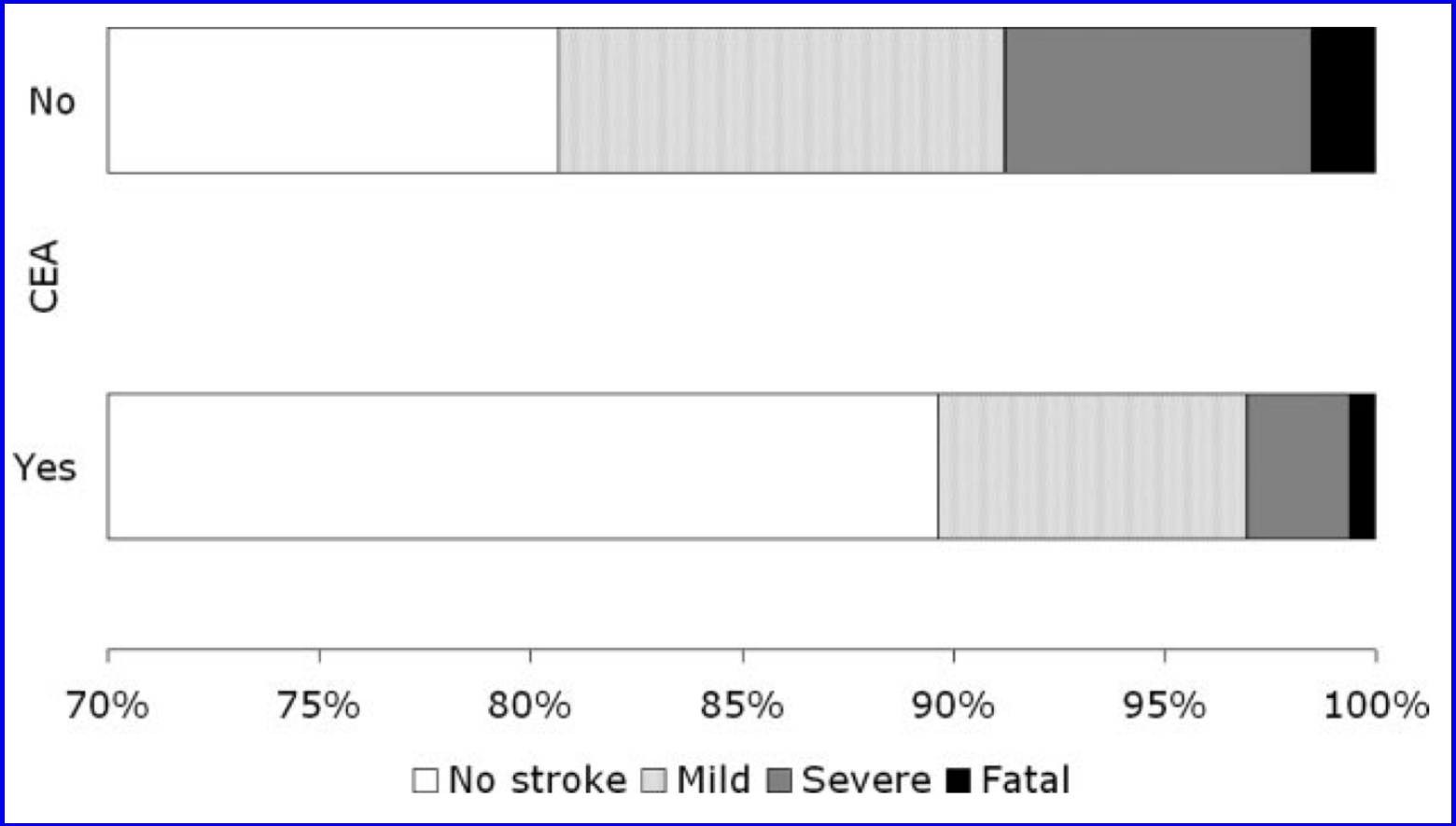
- Fatal, non fatal, no event = 3 levels
- Fatal, severe non fatal, mild, no event = 4 levels
- Fatal, severe non fatal, mild, TIA, no event = 5 levels



OA-prevention: Summary data

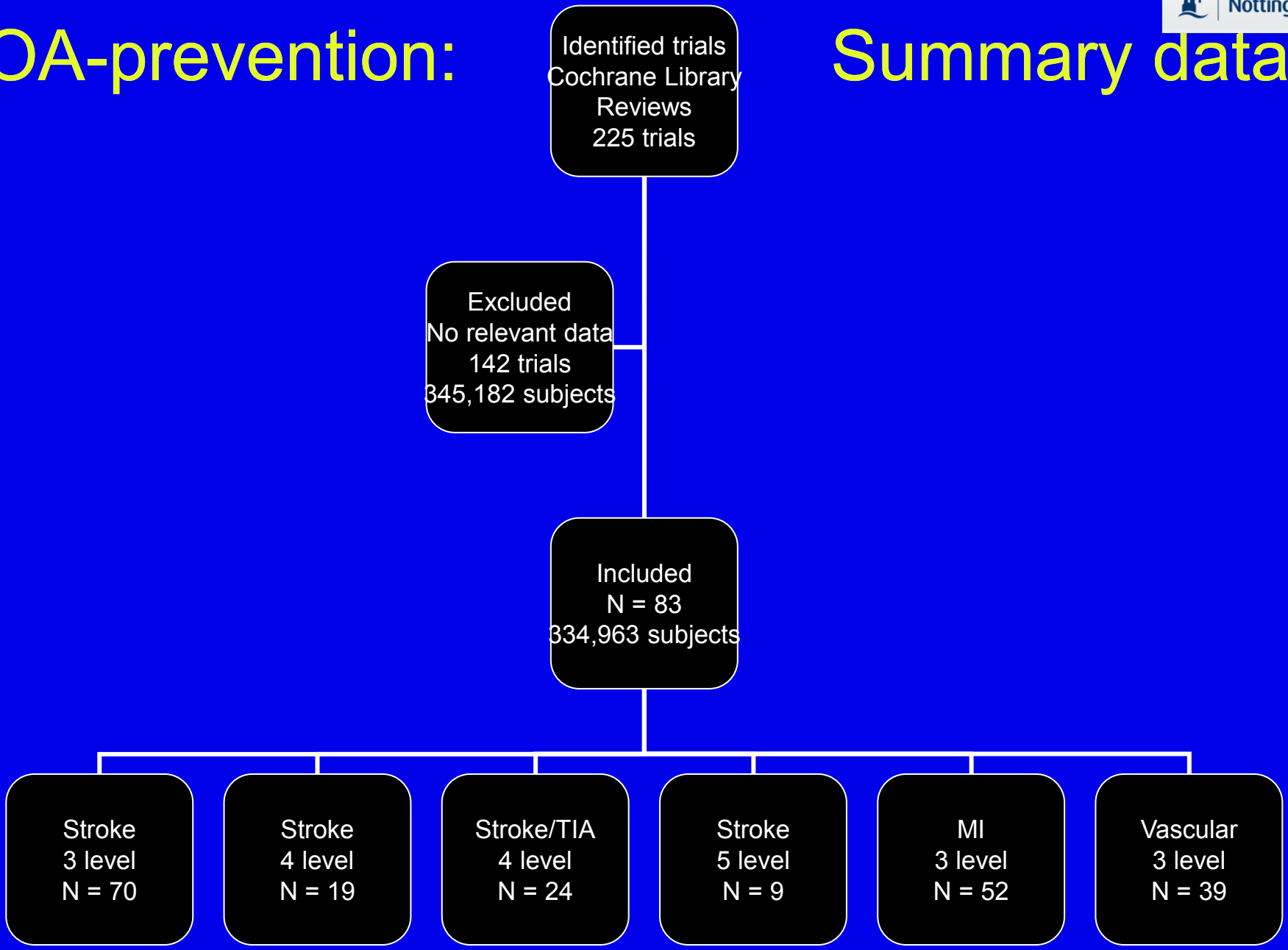
Ordinalised 4-level stroke data from NASCET

p: 2-level 0.002, 3-level 0.001, 4-level 0.0009



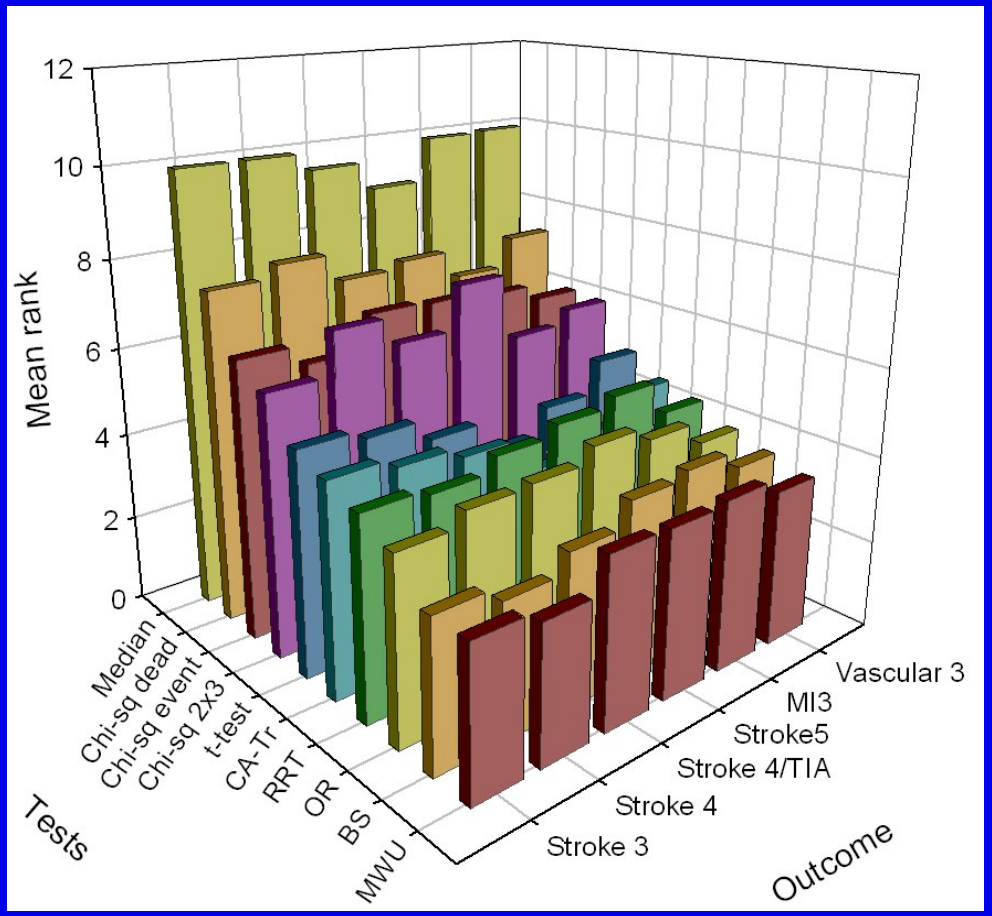
Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data



Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data



- The statistical tests differed substantially for each outcome in their efficiency (ANOVA p <0.0001)
- Ordinal analyses ranked above dichotomous approaches
- Similar results for stroke, MI, vascular events

Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data

Aim:

- To assess whether prevention outcomes can be ordinalised

Methods:

- Empirical
- Trial published summary data
- Analyse each trial using each approach
- Rank analysis results

85 trials, 10 tests

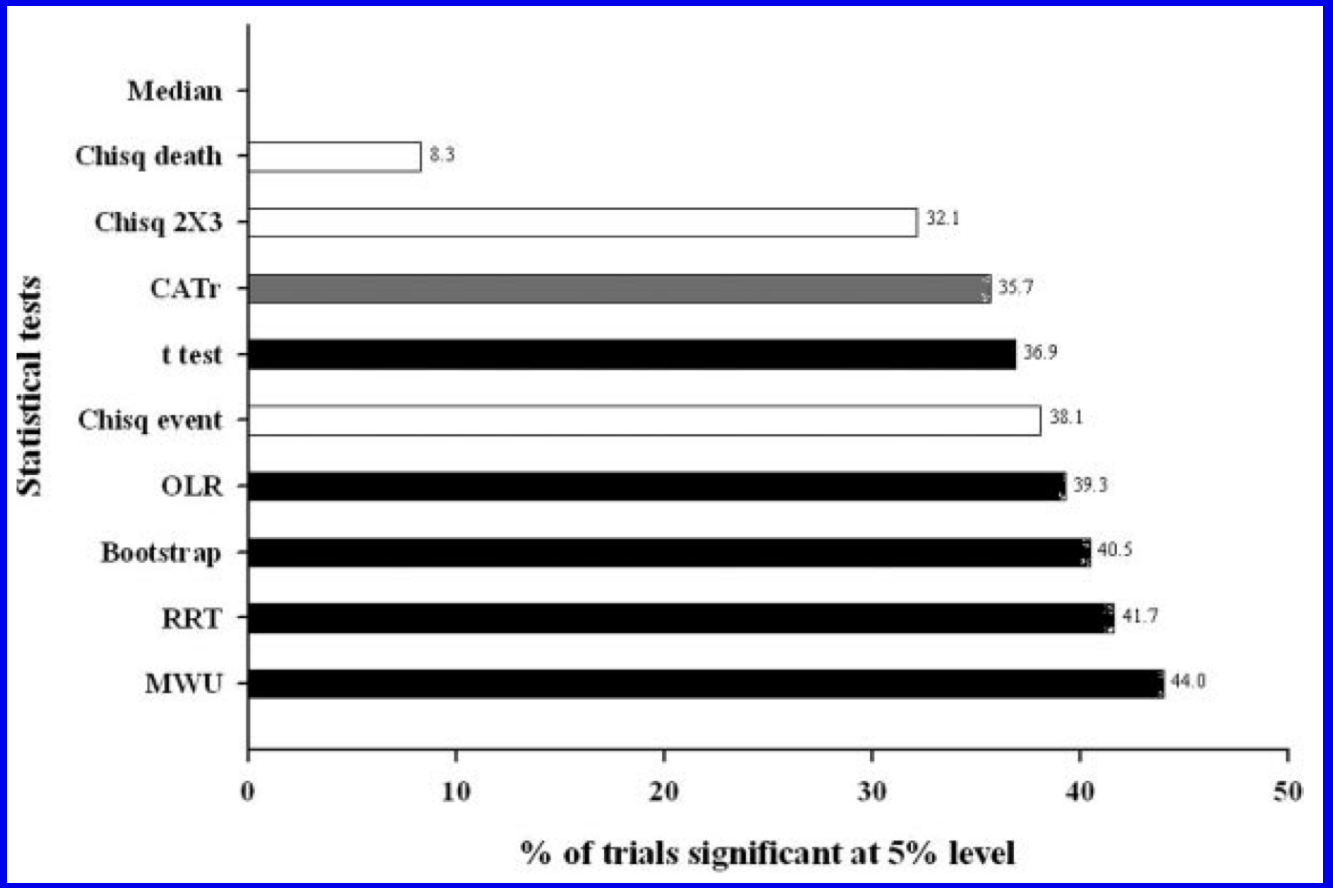
Test	Mean rank	Banding
Mann-Whitney U test	3.32	
Bootstrap (difference in mean rank)	3.32	
Ordinal logistic regression	4.12	
Robust ranks test	4.51	
Cochran-Armitage trend test	4.80	
t-test	5.08	
Pearson's Chi Sq – 2x3 test	5.94	
Pearson's Chi Sq – stroke vs. no stroke	6.37	
Pearson's Chi Sq – death vs. alive	7.58	
Median test	9.97	

Analysis by 2-way ANOVA ($P<0.0001$) on the ranked data (1 to 10 with 1 "best"); comparison of tests by Duncan's multiple range test—those tests joined by the same band are not significantly different from each other at $P<0.01$.

Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data

Number of trials that are ‘significant’ increases with optimal analysis



Bath *et al*. *Stroke* 2008;39:2817-23.

Hazard: Bleeding

Same approach applies:

- 3 level bleeding: major / minor / none
- 16 trials
- Significant difference in tests, ANOVA $p<0.00001$
- Most efficient tests
  - Ordinal logistic regression, bootstrapping, MWU

Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data

Testing assumption of proportionality of odds

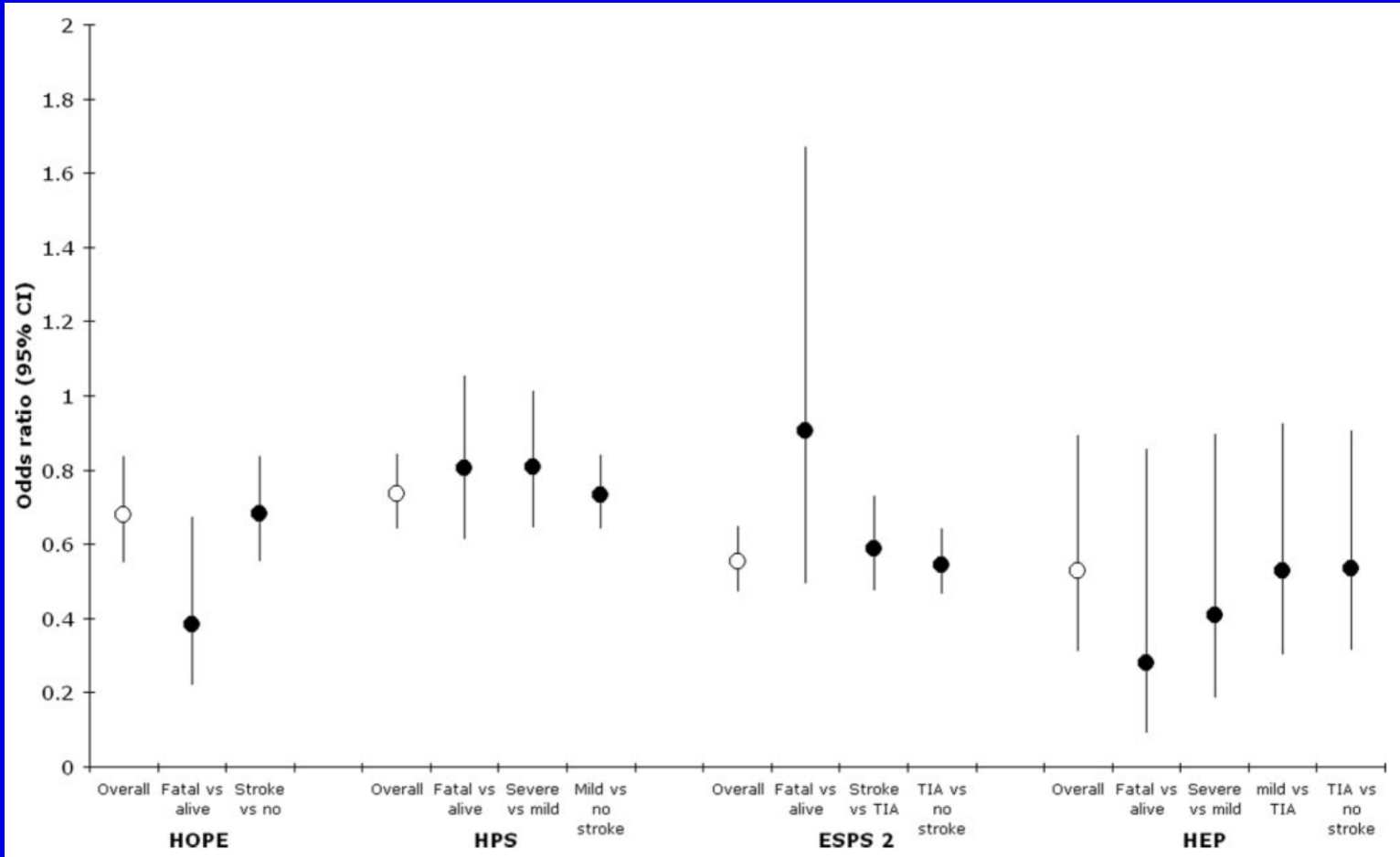
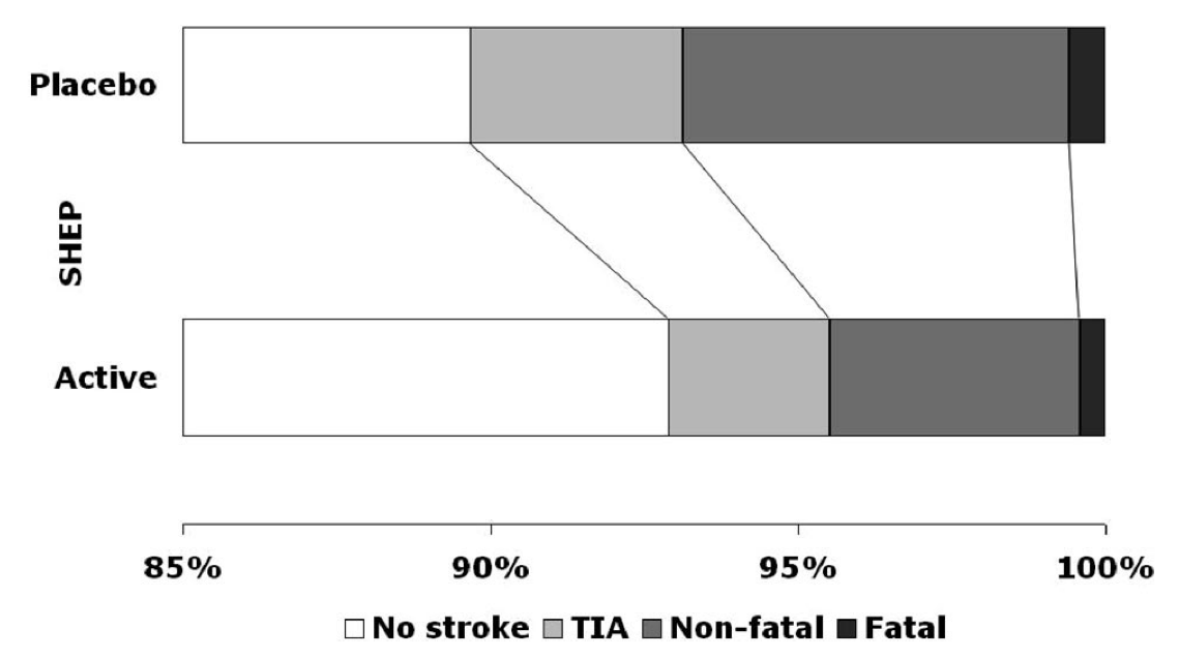


Figure 2. Odds ratios across trial (by ordinal logistic regression) and by individual outcome levels for 4 trials to illustrate the assumption of proportionality of odds.

Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: SHEP

Chlorthalidone



Stroke outcome

Levels	z	p
2	-3.53	0.0005
3	-3.53	
4 (TIA)	-3.94	0.00009

Bath *et al*. *Stroke* 2008;39:e145.

OA-prevention: Triple antiplatelets

Chronic aspirin, clopidogrel, dipyridamole vs aspirin

N=17

Adverse events

4-level: $p<0.01$

Bleeding

3-level: $p<0.01$

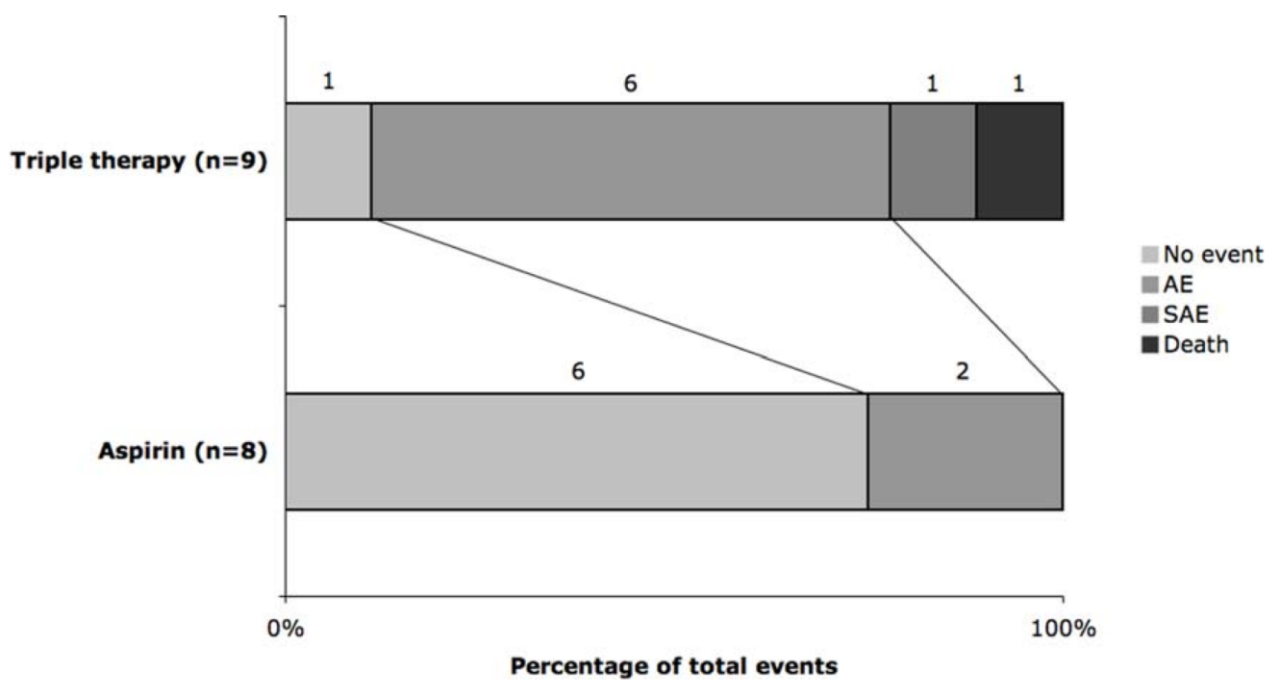
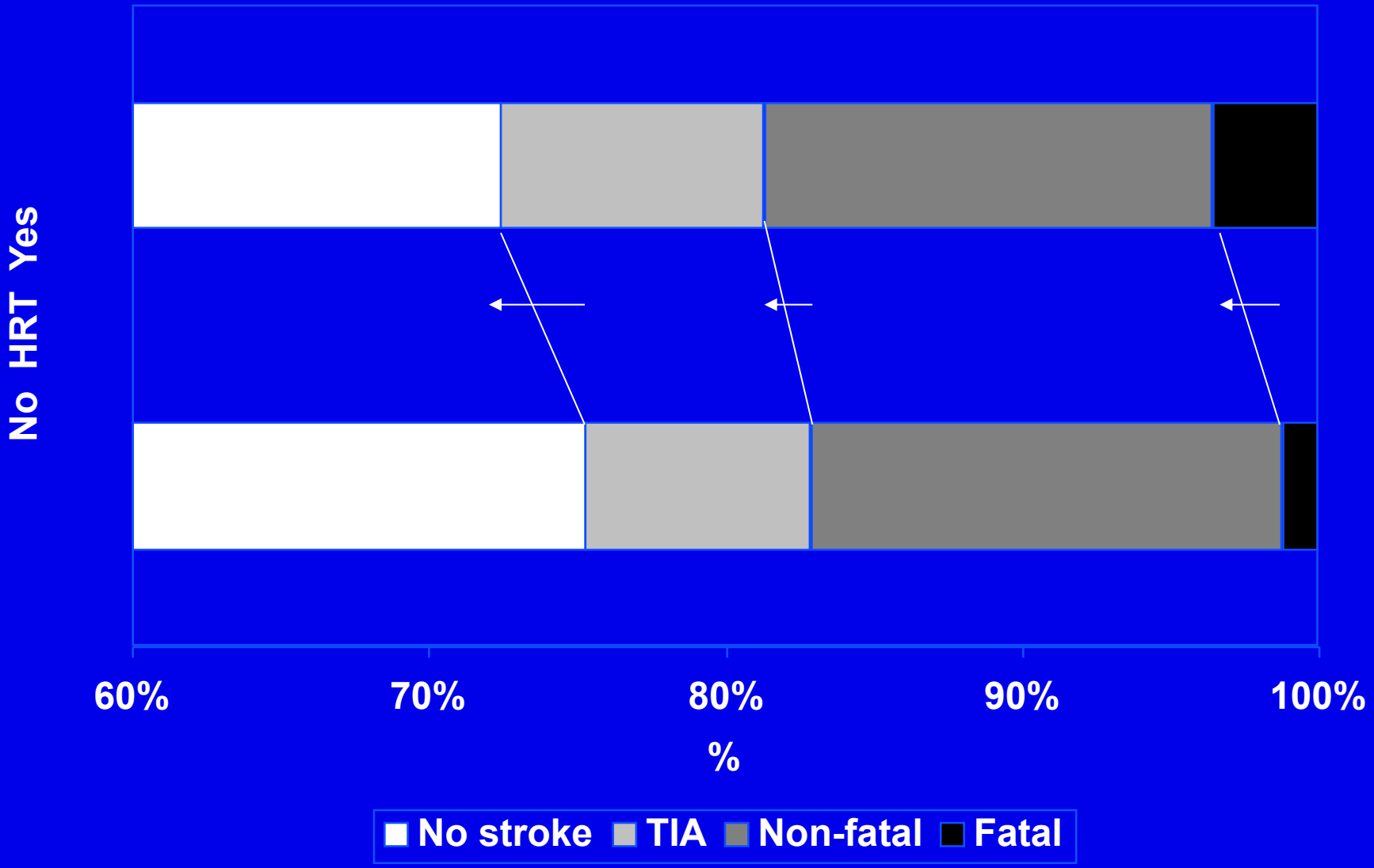


Figure 2. Frequencies of adverse events in aspirin and triple therapy groups.

Sprigg *et al*. *PLoS ONE* 2008;3:e2852.

OA-prevention – stroke/TIA: WEST

Stroke dichotomy p=0.61, Stroke/TIA ordinal p=0.37



OA-prevention: Summary data

Ordinal approach worked for different outcomes:

- Stroke: 3-, 4- (stroke), 4- (TIA), and 5-levels
- MI: 3-level – fatal / non-fatal / none
- MACE: 3-level - fatal / non-fatal / none
- Bleeding: 3-level – major / minor / none

Ordinal approach worked for different patient groups:

- Age; Recent / distant event
- Hypertension, HRT, post-stroke, post-MI
- Low / high risk of death; Low / high risk of stroke

Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Summary data, 2

Ordinal approach worked for different interventions:

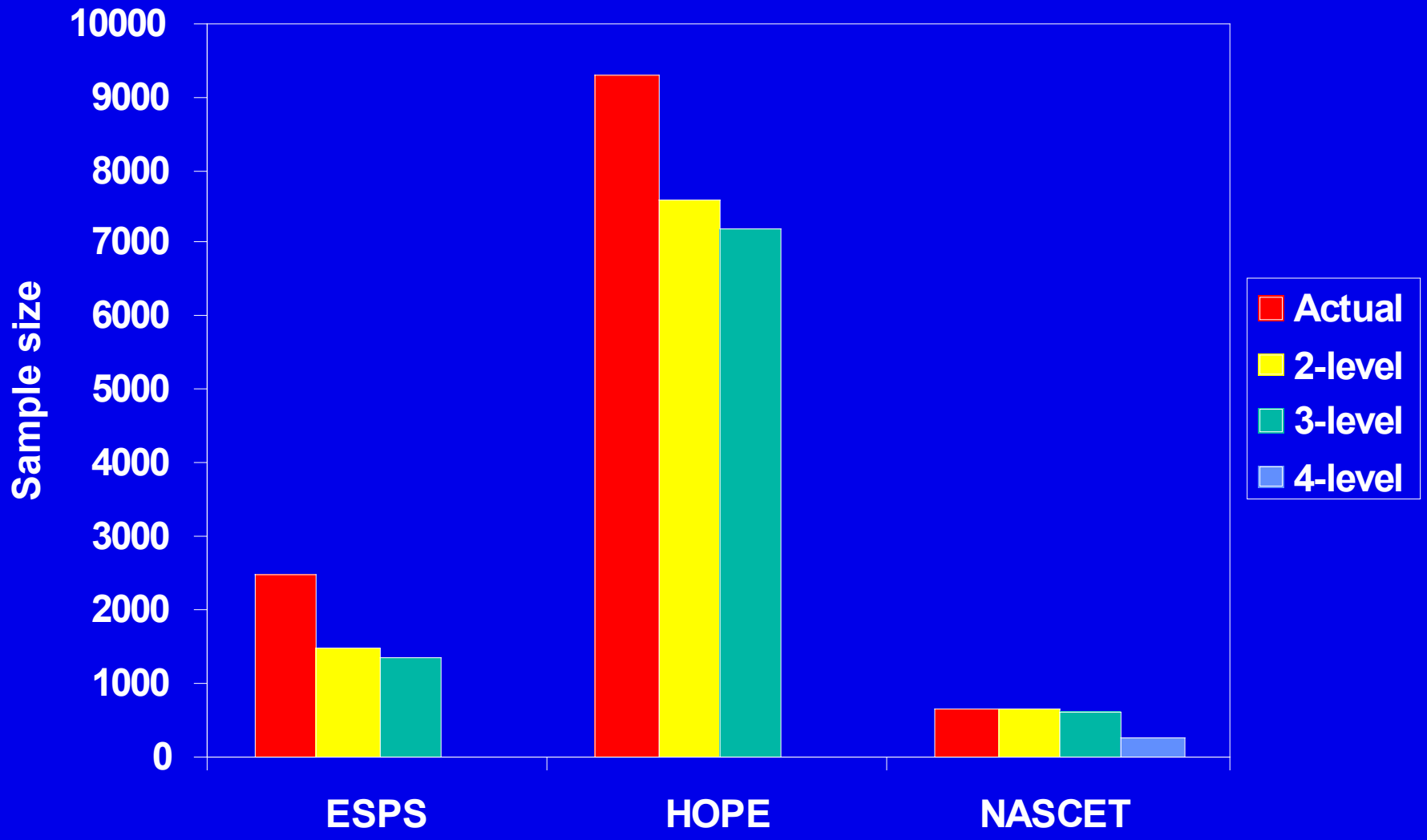
- Prevention: primary, secondary
- Interventions: anticoagulants, antiplatelets, antihypertensives, lipid lowering, endarterectomy, hormone replacement
- Effect direction: positive, negative (HRT)

Approach worked for different trial designs:

- Small / large size
- Short / long follow-up

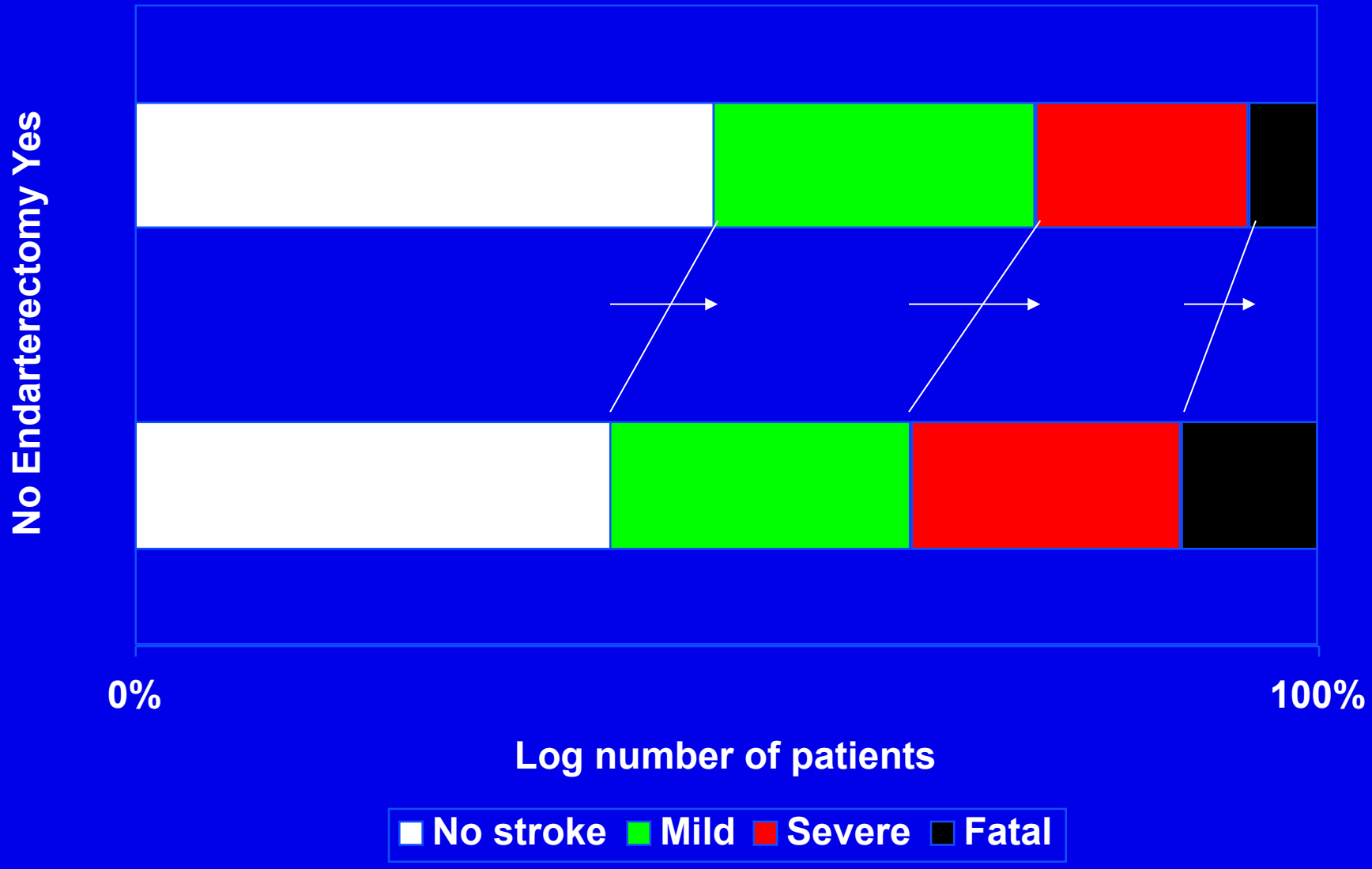
Bath *et al*. *Stroke* 2008;39:2817-23.

OA-prevention: Sample size calculations



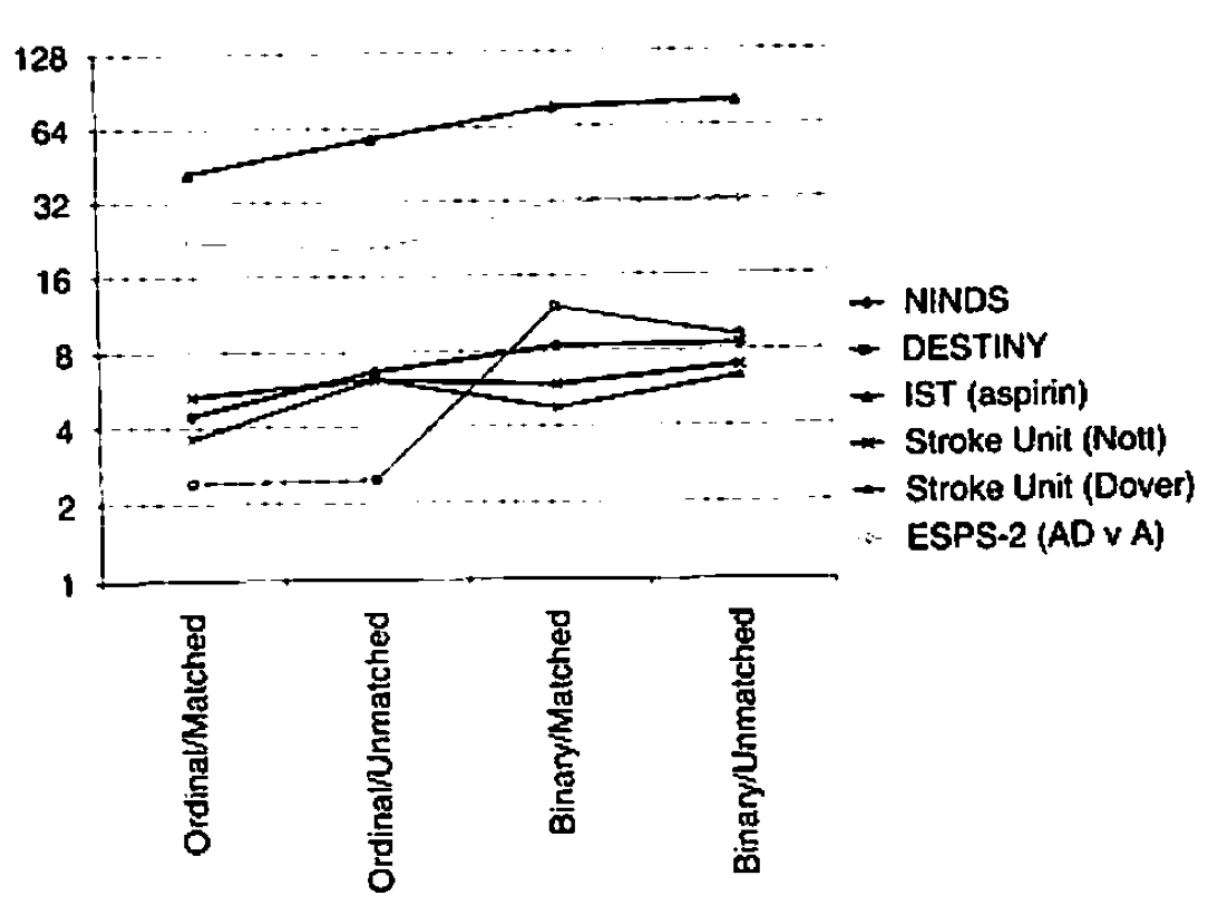
Number-needed-to-treat: NASCET

Acute stroke, rt-PA: dichotomy NNT~9, ordinal NNT~3



OAST: Prevention trial

Ordinal Numbers Needed to Treat (NNT) can be calculated and are smaller, i.e. better, than binary



Bath *et al*. *Int J Stroke* 2011;6:472-9.

OA-prevention: Ordinal meta-analysis

STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:2243–2260 (DOI: 10.1002/sim.919)

Meta-analysis of ordinal outcomes using individual patient data

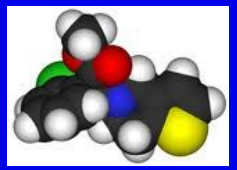
Anne Whitehead[1,*,†], Rumana Z. Omar[2], Julian P. T. Higgins[3], Elly Savaluny[4], Rebecca M. Turner[5] and Simon G. Thompson[3]

[1] *Medical and Pharmaceutical Statistics Research Unit, The University of Reading, P.O. Box 240, Earley Gate, Reading RG6 6FN, U.K.*
[2] *Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.*
[3] *MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, U.K.*
[4] *Biometrics Department, Pfizer Central Research, Ramsgate Road, Sandwich, Kent CT13 9NJ, U.K.*
[5] *MRC Clinical Trials Unit, 222 Euston Road, London NW1 2DA, U.K.*

TARDIS: Ordinal outcomes

First vascular trial to be designed with ordinal rather than binary outcome:

- Stroke (5):
  - Fatal / mRS 3-5 / mRS 0-2 / TIA / none
- MI (3):
  - Fatal / non-fatal / none
- Vascular (3):
  - Fatal / non-fatal / none
- Bleeding (5):
  - Fatal / severe / moderate / mild / none
- Adverse events (4):
  - Fatal / SAE / AE / none




TARDIS Sample size calculation, 1

Stroke/TIA: Distribution of events

mRS (%)	Index event	N	n	Stroke	6/Dead	2-5	0,1	TIA	No event
TARDIS	Both	392	25	2.81	0.51	0.77	1.53	3.57	93.62
	TIA	160	10	3.13	0.63	0.63	1.88	3.13	93.75
	Stroke	232	15	2.59	0.43	0.86	1.29	3.88	93.53
ENOS [43]	Stroke	2,239	82	3.66	1.30	2.01	0.36	NA	NC
FASTER [28]	Both	392	35	8.9	-	-	-	NA	NC
EARLY [29]	Stroke	543	49	9.00	-	-	-	1.84	89.16
PRoFESS early [30]	Stroke	1,360	31	2.28	-	-	-	NA	NC

Alpha 0.05, power 0.90

Odds ratio 0.68 (= OR 0.57 if binary)

(EARLY 0.54, CARESS 0.40, CLAIR 0.74)

Crossovers 5% (2.1%), losses to follow-up 2% (0.8%)

N=4,100 (2,050 per group)

Would need 8,900 if stroke alone with OR 0.68

Power for stroke alone as is = 0.626

Bath *et al. HTA Grant Application* 2011

TARDIS Sample size calculation, 2

Bleeding: Distribution of events

Bleeding (%)	Index	N	n	Fatal	Major	Moderate	Minor	None
TARDIS	Both	392	54	0.00	1.79	1.28	10.71	86.22
	TIA	160	20	0.00	0.00	1.88	10.63	87.50
	Stroke	232	34	0.00	3.02	0.86	10.78	85.34

With N=4,100 and for alpha 0.05, power 0.90

Crossovers 5%, losses to follow-up 2%, covariate adjustment

Can detect odds ratio 1.32

Bath *et al*. *HTA Grant Application* 2011

OA-prevention: Individual patient data

To assess the relative efficiency of dichotomous versus ordered categorical outcomes and their analysis using ordinal and binary statistical tests

- Individual patient data from vascular prevention trials
- Same outcomes
- Same number of levels
  - But more trials with 5+ levels based on more datasets
- Same statistical tests
- Same questions: efficiency of tests, sample size benefits, NNT, covariate adjustment

Issue 1: Statistical assumptions

- Dichotomous - few
- Ordinal – some, e.g. proportional odds
  - Reasonably robust
  - Can use alternative ordinal approaches
- Continuous – many, e.g. normal distribution
  - Robust if large samples, even with ordinal data (central limit theorem)

- To be examined in OA-prevention using IPD

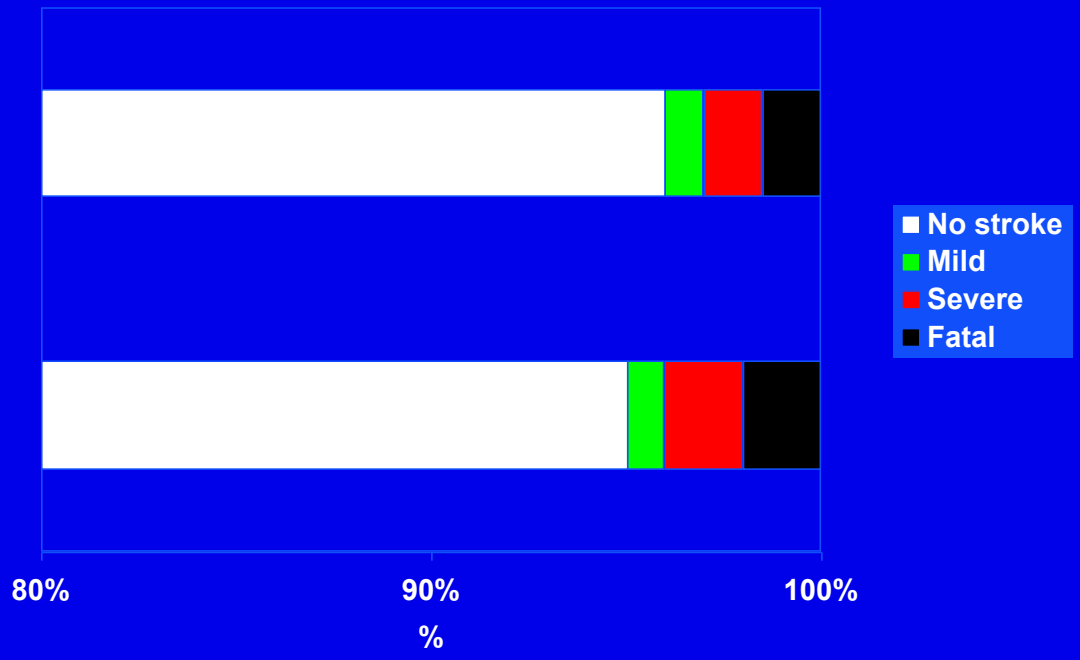
Issue 2: Interpretation of results

- Dichotomous – easy, use absolute risk reduction
- Ordinal – can be challenging, use odds ratio (vs relative risk) or difference in medians
  - 'Lowering BP reduces recurrence and its severity'
- Continuous – moderately challenging; use difference in means
  - Is difference of mRS 0.3 useful? Minimum Important Difference

- Can use NNT for any of the above
- Can use QALYs or DALYs
- Health economics

- To be examined in OA-prevention using IPD

Issue 3: Low event rates



- If the total rate is very low, say <5%, ordinalising will not help!
- No binary, ordinal or continuous system can be sensitive to treatment effects if very low event rate and sample size not massive
- To be examined in OA-prevention using IPD

Issue 4: Late events

- Events occurring late in follow-up will not change the event frequency but will bias severity since little chance for patient to recover to stable state
- Unlikely to be a major issue because:
  - Acute stroke - most events occur early, not late, in 3 month follow-up
  - Chronic stroke – events spread out across all of follow-up so few proximal to end
- To be examined in OA-prevention using IPD

Issue 5: Death

- Death or vascular death?
- mRS includes all death
- Not using all death will mean missing data
- So use all death

Issue 6: Ordinal versus dichotomous

- Ordinal is not a panacea for poor interventions or trial design
- Ordinal/dichotomous design/analysis does not always beat dichotomous - no guarantee!
  - ECASS-I

Conclusion: Ordinal (vs dichotomous)

- Applies to acute stroke
- May apply to vascular and stroke prevention trials
- Advantages probably outweigh disadvantages
- Smaller less complex trials (for a given power)
- Provides additional information on outcome severity
- Net NNTs better
- Interpretation and presentation of results might be more challenging
  - Use mixed model – design/analyse as ordinal/continuous but present (non-significant) binary results
- More work needed to assess pros and cons