

Can we improve the statistical analysis of stroke trials? Statistical re-analysis of functional outcomes in stroke trials.

The Optimising Analysis of Stroke Trials (OAST) Collaboration

Division of Stroke Medicine, University of Nottingham, Nottingham, UK; Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK

Correspondence and requests for reprints to:

Professor Philip Bath

Division of Stroke Medicine

University of Nottingham

D Floor, South Block

Queen's Medical Centre

Nottingham NG7 2UH UK

Tel: +44 115 8231023

Fax: +44 115 8231022

E-mail: philip.bath@nottingham.ac.uk

ACKNOWLEDGEMENTS

Secretariat and writing committee: Philip MW Bath (chief investigator, Nottingham, UK); Laura J Gray (lead statistician, Nottingham, UK); Timothy Collier (statistical advisor, London, UK); Stuart Pocock (statistical advisor, London, UK). We declare no conflicts of interest.

Statistical advisor: James Carpenter (Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK)

The following contributors provided individual patient data from their trial, and commented on the draft manuscript:

Abciximab: H Adams (USA), K Dougherty (USA); W Hacke (Germany),

ASK: G Donnan (Australia)

ASSIST 07 & 10: S Davis (Australia)

ATLANTIS A & B: G Albers, S Hamilton (USA)

BEST Pilot & Main: D Barer (UK)

Citicoline 1, 7, 10, 18: A Davalos (Spain)

Corr: S Corr (UK)

Dover Stroke Unit: P Langhorne (UK)

DCLHb: P Koudstaal, R Saxena (Netherlands)

Ebselen: T Yamaguchi (Japan)

ECASS II: W Hacke, E Bluhmki (Germany)

Factor VII: S Mayer (USA), K Begtrup (Denmark)

FISS: R Kay (Hong Kong)

FOOD 3: M Dennis (UK)

Gilbertson: L Gilbertson (UK)

INWEST: N-G Wahlgren, N Ahmed (Sweden)

IST: P Sandercock (UK)

Kuopio Stroke Unit: J Sivenius (Finland)

Logan: P Logan (UK)

MAST-I: L Candelise (Italy), J Wardlaw (UK)

Newcastle Stroke Unit: H Rodgers (UK)

NINDS: J Marler (USA)

Parker: C Parker (UK)

Nottingham Stroke Unit: N Lincoln, P Berman (UK)

RANNTAS I & II, STIPAS, TESS I & II: P Bath (UK), B Musch (USA)

Walker 1 & 2: M Walker (UK)

Young: J Young, A Forster (UK)

We thank the patients who took part in these studies, and the trialists who shared their data. The study was conceived, initiated, managed, analysed, and interpreted independently of any pharmaceutical company. Each collaborator listed above commented on the draft manuscript.

FUNDING

LJG is funded, in part, by BUPA Foundation and The Stroke Association (UK). PB is Stroke Association Professor of Stroke Medicine. The funding sources had no involvement in this project.

TITLE PAGE

Full title: Can we improve the statistical analysis of stroke trials?

Cover title: Can we improve the statistical analysis of stroke trials?

Table 1: Comparison of rank scores for 16 statistical tests

Figure 1: Distribution in Modified Rankin Scale in the ECASS II trial showing the primary and post hoc analyses

Figure 2: Identification of included trials

Figure 3: Percentage of trials significant at the 5% level with each statistical test

Keywords: stroke; randomised controlled trial; statistical analysis

Word count: 3,925

ABSTRACT

Background

Most large acute stroke trials have been neutral. Functional outcome is usually analysed using a yes or no answer, e.g. death or dependency vs. independence. We assessed which statistical approaches are most efficient in analysing outcomes from stroke trials.

Methods

Individual patient data from acute, rehabilitation and stroke unit trials studying the effects of interventions which alter functional outcome were assessed. Outcomes included modified Rankin Scale, Barthel Index, and '3 questions'. Data were analysed using a variety of approaches which compare two treatment groups. The results for each statistical test for each trial were then compared.

Results

Data from 55 datasets were obtained (47 trials, 54,173 patients). The test results differed substantially so that approaches which use the ordered nature of functional outcome data (ordinal logistic regression, t-test, robust ranks test, bootstrapping the difference in mean rank) were more efficient statistically than those which collapse the data into 2 groups (chi square) (ANOVA $p < 0.001$). The findings were consistent across different types and sizes of trial and for the different measures of functional outcome.

Conclusions

When analysing functional outcome from stroke trials, statistical tests which use the original ordered data are more efficient and more likely to yield reliable results.

Suitable approaches included ordinal logistic regression, t-test, and robust ranks test.

BACKGROUND

The management of patients with acute or recent stroke has benefited significantly from the results of randomised controlled trials and meta-analyses of these. For example, functional outcome is improved with alteplase, aspirin, management in a Stroke Unit, and community occupational therapy.¹⁻⁷ In contrast, some studies were overtly negative finding that treatment worsened outcome, e.g. DCLHb, enlimomab, selfotel, or tirilazad.⁸⁻¹¹ However, the majority of acute stroke trials were neutral in spite of positive preclinical findings. The failure of these latter studies can be attributed to multiple causes, including the relevance of laboratory findings to clinical stroke,¹² inadequate sample size,¹³ choice of primary outcome, and its statistical analysis.

Measures of functional outcome such as the modified Rankin Scale (mRS)¹⁴, Barthel Index (BI)¹⁵ and '3-questions'¹⁶ are ordinal in nature, that is, they consist of 3 or more categories which have a natural ordering, e.g. the mRS has 7 categories ranging from no symptoms to dead. It might then be expected that statistical analysis would preserve and utilise the data in this ordinal form. However, most published trials have used a 'yes/no' (dichotomised) analysis of functional outcome, e.g. combining categories within the mRS into two groups, such as 'dead or dependent' (e.g. mRS 3-6) and 'independent' (mRS 0-2), and then comparing these between the treatment groups. Unfortunately, there is little agreement where mRS data should be divided: i.e. 0,1 vs. 2-6,¹ 0-2 vs. 3-6,¹⁷ or 0-3 vs. 4-6,¹⁸ and whether this matters.¹⁹ Further, collapsing data in this way generally lowers statistical power and therefore reduces the chance of finding a significant treatment effect since information from many subjects are ignored. For example, patients responding to treatment and achieving a

mRS of 3 rather than 4 or 0 rather than 1 are not detected in a analysis comparing mRS 0-2 with 3-6.

Inadequacies in the statistical analysis of trials in acute stroke are apparent in two examples. First, the ECASS II trial of alteplase showed no treatment effect for its primary outcome (when comparing mRS 0,1 with mRS 2-6) but was positive when re-analysed using the data collapsed in a different place (mRS 0-2 vs. 3-6)²⁰ or when analysed using a 'bootstrapping' technique (figure 1).²¹ Second, five trials of tirilazad individually showed no treatment effect when analysed using dichotomous outcomes²²⁻²⁴ although a meta-analysis found that the intervention was associated with a worse outcome;²⁵ *post hoc* analysis then suggested that one of these trials was negative²⁴ (not neutral) when analysed using a method which preserved the original ordered data (P Bath, unpublished data).

We aimed to identify which statistical methods might optimise the analysis of data from functional outcome scales in stroke trials.

METHODS

Identification of trials

We sought individual patient data from randomised controlled trials assessing functional outcome after stroke for interventions which were either positive or negative according to the trial publication, or were included in a meta analysis showing benefit or harm; neutral trials in a neutral meta-analysis were excluded. Published studies (full paper or abstract) fulfilling these criteria were identified from electronic searches of the Cochrane Library (to end of 2005). In each case, we invited the chief investigator to join the collaboration and share their data. In some cases where individual data could not be obtained it was possible to extract it from the original publication.

Trial data

Demographic (age, gender), trial (setting, intervention, length of follow up, result), patient severity, and functional outcome (BI, mRS, '3 question' scale [3Q, a derivative of mRS], or another measure) data were collected for each trial. In factorial trials or those having more than two treatment groups, data were analysed for each comparison of active therapy versus control. Where outcome data were scored at several time points (e.g. 1, 3 and 6 months) the time point used for the primary outcome was included.

Statistical tests

We compared different statistical tests for assessing treatment effect. Some of these required the data to be collapsed into groups (such as the chi square test) while others used the original ordinal data (such as Wilcoxon test and t-test). Statistical tests which dichotomised ('yes/no') data were assessed multiple times collapsing the

data in different places, e.g. mRS 0,1 vs. 2-6, 0-2 vs. 3-6 and 0-5 vs. 6. A description of the statistical tests used is given in <http://www.nottingham.ac.uk/stroke-medicine/oast/oastappendix1.doc>.

Comparison of statistical tests

Each data set was analysed using each statistical test. These results were then ordered within each trial and given a rank, with the lowest rank given to the test which produced the most significant result, i.e. the largest z score, within that trial. A two-way analysis of variance test was then used to see on average which statistical test had produced the lowest ranks. We were then able to order the statistical tests in terms of their efficiency in identifying treatment effects. We also assessed how many statistically significant (at 5%) results each test found.

To assess the validity and reliability of the results, a number of supplementary analyses were carried out. First, the comparison of statistical tests was repeated within sub-groups of trials sharing similar characteristics; second, the statistical assumptions of the tests were assessed; and last, the sensitivity of the tests was explored to make sure treatment effects were only detected when they truly existed (the type one error rate). Technical details of these supplementary analyses can be found in <http://www.nottingham.ac.uk/stroke-medicine/oast/oastappendix2.doc>.

Analyses were carried out in SAS (version 8.2) and Stata (version 7) and significance was taken at $p < 0.05$.

RESULTS

Trials characteristics

A total of 55 comparisons of active versus control treatment (54,173 patients) were included, these comprising individual patient data from 38 trials and summary data extracted from the publications of a further 9 studies; six trials had two active treatment groups, and one had three active groups so a further 8 comparisons were available (figure 2). The data related to 34 acute stroke trials, 7 trials of rehabilitation (1,164 patients) and 6 trials of stroke units (1,399 patients). BI was used to measure functional outcome in 22 trials, 18 used the mRS, 3 used the 3Q scale, 1 used the Rivermead scale, 2 related trials used the Nottingham ADL scale, and 1 trial used its own ordinal measure.²⁶ Included trials studied the following interventions: abciximab (AbESTT); alteplase (ATLANTIS A & B, ECASS II, NINDS); aspirin (CAST, IST); atenolol (BEST); citicoline; DCLHb; ebselen; edaravone; enlimomab (EAST); factor VIIa; feeding (FOOD 3); nadroparin (FISS, FISS-TRIS); nimodipine (INWEST); occupational therapy (Corr, Gilbertson, Logan, TOTAL, Walker); physiotherapy (Young); pro-urokinase (PROACT II); selfotel (ASSIST); streptokinase (ASK, MAST-E, MAST-I); stroke unit (Dover, Helsinki, Kuopio, Nottingham, Orpington, Newcastle); and tirilazad (RANTTAS I & II, STIPAS, TESS I & II). Data relating to 16 trials or interventions which fulfilled the inclusion criteria were not made available.

The method of analysing functional outcome used in the original trial publication varied considerably, see <http://www.nottingham.ac.uk/stroke-medicine/oast/oastappendix3.doc>. 23 (48.9%) trials assessed the treatment effect using a method which required the data to be collapsed into groups, e.g. chi-square test; 17 (36.2%) used a test based on comparing medians and 4 (8.5%) used a test

which compared means; the remaining trials were unpublished so the method of analysis is not known.

Comparison of statistical tests

The statistical tests assessed differed significantly in the results they gave for each trial (2 way ANOVA $p < 0.0001$). The ordering of the tests showed that those which analyse the original ordinal data generally perform better than those which collapse the data into 2 or more groups. The most efficient tests included ordinal logistic regression, t-test, robust rank test and bootstrapping the difference in mean rank (table 1). The sub group analysis showed the same ordering of tests irrespective of type of intervention (acute, rehabilitation, stroke unit), trial size, time between randomisation and onset, patient age, baseline severity, outcome measure, length of follow up, and trial result (<http://www.nottingham.ac.uk/stroke-medicine/oast/oastappendix4.doc>).

When assessed by how many trials were statistically significant, those tests which did not collapse the data into groups again out-performed the other approaches; for example, ordinal logistic regression (using raw data) gave a statistically significant result in 25.9% of trials whereas the 2x2 chi-square test comparing death or poor outcome to an excellent outcome only gave a significant result in 9.3% of the trials (figure 3).

Test assumptions and sensitivity

The statistical assumptions of the t-test were not met for the majority of trials and the assumptions of the ordinal logistic regression analysis failed for 8 out of the 55 data sets; in contrast, the assumptions for the other tests were maintained. The sensitivity analysis showed that the top performing statistical tests were not overly sensitive and

statistically significant treatment effects were only found where they truly existed; see <http://www.nottingham.ac.uk/stroke-medicine/oast/oastappendix5.doc> for detailed results.

DISCUSSION

These results show that statistical approaches which analyse the original ordinal data for functional outcome are more efficient than those which work on pre-processed data which has been collapsed into 2 or more groups. Interestingly, this point was originally demonstrated mathematically by Shannon in 1948²⁷. In particular, ordinal logistic regression, t-test, robust ranks test, and bootstrapping (the difference in mean rank) performed well and appear to be useful irrespective of the type of stroke trial, patient or intervention. Although individual tests based on dichotomised data using Chi-square analysis (e.g. 'dead/dependent' versus 'independent') were effective for some data sets, they performed poorly in many and therefore cannot be recommended as general solutions for analysing stroke trials. From an historical perspective, it is quite possible that trials which collapsed mRS or BI in two groups may have used a sub-optimal analysis, and this may have contributed to false neutral findings in some cases in the past. For example, MAST-E²⁸ and STIPAS²⁴ were neutral as reported using dichotomous analysis but negative when assessed with ordinal approaches.

Several comments can be made about this study. First, it aimed to include data from all stroke trials assessing a beneficial or harmful intervention. Unfortunately, data were not made available for all identified trials; where possible, we created individual data from publications which provided patient numbers by outcome score. Data were missing for a variety of trial types (acute/rehabilitation/stroke unit) and sizes, and functional outcome measure (mRS/BI), so it is unlikely that a systematic bias was introduced into the findings; however, the precision of the results may have been attenuated by the missing trials. Second, we did not exhaustively search for all possible statistical tests relevant to the problem of analysing ordered categorical data;

instead, we focussed on those approaches which are available in standard statistical textbooks and computer packages. Additionally, we could not include some tests used in recent trials, e.g. patient specific outcomes ²⁹ and Cochran Mantel Haenszel test ³⁰ since these require access to individual data for both baseline and outcome variables, and these data were not available uniformly. Third, some of the statistical assumptions underlying the more efficient tests were not met in all trials; for example, the t test assumes data are normally distributed while ordinal logistic regression assumes that any treatment effect is similar across outcome levels ('proportionality of odds', i.e. the odds of moving a treated patient from mRS 2 to 1 is similar to that for moving them from 5 to 4). Nevertheless, the robustness of these tests to deviations from their underlying assumptions means that they remain relevant for analysing functional outcome data from stroke trials.

If alternative approaches to analysing functional outcome data are to be used in the future, it is pertinent to ask how sample size should be calculated at the trial design stage. Historically, most calculations assumed that functional outcome would be dichotomised and analysed using a Chi-square test approach.¹³ Although future trials could continue to calculate sample size in the same way (and then gain extra power by analysing their data using an ordinal approach), specific sample size calculations are available when data are to be analysed using ordinal logistic regression³¹ or the t-test. Ideally, the extra power gained by using an ordinal statistical approach should not be used to reduce sample size; stroke trials have been too small in the past, as shown in a recent meta analysis,¹³ and this may also have contributed to the failure of some of them.

A further issue with using a statistical test which analyses ordered categorical data is how to report the results to patients, carers, clinicians, and health policy makers. The

results of dichotomous tests may be summarised easily as the proportion of patients who benefit (or suffer) with a treatment, i.e. alteplase reduced absolute death or dependency (mRS>1) by 13% in the NINDS part 2 trial.¹ In contrast, ordinal tests will need to be presented as the average absolute improvement in outcome, e.g. alteplase improved the mRS by 1 (of 7) point and BI by 22.5 (of 100) points. Alternatively, the combined odds ratio and its confidence intervals would be reported if ordinal logistic regression was used. In this respect, health consumers will need to decide what differences in mRS and BI are worthwhile, both clinically and in terms of health economics. In reality, it is reasonable to present the effect on functional outcome using both absolute percentage change and mean or median change in functional outcome score, and show this data graphically (as in figure 1).

In summary, we suggest that ongoing and future trials should consider using statistical approaches which utilise the original ordered categorical data in the primary analysis of functional outcome measures. Such ordinal tests include ordinal logistic regression, and the robust ranks test; the t-test may also be used although its assumptions were not meant in the majority of trials.

REFERENCES

1. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute stroke. *New England Journal of Medicine*. 1995;333:1581-1587
2. CAST (Chinese Acute Stroke Trial) Collaborative Group. Cast: Randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. *Lancet*. 1997;349:1641-1649
3. Walker M, Gladman J, Lincoln N, Siemonsma P, Whiteley T, 1996. Occupational therapy for stroke patients not admitted to hospital: A randomised controlled trial. *The Lancet*. 1999;354:278-280
4. Chen ZM, Sandercock P, Pan HC, Counsell C, Collins R, Liu LS, Xie JX, Warlow C, Peto R, On behalf of the CAST and IST Collaborative Groups. Indications for early aspirin use in acute ischemic stroke - a combined analysis of 40 000 randomized patients from the chinese acute stroke trial and the international stroke trial. *Stroke*. 2000;31:1240-1249
5. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *The cochrane library*. Oxford: Update Software; 2002.
6. The ATLANTIS ECASS and NINDS rt-PA Study Group Investigators. Association of outcome with early stroke treatment: Pooled analysis of atlantis, ecass, and ninds rt-pa stroke trials. *The Lancet*. 2004;363:768-813
7. Walker MF, Leonardi-Bee J, Bath P, Langhorne P, Corr S, Drummond A, Gilbertson L, Gladman JRF, Jongbloed L, Parker C. An individual patient data meta-analysis of randomised controlled trials of community occupational therapy for stroke patients. *Stroke*. 2004;35:2226-2232
8. Saxena R, Wijnhoud AD, Carton H, Hacke W, Kaste M, Przybelski RJ, Stern KN, Koudstaal PJ. Controlled safety study of a hemoglobin-based oxygen carrier, dclhb, in acute ischemic stroke. *Stroke*. 1999;30:993-996
9. Davis SM, Lees KR, Albers GW, Diener HC, Markabi S, Karlsson G, Norris J, for the ASSIST Investigators. Selfotel in acute ischemic stroke. Possible neurotoxic effects of an nmda antagonist. *Stroke*. 2000;31:347-354
10. Enlimomab acute stroke trial investigators. Use of anti-icam-1 therapy in ischemic stroke: Results of the enlimomab acute stroke trial. *Neurology*. 2001;57:1428-1434
11. Bath PMW, Blecic S, Bogousslavsky J, Boysen G, Davis S, Diez-Tejedor E, Ferro JM, Gommans J, Hacke W, Indredavik B, Norrving B, Orgogozo JM, Ringelstein EB, Sacchetti ML, Iddenden R, Bath FJ, Musch BC, Brosse DM, Naberhuis-Stehouwer SA. Tirilazad mesylate in acute ischemic stroke - a systematic review. *Stroke*. 2000;31:2257-2265
12. Grotta J. Why do all drugs work in animals but none in stroke patients? 2 neuroprotective therapy. *Journal of Internal Medicine*. 1995;237:89-94
13. Weaver CS, Leonardi-Bee J, Bath-Hexall FJ, Bath PMW. Sample size calculations in acute stroke trials: A systematic review of their reporting, characteristics, and relationship with outcome. *Stroke*. 2004;35:1216-1224
14. Rankin J. Cerebral vascular accidents in patients over the age of 60. 2. Prognosis. *Scottish Medical Journal*. 1957;2:200-215
15. Mahoney FI, Barthel DW. Functional evaluation: The barthel index. *Maryland State Medical Journal*. 1965:61-65
16. Lindley RI, Waddell F, Livingstone M, Sandercock P, Dennis MS, Slattery J, Smith B, Warlow C. Can simple questions assess outcome after stroke? *Cerebrovascular Diseases*. 1994;4:314-324
17. Bath P, Lindenstrom E, Boysen G, De Deyn P, Friis P, Leys D, Marttila R, Olsson J-E, O'Neill D, Orgogozo J-M, Ringelstein B, van der Sande J-J, Turpie AGG. Tinzaparin in acute ischaemic stroke (taist): A randomised aspirin-controlled trial. *Lancet*. 2001;358:702-710

18. Horn J, deHaan R, Vermeulen M, Limburg M. Very early nimodipine use in stroke (venus). A randomised, double-blind, placebo-controlled trial. *Stroke*. 2001;32:461-465
19. Wardlaw J, Sandercock P, Warlow C, Lindley RI. Trials of thrombolysis in acute ischemic stroke. Does the choice of primary outcome measure really matter? *Stroke*. 2000;31:1133-1135
20. Hacke W, Markku K, Fieschi C, von Kummer R, Davalos A, Meier D, Larrue V, Bluhmki E, Davis S, Donnan G, Schneider D, Diez-Tejedor E, Trouillas P. Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ecass ii). *Lancet*. 1998;352:1245-1251
21. Stingele R, Bluhmki E, Hacke W. Bootstrap statistics of ecass ii data: Just another post hoc analysis of a negative stroke trial? *Cerebrovascular Diseases*. 2001;11:30-33
22. Haley EC. High-dose tirilazad for acute stroke (ranttas ii). *Stroke*. 1998;29:1256-1257
23. The RANTTAS Investigators. A randomized trial of tirilazad mesylate in patients with acute stroke (ranttas). *Stroke*. 1996;27:1453-1458
24. The STIPAS Investigators. Safety study of tirilazad mesylate in patients with acute ischemic stroke (stipas). *Stroke*. 1994;25:418-423
25. The Tirilazad International Steering Committee. Tirilazad for acute ischaemic stroke (cochrane review). Oxford: Update Software; 2002.
26. Barer DH, Cruickshank JM, Ebrahim SB, Mitchell JR. Low dose beta blockade in acute stroke ("Best" Trial): An evaluation. *British Medical Journal*. 1988;296:737-741
27. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948; 27:379-423.
28. The Multicenter Acute Stroke Trial - Europe Study Group. Thrombolytic therapy with Streptokinase in acute ischemic stroke. *The New England Journal of Medicine*. 1996; 335(3): 145-149.
29. Adams HP, Leclerc JR, Bluhmki E, Clarke WR, Hansen MD, Hacke W. Measuring outcomes as a function of baseline severity of ischemic stroke. *Cerebrovascular Diseases*. 2004;18:124-129
30. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener H-C, Grotta J, Lyden P, Shuaib A, Hardernark H-G, Wasiewski WW, For the Stroke-Acute Ischemic NXY Treatment (SAINT I) Trial Investigators. Nxy-059 for acute ischemic stroke. *New England Journal of Medicine*. 2006;354:588-600
31. Whitehead J. Sample-size calculations for ordered categorical-data *Statistics in Medicine*. 1993;12:2257-2271

Table 1. 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 $p < 0.05$.

Test	Mean rank	No. of datasets	Banding
Ordinal logistic regression	6.11	54	
t-test	6.51	55	
Robust ranks test	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	
Chi Sq – death or poor outcome vs good	8.87	55	
Chi Sq – death or poor outcome vs excellent	9.24	54	
Median test	9.47	55	
Chi Sq – 2x3 test	9.96	51	
Chi Sq – death vs alive	9.98	51	
Chi Sq – 2x4 test	10.02	50	
Kolmogorov-Smirnov test	11.29	55	

FIGURE LEGENDS

Figure 1

Distribution in Modified Rankin Scale in the ECASS II trial showing the primary and post hoc analyses ²⁰

Figure 2

Identification of included trials

Figure 3

Percentage of trials significant at the 5% level with each statistical test

Figure 1

Distribution in Modified Rankin Scale in the ECASS II trial showing the primary and post hoc analyses ²⁰

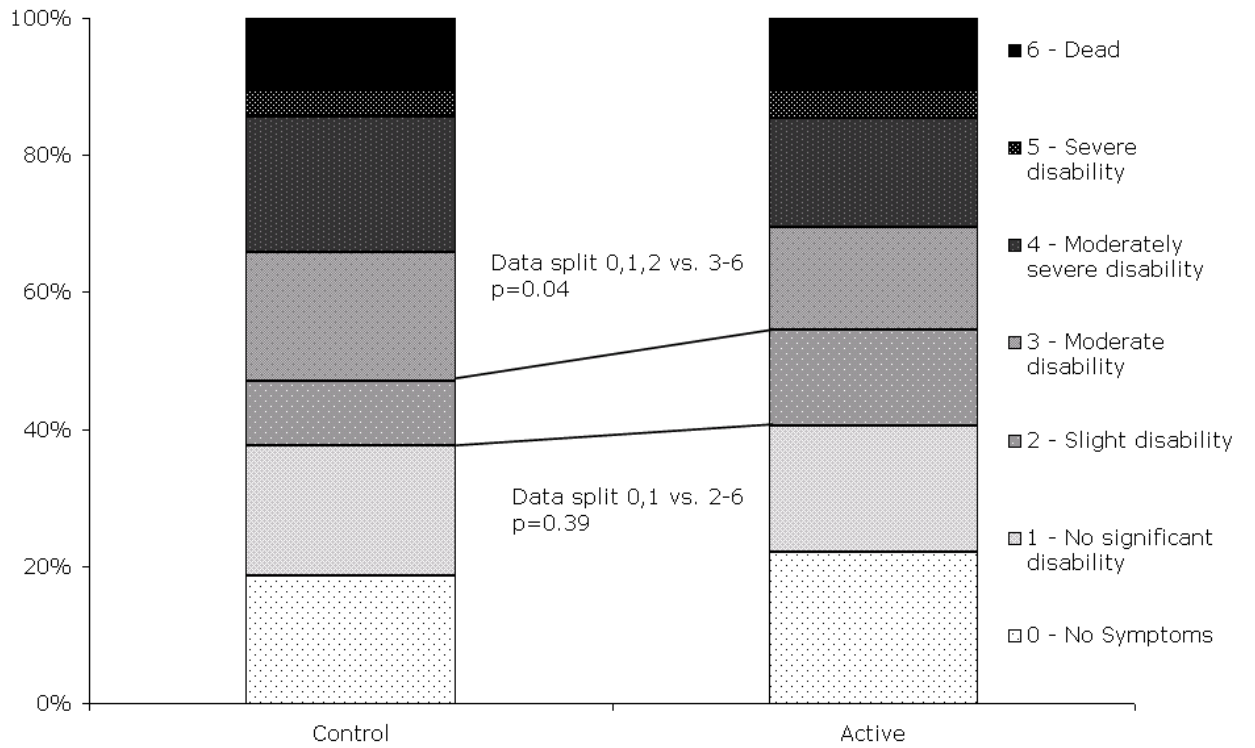


Figure 2

Identification of included trials

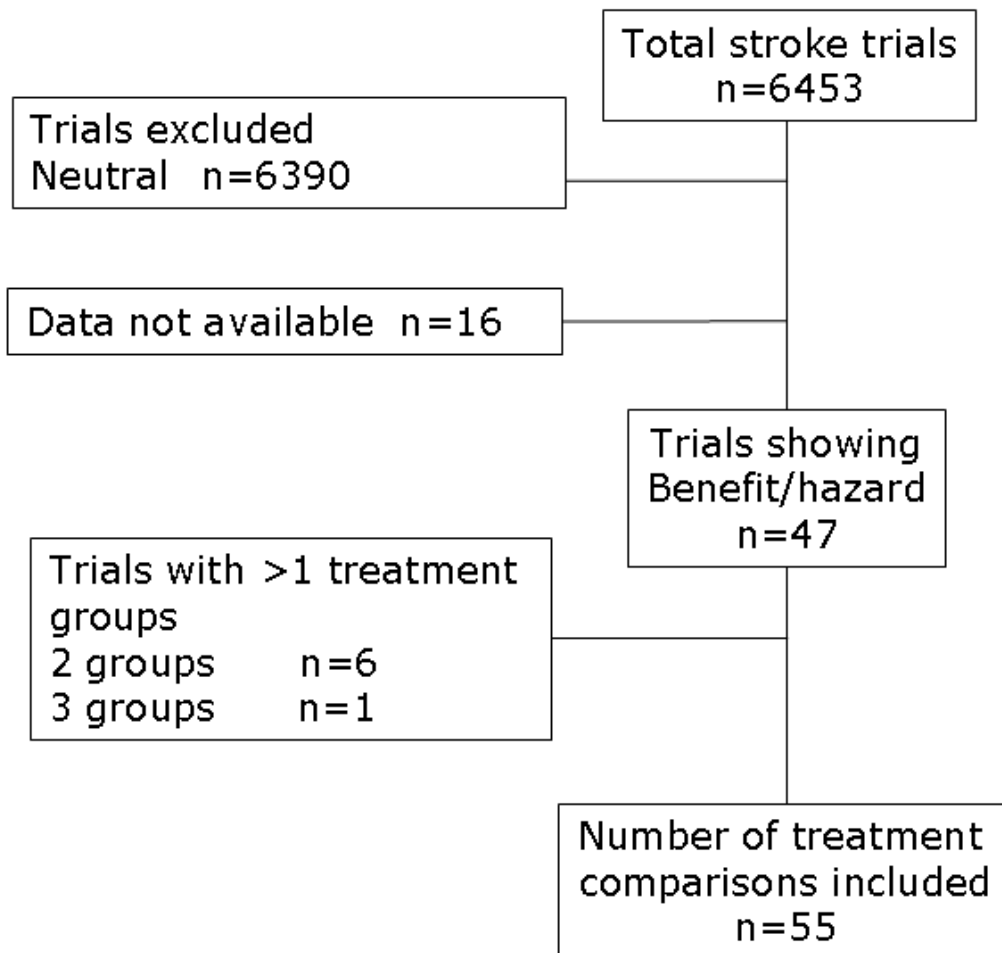
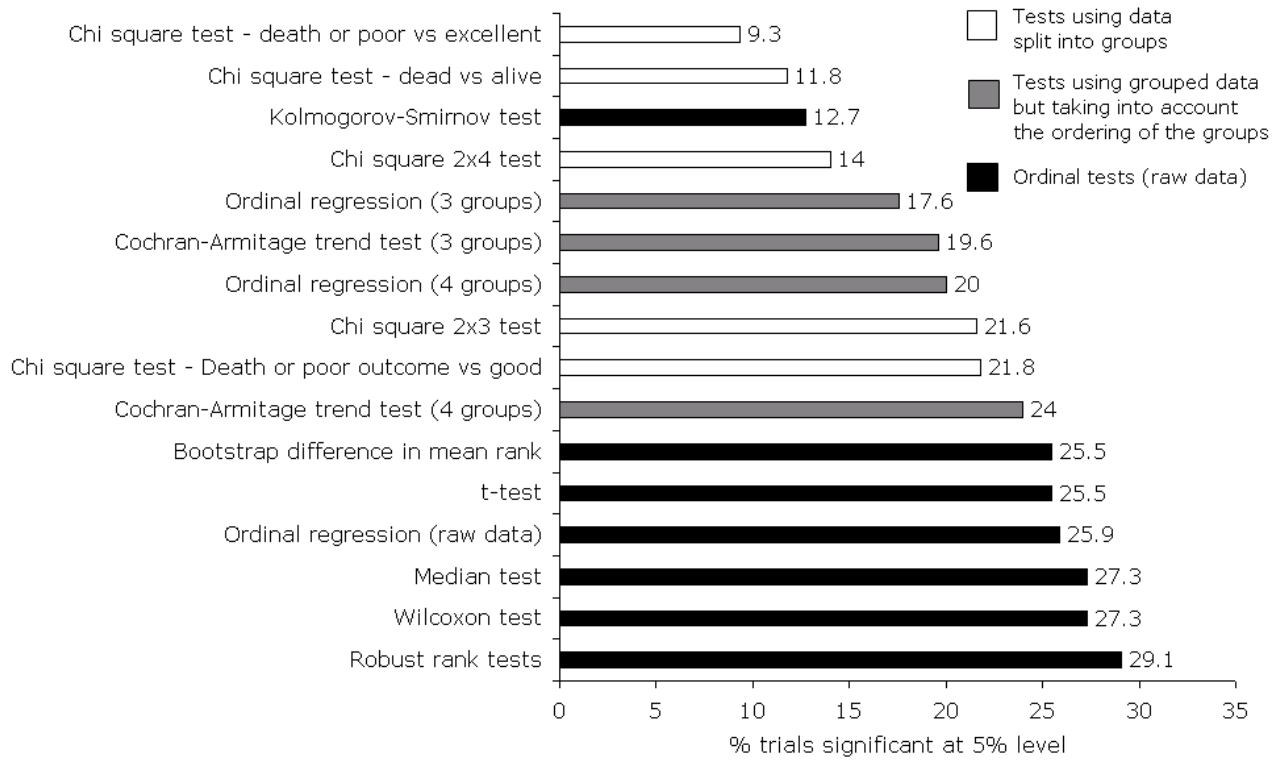


Figure 3

Percentage of trials significant at the 5% level with each statistical test



APPENDICES

OAST Appendix 1

Statistical tests compared

OAST Appendix 2

Supplementary analyses

OAST Appendix 3:

Trial data

OAST Appendix 4

Results

OAST Appendix 5

Results

OAST Appendix 1: Statistical tests compared

Included tests

Univariate statistical approaches for analysing dichotomous and ordinal data comprised tests based on Chi-square, ordinal, and bootstrap approaches.[1-3] Sixteen statistical approaches were assessed: (i) Chi-square 2x2 test - death or poor outcome vs. good outcome (BI <60 vs. 60-100, mRS 3-6 vs. 0-2, 3Q 1/2 vs. 3/4); (ii) Chi-square 2x2 test - death or poor outcome vs. excellent outcome (BI <95 vs. 95/100, mRS 2-6 vs. 0/1, 3Q 1-3 vs. 4); (iii) Chi-square 2x2 test - death vs. alive; (iv) Chi-square 2x3 test (unordered data) - death vs. poor vs. good outcome; (v) Chi-square 2x4 test (unordered data) - death vs. poor outcome vs. good outcome vs. excellent outcome; (vi) Cochran-Armitage trend test (ordered data with 3 levels) - death vs. poor vs. good outcome); (vii) Cochran-Armitage trend test (ordered data with 4 levels) - death vs. poor vs. good vs. excellent outcome); (viii) ordinal logistic regression (raw data); (ix) ordinal logistic regression (3 levels) (x) ordinal logistic regression (4 levels); (xi) median test; (xii) Wilcoxon/Mann-Whitney U test (adjusted for ties); (xiii) robust ranks test (RRT [4]); (xiv) Kolmogorov-Smirnov test; (xv) t-test (unpooled variances); (xvi) bootstrap of difference in mean rank (with 3x3000 cycles [5, 6]). Chi-square tests were performed without continuity correction since most trials enrolled more than 100 patients.

Excluded tests

Three non-parametric tests were excluded: Wald-Wolfowitz runs test; Siegel-Tukey test; and the Cramer-von Mises two-sample test, on methodological grounds.[2]

Statistical detail for non-standard tests

Robust rank test

The Robust rank test is an alternative to the Wilcoxon test, it tests whether the median of one group is equal to another, but unlike the Wilcoxon test it does not assume that the distributions of the two groups are equal, i.e. it makes no assumptions about the variance of the two groups. [3, 4]

Bootstrapping

Bootstrapping is a computationally intensive method which involves resampling data from a given sample. The main advantage of bootstrapping over more traditional methods is that it does not make assumptions about the distribution of the data. In this report we bootstrap the difference in mean rank; the procedure for doing this is outlined below: [5]

1. Take a dataset, which contains N observations
2. Draw a sample with replacement of size N (using replacement means that some of the original observations may appear in the new sample more than once and some not at all)
3. Estimate the parameter of interest (here the difference in mean rank) and store the result
4. Repeat 2 and 3 many times, here we use 3 sets of 3,000 as used in the ECASS II trial [6]
5. Compare the distribution of the stored results to the actual point estimate from the original dataset

Ordinal logistic regression

Ordinal logistic regression can be used when the dependent variable is ordered categorical. It is similar to logistic regression but it simultaneously estimates multiple endpoints instead of just one. The number of endpoints it estimates is equivalent to the number of ordered categories minus one. For example if the mRS was the dependent variable of interest it would compare the following j categories:

0	vs.	1,2,3,4,5,6
0,1	vs.	2,3,4,5,6
0,1,2	vs.	3,4,5,6
0,1,2,3	vs.	4,5,6
0,1,2,3,4	vs.	5,6
0,1,2,3,4,5	vs.	6

Ordinal logistic regression provides one overall estimate for each covariate in the model and not one for each cut point. This assumes that the overall odds ratio is constant no matter which cut is taken. So, for example the odds ratio for the treatment effect would be interpreted as the odds of being in category j or above for all choices of j comparing treatment 1 to treatment 0. [7]

Supplementary tables

Table 1

Sixteen statistical tests compared and the data used for the three main scales compared. [1-3] Blank cells indicate that the data has not been collapsed into groups.

	Barthel Index	Modified Rankin Scale	3 Questions Scale
Chi square test			
Good vs. poor	≥60 vs. <60	≤2 vs. >2	≥3 vs. <3
Excellent vs. poor	≥95 vs. <95	≤1 vs. >1	4 vs. <4
Alive vs. dead	≥0 vs. -5	≤5 vs. 6	>1 vs. 1
Good vs. poor vs. dead	≥60 vs. <60-≥0 vs. -5	≤2 vs. >2-≤5 vs. 6	≥3 vs. <3->1 vs. 1
Excellent vs. good vs. poor vs. dead	≥95 vs. <95-≥60 vs. <60-≥0 vs. -5	≤1 vs. <1-≤2 vs. >2-≤5 vs. 6	4 vs. 3 vs. <3->1 vs. 1
Cochran-Armitage test for trend			
Good vs. poor vs. dead	≥60 vs. <60-≥0 vs. -5	≤2 vs. >2-≤5 vs. 6	≥3 vs. <3->1 vs. 1
Excellent vs. good vs. poor vs. dead	≥95 vs. <95-≥60 vs. <60-≥0 vs. -5	≤1 vs. <1-≤2 vs. >2-≤5 vs. 6	4 vs. 3 vs. <3->1 vs. 1
Non parametric 2 sample tests			
Wilcoxon test			
Median test			
Robust rank test[4]			
Kolmogorov Smirnov			
Parametric 2 sample tests			
t-test			
Ordinal logistic regression			
Original data			
Good vs. poor vs. dead	≥60 vs. <60-≥0 vs. -5	≤2 vs. >2-≤5 vs. 6	≥3 vs. <3->1 vs. 1
Excellent vs. good vs. poor vs. dead	≥95 vs. <95-≥60 vs. <60-≥0 vs. -5	≤1 vs. <1-≤2 vs. >2-≤5 vs. 6	4 vs. 3 vs. <3->1 vs. 1
Bootstrap difference in mean rank[5]			

References

1. Altman, D.G., *Practical statistics for medical research*. 1991, London: Chapman & Hall.
2. Conover, W.J., *Practical nonparametric statistics*. 2nd ed. 1971, New York: John Wiley & Sons.
3. Siegel, S. and N.J. Castellan, *Nonparametric statistics for the behavioral sciences*. 2 ed. 1988, Singapore: McGraw-Hill. 1-399.
4. Fligner, M.A. and G.E. Policello, *Robust Rank procedures for the Behrens-Fisher problem*. *Journal of the American Statistics Association*, 1981. **76**: p. 162-168.
5. Efron, B. and R.J. Tibshirani, *An introduction to the Bootstrap*. 1 ed. *Monographs on Statistics and Applied Probability*, ed. D.R. Cox, et al. Vol. 57. 1993, New York: Chapman & Hall. 1-436.
6. Stingele, R., E. Bluhmki, and W. Hacke, *Bootstrap statistics of ECASS II data: just another post hoc analysis of a negative stroke trial?* *Cerebrovascular Diseases*, 2001. **11**: p. 30-33.
7. Agresti, A., *Modelling ordered categorical data: Recent advances and future challenges*. *Statistics in Medicine*, 1999. **18**(17-18): p. 2191-2207.

OAST Appendix 2: Supplementary analyses

Supplementary analyses

Sub group analysis

Sub group analyses were performed by assessing the efficiency of the different tests for differing trial characteristics: type of intervention (acute drug treatment, rehabilitation, stroke unit); trial size (<500, ≥500 participants); time between randomisation and stroke onset (≤6, >6 hours); patient age (median ≤70, >70 years); baseline severity (control group death rate adjusted for length of follow up, ≤median (0.05) ,>median); outcome measure (BI, mRS, 3Q); length of follow up (≤3 months, >3 months); and trial result (positive, negative).

Statistical assumptions

The principal statistical assumptions underlying the tests which performed well were assessed to ensure that their use was appropriate for stroke trial data. Assumptions included: ordinal logistic regression - proportionality of odds across response categories (i.e. the magnitude of improvement or hazard, with a treatment, would be similar irrespective of baseline severity, age etc); t-test – normal distribution of outcome scores (the use of the unpooled t-test means that homogeneity of variances between the treatment groups was not a necessary assumption); robust ranks test – independence of treatment groups.[1, 2]

Type 1 error rate

It is conceivable that an overly sensitive statistical test might find significance in a trial when no real difference existed, a type 1 error. We assessed the type I error rate for the three most efficient statistical tests, using data from three representative trials including one of the three measures of functional outcome (BI: RANTTAS,[3] mRS: NINDS,[4] 3Q: IST [5]). From these we generated 1000 data sets, using random sampling with replacement, in which any treatment difference could have occurred only by chance. Tests maintaining adherence to the nominal type I error rate would expect to see a significant result in around 50 of the 1000 data sets.

References

1. Altman, D.G., *Practical statistics for medical research*. 1991, London: Chapman & Hall.

2. Siegel, S. and N.J. Castellan, *Nonparametric statistics for the behavioral sciences*. 2 ed. 1988, Singapore: McGraw-Hill. 1-399.
3. The RANTTAS Investigators, *A randomized trial of tirilazad mesylate in patients with acute stroke (RANTTAS)*. *Stroke*, 1996. **27**: p. 1453-1458.
4. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *Tissue plasminogen activator for acute stroke*. *New England Journal of Medicine*, 1995. **333**: p. 1581-1587.
5. International Stroke Trial Collaborative Group, *The International Stroke Trial (IST); a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke*. *Lancet*, 1997. **349**: p. 1569-1581.

OAST Appendix 3: Trial data

Table 1

Baseline data for included trials

Trial	Sample size	Trial characteristics				Baseline		
		Intervention	Time (hr)	Active groups	Follow up (mo)	Age (median [IQR])	Male (%)	Baseline severity (NIHSS) (median [IQR])
Acute								
AbESTT [1]	400	Abciximab	6	1	3	69 [58-78]	56	9 [6-14]
ASK [2]	340	Streptokinase	4	1	3	71 [64-77]	61	-
ASSIST 07 [3]	138	Selfotel	6	1	3	70 [63-76]	62	14 [8-18]
ASSIST 10 [3]	432	Selfotel	6	1	3	72 [63-77]	55	14 [9-20]
ATLANTIS A [4]	142	Alteplase	0-6	1	3	70 [60-76]	68	11 [7-17]
ATLANTIS B [5]	613	Alteplase	3-5	1	3	67 [59-75]	59	10 [6-15]
BEST pilot [6]	65	Atenolol-propranolol	48	2	6	71 [62-81]	54	-
BEST [6]	302	Atenolol-propranolol	48	2	6	71 [62-77]	52	-
CAST [7]	20,655	Aspirin	48	1	1	-	-	-
Citicoline 1 [8]	259	Citicoline	24	1	3	70 [60-76]	47	11 [7-18]
Citicoline 7 [9]	394	Citicoline	24	1	3	73 [65-79]	47	11 [7-18]
Citicoline 10 [10]	100	Citicoline	24	1	3	74 [62-79]	49	12 [9-16]
Citicoline 18 [11]	899	Citicoline	24	1	3	71 [60-77]	52	13 [10-17]
DCLHb [12]	85	DCLHb	18	1	3	-	-	-
EAST [13]	623	Enlimomab	6	1	3	-	-	-
Ebselen [14]	298	Ebselen	48	1	3	67 [59-74]	63	-
ECASS II [15]	800	Alteplase	6	1	3	68 [59-74]	59	12 [8-16]
Edaravone [16]	250	Edaravone	72	1	3	-	-	-
Factor VIIa [17]	399	Factor VII	3	1	3	-	-	-
FISS [18]	308	Nadroparin	48	2	3	68 [62-73]	58	-
FISS-TRIS [19]	599	Heparin	48	1	6	-	-	-
FOOD 3 [20]	321	NG tube	-	1	6	78 [71-84]	45	-
INWEST [21]	295	Nimodipine	24	2	3	73 [65-79]	46	-
IST [22]	19,435	Aspirin	48	1	6	73 [65-80]	54	-
MAST-E [23]	310	Streptokinase	6	1	6	-	-	-
MAST-I [24]	622	Streptokinase-aspirin	6	3	6	71 [62-78]	54	-
Streptokinase pilot	20	Streptokinase	6	1	3	66 [62-75]	-	-

[25]													
NINDS [26]	624	Alteplase	3	1	3	69 [60-75]	58	14 [9-20]					
PROACT II [27]	180	Prourokinase	6	1	3	-	-	-					
RANTTAS [28]	660	Tirilazad	6	1	3	74 [62-78]	55	9 [5-17]					
RANTTAS II [29]	126	Tirilazad	6	1	3	73 [65-81]	58	13 [9-18]					
STIPAS [30]	111	Tirilazad	12	1	3	68 [58-75]	56	8 [5-16]					
TESS [31]	450	Tirilazad	6	1	3	72 [64-78]	56	-					
TESS II [32]	355	Tirilazad	?	1	3	70 [62-76]	60	-					
Subtotal	51,610			40		72 [64-79]	54	12 [8-17]					
Rehabilitation													
Corr[33]	110	OT	-	1	12	75 [70-81]	37	-					
Gilbertson[34]	138	OT	-	1	6	71 [64-78]	45	-					
Logan[35]	111	OT	-	1	3	72 [66-79]	51	-					
Parker[36]	466	OT	-	2	12	72 [65-79]	58	-					
Walker I[37]	30	OT	-	1	3	70 [62-74]	53	-					
Walker 2[38]	185	OT	1 mo	1	6	75 [70-80]	51	-					
Young[39]	124	PT	-	1	6	70 [65-76]	56	-					
Subtotal	1,164			8		72 [66-79]	52	-					
Stroke unit:													
Dover[40]	235	SU	1-2 w	1	3	74 [67-79]	41	-					
Helsinki[41]	232	SU	-	1	12	-	-	-					
Kuopio[42]	94	SU	1 w	1	3	72 [67-78]	38	-					
Nottingham[43]	315	SU	5 w	1	3	69 [62-75]	59	-					
Orpington[44]	457	SU	-	2	12	-	-	-					
Newcastle[45]	66	SU	72	1	6	77 [73-82]	50	-					
Subtotal	1,399			7		72 [65-78]	49	-					
Total	54,173			55		72 [64-79]	54	12 [8-17]					

IQR: Inter quartile range; NIHSS: National Institute of Health Stroke Scale; hr: hours; w: weeks; mo: months; SU: stroke unit;

OT: occupational therapy; PT: physiotherapy

Table 2

Primary outcome for included trials

	Barthel Index (median [IQR])	Rankin Scale (median [IQR])	3Q (median [IQR])	Death rate (%) per month (control group)	Outcome scale	Type of analysis	Analysis test	Trial result (+/0/-)
Acute								
AbESTT	-	2 [1-4]	-	4.2	mRS	O	Ordinal logistic regression	0
ASK	62.5 [-5-100]	-	-	6.8	BI	D	Chi square test (BI <60)	0
ASSIST 07	70 [15-100]	-	-	5.0	BI	D	Cochran Mantel Haenszel test	0
ASSIST 10	55 [5-100]	-	-	6.1	BI	D	Cochran Mantel Haenszel test	0
ATLANTIS A	90 [20-100]	-	-	2.3	BI	D	Binomial test (BI <100)	0
ATLANTIS B	95 [50-100]	2 [1-4]	-	2.3	mRS	D	Binomial test (mRS >1)	0
BEST pilot	-	-	-	5.8	Nottingham ADL	?	Unpublished	0
BEST	-	-	-	3.8	Nottingham ADL	O	Kruskal-Wallis test	0
CAST	-	-	3 [2-4]	3.9	3Q Scale	D	Chi square test	+
Citicoline 01	80 [20-100]	4 [2-5]	-	5.1	BI	D	Logistic regression (0, 5-40, 45-60, 60-80, 85-100)	+
Citicoline 07	75 [5-100]	2 [1-4]	-	6.4	BI	D	Logistic regression (0, 5-40, 45-60, 60-80, 85-100)	0
Citicoline 10	70 [15-100]	3 [2-4]	-	2.8	mRS	O	Wilcoxon test	0
Citicoline 18	75 [10-100]	3 [1-4]	-	5.9	BI	D	Cochran Mantel Haenszel test	0
DCLHb	-	3 [2-4]	-	3.0	mRS	D	Chi square test (mRS >2)	-
EAST	-	3 [1-5]	-	5.4	mRS	O	Wilcoxon test	-
Ebselen	-	-	-	2.8	BI	O	Wilcoxon test	0
ECASS II	-	2 [1-4]	-	3.4	mRS	D	Fishers exact test (mRS >1)	0
Edaravone	-	2 [1-3]	-	1.5	mRS	O	Wilcoxon test	+

Factor VIIa	-	4 [2-5]	-	9.7	mRS	O	Adjusted cumulative logit model	+
FISS	-	-	2 [2-3]	4.8	3Q Scale	O	Chi square test for trend (dichotomised >2)	+
FISS-TRIS	-	2 [1-3]	-	0.9	mRS	?	Unpublished	0
FOOD 3	-	5 [5-6]	-	8.1	mRS	D	Logistic regression (dichotomised >3)	0
INWEST	25 [-5-85]	-	-	9.3	BI	O	Wilcoxon test	-
IST	-	-	2 [2-3]	3.7	3Q Scale	D	Chi square test	0
MAST-E	-	5 [3-6]	-	6.4	mRS	D	Chi square test (mRS >2)	0
MAST-I	-	4 [1-6]	-	4.8	mRS	D	Chi square test (mRS >2)	0
Streptokinase pilot	48.5 [-5-100]	-	-	10	BI	O	Kruskal-Wallis test	0
NINDS	85 [7.5-95]	3 [1-5]	-	6.8	mRS	D	GEE global outcome (BI <95, RS >1, GOS >1 NIH >1)	+
PROACT II	-	3 [2-6]	-	9.0	mRS	D	Cochran Mantel Haenszel test	+
RANTTAS	95 [25-100]	-	-	4.9	BI	O	Kruskal-Wallis test	0
RANTTAS II	65 [-5-100]	-	-	10.8	BI	O	Kruskal-Wallis test	0
STIPAS	95 [55-100]	-	-	1.2	BI	D	Chi square test (BI <60)	0
TESS	65 [0-100]	-	-	7.2	BI	D	Chi square test	0
TESS II	75 [10-100]	-	-	5.6	BI	?	Unpublished	0
Rehabilitation Corr	55 [15-75]	-	-	1.8	BI	O	Mann Whitney U test	0
Gilbertson	80 [60-90]	3 [2-4]	-	1.2	BI	C	t-test	0
Logan	80 [55-90]	-	-	4.5	BI	O	Wilcoxon test	0
Parker	80 [60-95]	3 [1-4]	-	0.8	mRS	C	Multiple linear regression	0
Walker I	-	-	-	0	Rivermead BI	O	Wilcoxon test	+
Walker II	95 [85-100]	-	-	0	BI	O	Wilcoxon test	+
Young	85 [67.5-100]	-	-	0	BI	O	Mann Whitney U test	+

Stroke unit	95]											
Dover	-	4 [2-6]	-	14.3	mRS	C	Comparison of average score	0				
Helsinki	-	2 [1-5]	-	1.8	mRS	O	Mann Whitney U test	+				
Kuopio	-	-	-	1.6	Trial specific ADL	C	ANCOVA	+				
Nottingham	75 [45-90]	-	-	2.9	BI	O	Wilcoxon test	+				
Orpington	-	2 [1-4]	-	1.3	mRS	D	Chi square test (mRS >*)	+				
Newcastle	30 [15-50]	-	-	6	BI	O	Wilcoxon test	0				

D: Dichotomised or data collapsed into multiple groups; O: Ordinal method; C: Continuous method

+: Beneficial intervention effect; -: Harmful intervention effect; 0: No intervention effect but part of positive or negative meta analysis

Excluded trials

Unavailable data: aptiganel, Day hospital, DIAS, ECASS I, Glycine, Hyperbaric oxygen, LUB-INT-9, Norwegian, Orpington (1993 & 1995), Ronning, Sulter and Goteberg stroke unit trials PROACT I, STAT, ZK200775.

References

1. Abciximab Emergent Stroke Treatment Trial (AbESTT) Investigators, *Emergency treatment of Abciximab for treatment of patients with acute ischemic stroke. Results of a randomized phase 2 trial* Stroke, 2005. 36: p. 880-890.
2. Donnan, G.A., et al., *Streptokinase for acute ischemic stroke with relationship to time of administration: Australian Streptokinase (ASK) Trial Study Group*. Journal of the American Medical Association, 1996. 276: p. 995-996.
3. Davis, S.M., et al., *Selfotel in acute ischemic stroke. Possible neurotoxic effects of an NMDA antagonist*. Stroke, 2000. 31: p. 347-354.
4. Clark, W.M., et al., *The rtPA (alteplase) 0-6 hour acute stroke trial, part A (A0276g). Results of a double-blind, placebo-controlled, multicenter study*. Stroke, 2000. 31: p. 811-816.
5. Clark, W.M., et al., *Recombinant tissue-type plasminogen activator (alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS study: a randomized controlled trial*. Journal of the American Medical Association, 1999. 282: p. 2019-2026.
6. Barer, D.H., et al., *Low dose beta blockade in acute stroke ("BEST" trial): an evaluation*. British Medical Journal, 1988. 296: p. 737-741.
7. CAST (Chinese Acute Stroke Trial) Collaborative Group, *CAST: randomized placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke*. Lancet, 1997. 349: p. 1641-1649.
8. Clark, W.M., S.J. Warach, and L.C. Pettigrew, *A randomised dose-response trial of citicoline in acute ischemic stroke patients*. Neurology, 1977. 49: p. 671-678.
9. Clark, W.M., et al., *A randomised efficacy trial of citicoline in patients with acute ischaemic stroke*. Stroke, 1999. 30: p. 2592-2597.
10. Warach, S., et al., *Effect of citicoline on ischemic lesions as measured by diffusion-weighted magnetic resonance imaging*. Annals of Neurology, 2000. 48(5): p. 713-721.
11. Clark, W.M., et al., *A phase III randomized efficacy trial of 2000mg citicoline in acute ischemic stroke patients*. Neurology, 2001. 57: p. 1595-1602.
12. Saxena, R., et al., *Controlled safety study of a hemoglobin-based oxygen carrier, DCLHb, in acute ischemic stroke*. Stroke, 1999. 30: p. 993-996.
13. Enlimomab Acute Stroke Trial Investigators, *Use of anti-ICAM-1 therapy in ischemic stroke*. Neurology, 2001. 57: p. 1428-1434.
14. Yamaguchi, T., et al., *Ebselen in acute ischemic stroke*. Stroke, 1998. 29: p. 12-17.

15. Hacke, W., et al., *Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ECASS II)*. *Lancet*, 1998. 352: p. 1245-1251.
16. The Edaravone Acute Brain Infarction Study Group (Chair: Eiichi Otomo MD), *Effect of a Novel Free Radical Scavenger, Edaravone (MCI-186), on acute brain infarction*. *Cerebrovascular Diseases*, 2003. 15: p. 222-229.
17. Mayer, S.A., et al., *Recombinant activated factor VII for acute intracerebral hemorrhage*. *New England Journal of Medicine*, 2005. 352(8): p. 777-785.
18. Kay, R., et al., *Low-molecular-weight heparin for the treatment of acute ischemic stroke*. *New England Journal of Medicine*, 1995. 333: p. 1588-1593.
19. Wong, K.S., et al. *A randomized controlled study of low molecular weight heparin versus aspirin for the treatment of acute ischaemic stroke in patients with large artery occlusive disease*. in *14th European Stroke Conference*. 2005. Bologna, Italy.
20. The FOOD Trial Collaboration, *Effect of timing and method of enteral tube feeding for dysphagic stroke patients (FOOD): a multicentre randomised controlled trial*. *Lancet*, 2005. 365: p. 764-772.
21. Wahlgren, N.G., et al., *Intravenous Nimodipine West European Stroke Trial (INWEST) of nimodipine in the treatment of acute ischaemic stroke*. *Cerebrovascular Diseases*, 1994. 4: p. 204-210.
22. International Stroke Trial Collaborative Group, *The International Stroke Trial (IST); a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke*. *Lancet*, 1997. 349: p. 1569-1581.
23. The Multicenter Acute Stroke Trial - Europe Study Group, *Thrombolytic therapy with streptokinase in acute ischemic stroke*. *New England Journal of Medicine*, 1996. 335: p. 145-150.
24. Multicentre Acute Stroke Trial-Italy (MAST-I) Group, *Randomised controlled trial of streptokinase, aspirin, and combination of both in treatment of acute ischaemic stroke*. *Lancet*, 1995. 346: p. 1509-1514.
25. Morris, A.D., et al., *A pilot study of streptokinase for acute cerebral infarction*. *Quarterly Journal of Medicine*, 1995. 88: p. 727-731.
26. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *Tissue plasminogen activator for acute stroke*. *New England Journal of Medicine*, 1995. 333: p. 1581-1587.
27. Furlan, A., et al., *Intra-arterial prourokinase for acute ischemic stroke. The PROACT II study: a randomized trial*. *Journal of the American Medical Association*, 1999. 282: p. 2003-2011.
28. The RANTTAS Investigators, *A randomized trial of tirilazad mesylate in patients with acute stroke (RANTTAS)*. *Stroke*, 1996. 27: p. 1453-1458.
29. Haley, E.C. *High-dose tirilazad for acute stroke (RANTTAS II)*. in *Stroke*. 1998.

30. The STIPAS Investigators, *Safety study of tirilizad mesylate in patients with acute ischemic stroke (STIPAS)*. Stroke., 1994. 25: p. 418-423.
31. Peters, G.R., et al., *Safety and efficacy of 6 mg/kg/day tirilizad mesylate in patients with acute ischemic stroke (TESS study)*. Stroke, 1996. 27: p. 195.
32. Orgogozo, J.M., *TESS II*. Unpublished Work, 1995.
33. Corr, S., *Occupational therapy for stroke patients after hospital discharge - a randomised controlled trial*. Clinical Rehabilitation, 1995. 9: p. 291-296.
34. Gilbertson, L., et al., *Domiciliary occupational therapy for patients with stroke discharged from hospital: randomised controlled trial*. British Medical Journal, 2000. 320(7235): p. 603-606.
35. Logan, P.A., et al., *A randomised controlled trial of enhanced social service occupational therapy for stroke patients*. Clinical Rehabilitation, 1997. 11: p. 107-113.
36. Parker, C.J., et al., *A multicentre randomized controlled trial of leisure therapy and conventional occupational therapy after stroke*. Clinical Rehabilitation, 2001. 15(1): p. 42-52.
37. Walker, M., A. Drummond, and N. Lincoln, *Evaluation of dressing practice for stroke patients after discharge from hospital: a crossover design study*. Clinical Rehabilitation, 1996. 10: p. 25-23.
38. Walker, M., et al., *Occupational therapy for stroke patients not admitted to hospital: a randomised controlled trial*. The Lancet, 1999. 354: p. 278-280.
39. Young, J.B. and A. Forster, *The Bradford community stroke trial: results at six months*. British Medical Journal, 1992. 304: p. 1085-1092.
40. Stevens, R.S. and N.R. Ambler, *The Dover Stroke Rehabilitation Unit: a randomised controlled trial of stroke management*, in *Advances in Stroke Therapy*, F.C. Rose, Editor. 1982, Raven Press: New York. p. 257-261.
41. Kaste, M., H. Palomaki, and S. Sarna, *Where and how should elderly stroke patients be treated? A randomized trial*. Stroke., 1995. 26: p. 249-253.
42. Sivenius, J., et al., *The significance of intensity of rehabilitation of stroke - A controlled trial*. Stroke, 1985. 16(6): p. 928-931.
43. Juby, L.C., et al., *The effect of a stroke rehabilitation unit on functional and psychological outcome: a randomised controlled trial*. Cerebrovascular Diseases, 1996. 6: p. 106-110.
44. Kalra, L., et al., *Alternative strategies for stroke care: a prospective randomised controlled trial*. Lancet, 2000. 356(9233): p. 894-9.
45. Aitken, P.D., et al., *General medical or geriatric unit care for acute stroke? A controlled trial (Abstract)*. Age and Ageing, 1993. 22(supp 2): p. 4-5.

OAST Appendix 4: Results

Table 1

Comparison of the rankings of statistical tests by trial and patient characteristics. The top 3 ranked tests are given for each characteristic

Rank	Scale	Size	Setting	Follow-up	Severity †	Age	Recruitment time	Outcome							
BI	mRS	3Q	<500	Acute	SU	3 M	>3 M	Mild Severe	<70	>70	<6 h	>6 h	-	+	
1	OR	OR	BS	OR	TT	TT	OR	OR	OR	RR	OR	TT	OR	TT	
2	RR	Tr4	TT	RR	Tr3	OR	OR4	TT	BS	OR	TT	OR	RR	OR	
3	BS	TT	RR	BS	RR	X ² c	RR	BS	OR4	RR	TT	BS	Tr4	Tr3	RR
p	**	**	**	**	**	**	*	**	**	**	**	*	**	*	**

BI: Barthel Index; BS: bootstrap; mRS: modified Rankin Scale; 3Q: three questions; OR: ordinal logistic regression (raw data); OR 3: ordinal logistic regression (3 levels); OR 4: ordinal logistic regression (4 levels); RRT: robust ranks test; Tr 3: trend test (3 levels); Tr 4: trend test (4 levels); TT: t-test; X² c: Chi Sq – death or poor outcome vs good.
 *p<0.05; **p<0.01; ***p<0.0001; †Severity assessed as death rate per month of follow up in control group (baseline severity was not used since it was not available for many trials)

OAST Appendix 5: Results

Type 1 error rate

Analysis of 1000 re-sampled random datasets from the 3 trials [1-3] did not find any evidence of an increased type I error rate for ordinal logistic regression with the number of 'positive' data sets being: BI 39/1000 ($p=0.96$); mRS 57/1000 ($p=0.17$) and 3Q 56/1000 ($p=0.21$). Similar results were found for both the t-test and robust ranks test.

Test assumptions

When assessing ordinal logistic regression, the assumption of proportionality of odds (likelihood ratio test comparing the multinomial logistic model to the ordinal logistic regression model) was not met ($p<0.05$) in 8 of the 55 data sets (ASK, $p=0.001$; ASSIST 07, $p=0.002$; ATLANTIS A, $p=0.01$; citicoline 10, $p=0.004$; FOOD 3, $p=0.04$; MAST-I, $p=0.003$; Orpington Domiciliary care, $p=0.02$; Orpington Team, $p=0.02$). The assumption of normality required for the t-test did not hold for any of the data sets. In contrast, the assumption of the robust ranks test was met in all cases whilst the bootstrap approach is assumption free.

References

1. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *Tissue plasminogen activator for acute stroke*. New England Journal of Medicine, 1995. 333: p. 1581-1587.
2. The RANTTAS Investigators, *A randomized trial of tirilazad mesylate in patients with acute stroke (RANTTAS)*. Stroke, 1996. 27: p. 1453-1458.
3. International Stroke Trial Collaborative Group, *The International Stroke Trial (IST); a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke*. Lancet, 1997. 349: p. 1569-1581.