



University  
of Glasgow

Bath, P.M.W., Lees, K.R. , Schellinger, P.D., Altman, H., Bland, M., Hogg, C., Howard, G., and Saver, J.L. (2012) *Statistical analysis of the primary outcome in acute stroke trials*. *Stroke*, 43 (4). pp. 1171-1178. ISSN 0039-2499 (doi:10.1161/STROKEAHA.111.641456)

<http://eprints.gla.ac.uk/63072/>

Deposited on: 30 April 2012

# Stroke

American Stroke  
Association<sup>SM</sup>

JOURNAL OF THE AMERICAN HEART ASSOCIATION

A Division of American  
Heart Association



**Statistical Analysis of the Primary Outcome in Acute Stroke Trials**  
Philip M.W. Bath, Kennedy R. Lees, Peter D. Schellinger, Hernan Altman, Martin  
Bland, Cheryl Hogg, George Howard and Jeffrey L. Saver

*Stroke* published online March 15, 2012

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214  
Copyright © 2012 American Heart Association. All rights reserved. Print ISSN: 0039-2499. Online  
ISSN: 1524-4628

The online version of this article, along with updated information and services, is  
located on the World Wide Web at:

<http://stroke.ahajournals.org/content/early/2012/03/15/STROKEAHA.111.641456>

Subscriptions: Information about subscribing to Stroke is online at  
<http://stroke.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters  
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:  
410-528-8550. E-mail:  
[journalpermissions@lww.com](mailto:journalpermissions@lww.com)

Reprints: Information about reprints can be found online at  
<http://www.lww.com/reprints>

# 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.)

**Key Words:** acute Rx ■ outcomes ■ randomized controlled trials

See related articles, p 935 and p 1163.

The failure of most trials in acute stroke to identify new interventions has no single cause but does reflect, in some cases, the use of suboptimal methods for analyzing the primary outcome. Common outcome scales in acute stroke trials such as the modified Rankin Scale (mRS) and Barthel Index<sup>1</sup> are ordered categorical or pseudocontinuous in nature and yet have been analyzed, in the main, using dichotomous approaches whereby data are converted into binary measures such as “good” or “bad” outcome. Statistical theory,<sup>1a</sup> empirical studies, and mathematical modeling each confirm that functional outcomes measures such as mRS should be analyzed using approaches that maintain the original data so far as possible such as sliding dichotomy, ordinal, or continuous analyses. The following review, which summarizes the findings of a European Stroke Organisation workshop held in February 2011, explores and justifies this position. Acute trials in which the primary outcome is stroke (or vascular) recurrence, for example, Fast Assessment of Stroke and

Transient ischemic attack to prevent Early Recurrence—Pilot Study (FASTER),<sup>2</sup> are not included in this document, although ordinal or continuous approaches may be relevant to these trials as well.<sup>3</sup>

## Background

Historically, most acute stroke trials dichotomized ordinal (eg, mRS) or pseudocontinuous (eg, Barthel Index) functional outcome scales. The resulting binary measures (amounting to “good” and “bad” outcome) were analyzed using the  $\chi^2$  test (thereby providing a  $\chi^2$  statistic and probability value) or logistic regression (which generates an OR and CI). In some cases, the power of these analyses was improved with adjustment for covariates (such as age, sex, and severity).

When considering completed trials, it is apparent that several with positive (alteplase, aspirin) or negative (tirilazad) effects were statistically significant or neutral depending on how they were analyzed. Examples include: (1) European Cooperative Acute Stroke Study (ECASS)-1 (alteplase for

Received April 21, 2011; accepted November 7, 2011.

From the University of Nottingham (P.M.W.B., C.H.), Nottingham, UK; the University of Glasgow (K.R.L.), Glasgow, UK; JW Klinikum Minden (P.D.S.), Minden, Germany; Brainsgate (H.A.), Caesaria, Israel; the University of York (M.B.), York, UK; the University of Alabama at Birmingham (G.H.), Birmingham, AL; and Geffen School of Medicine at the University of California—Los Angeles (J.L.S.), Los Angeles, CA.

†ESO Outcomes Working Group (\*committee member): Hernan Altman, Philip M.W. Bath\* (Session 2 lead), Martin Bland, Natan Bornstein,\* John Boscardin, Stephen M. Davis, Avinoam Dayan,\* Geoffrey Donnan, Wolfgang Eisert,\* Gary A. Ford, Werner Hacke, George Howard, Markku Kaste,\* Michael Krams, Kennedy R. Lees\* (working group chair; Session 1 lead), Didier Leys, Patrik Lyden, David Matchar, Carlos Molina, John Norrie, Bo Norrving, Frank Rathgeb, Joshua Resnick, Steve Richieri, Jeffrey Saver, Peter D. Schellinger\* (Session 3 lead), Ruchir Sehra,\* Yoram Solberg, Danilo Toni,\* Thomas Truelsen, Nils Wahlgren, and Andrew Weiss.\*

Correspondence to Kennedy R. Lees, FRCP, FESO, University of Glasgow, Western Infirmary, 44 Church Street, Glasgow, UK G11 6NT. E-mail k.r.lees@clinmed.gla.ac.uk

© 2012 American Heart Association, Inc.

Stroke is available at <http://stroke.ahajournals.org>

DOI: 10.1161/STROKEAHA.111.641456

acute ischemic stroke, Phase III trial) was neutral when analyzed using a comparison of medians (the primary analysis)<sup>4</sup> but positive with a dichotomous analysis (comparison of mRS 0, 1 versus 2–6) or analysis of a global outcome<sup>5</sup>; (2) ECASS-2 (alteplase for acute ischemic stroke, Phase III trial) was neutral with a dichotomous analysis (comparison of mRS 0, 1 versus 2–6, the primary analysis)<sup>6</sup> but positive with an analysis using a different dichotomization (mRS 0–2 versus 3–6)<sup>6</sup> or with a bootstrap analysis<sup>7</sup>; (3) International Stroke Trial (IST; aspirin for acute ischemic stroke, Phase III megatrial) was neutral with a dichotomous analysis (primary analysis)<sup>8</sup> but positive using a variety of different ordinal analyses (data from<sup>9</sup>); and (4) Safety study of tirilazad mesylate in patients with acute ischemic stroke (STIPAS; tirilazad for acute ischemic stroke, Phase II trial) was neutral with a dichotomous analysis<sup>10</sup> but negative on ordinal analysis (data from the Optimising Analysis of Stroke Trials [OAST] Collaboration<sup>9</sup>).

When choosing a dichotomous analysis, it is clear that different interventions should be analyzed using a transition between “good” and “bad” outcomes that varies according to case mix, for example, mRS 0, 1 versus 2 to 6 or 0 to 2 versus 3 to 6 for early patients with moderate severity (as used for assessment of alteplase)<sup>6,11</sup>; mRS 0 to 3 versus 3 to 6 for dysphagic patients or others with severe stroke (as used for assessment of feeding route)<sup>12</sup>; and mRS 0 to 4 versus 5, 6 for patients with very severe stroke (as used for assessment of hemicraniectomy or factor VIIa).<sup>13,14</sup> If the anticipated case mix is not identified accurately during the design stage, then a suboptimal level for dichotomization may be chosen, as appears to have happened in ECASS-2.<sup>6</sup> Often, the most efficient dichotomous transition will split the trial population into approximately 2 equal halves, that is, around the grand median. In addition to these statistical drivers for choosing a suitable dichotomous transition, the relevance of the transition to patients, caregivers, and society is also important. For example, in a sliding dichotomous outcome based on initial stroke severity, success might be considered moving from mRS 5/6 to  $\leq 4$ ; however, those with a mRS of 4 might still require institutional care in contrast to those with mRS  $\leq 3$ , an outcome with very different perceptions to patient and society.

Improvements to the analysis of dichotomized mRS data have been suggested, including using sliding dichotomy in which individual patient characteristics such as age and severity determine what dichotomous transition should be used for individual patients<sup>15,16</sup>; or the use of a sequential design,<sup>17</sup> repeated measures,<sup>18</sup> or global statistic.<sup>19,20</sup> Finally, statistical theory suggests that analysis of ordered categorical scales such as mRS will be more powerful using ordinal or continuous rather than dichotomous approaches.<sup>1,9</sup>

## Types of Analysis

### Sliding Dichotomy Analysis

The method of sliding dichotomy is also known as responder analysis, prognosis-adjusted analysis, and patient-specific analysis. The method establishes different dichotomous success transitions for prespecified subgroups with distinct prognostic variables such as age and severity.<sup>15,16</sup> This results

**Table 1. Summarizing Treatment Effects Using Binary, Ordinal, and Continuous Differences for 3 Interventions**

mRS	Binary Absolute Risk Reduction	Median Difference	Mean Difference
Aspirin <sup>8</sup>	1.3 (IST* 4, 3 vs 2, 1)	0	0.03
Alteplase <sup>11</sup>	16 (mRS 0, 1 vs 2–6)	1	0.53
Hemicraniectomy <sup>13</sup>	43 (mRS 0–4 vs 5, 6)	2	1.10

mRS indicates modified Rankin Scale; IST, International Stroke Trial.

\*Four-level version of mRS.

in improved statistical power as compared with simple dichotomous analysis.<sup>15</sup> The concept of sliding dichotomy may be generalized to sliding trichotomy or tetrachotomy. Like with a fixed dichotomous analysis, summary statistics may be presented as proportions achieving a poor (or good) outcome and the size of treatment effect described as the difference in proportions having a poor outcome (absolute risk reduction) or as an OR and CI (Table 1).

Other benefits exist for the use of sliding dichotomous analysis. First, prognosis-based transitions significantly reduce the impact of unanticipated case mix, that is, where a trial recruits patients with greater or lesser severity than planned. Second, the  $\chi^2$  test and OR are easy to calculate using the same approaches as those used with a fixed dichotomous approach. This also applies to estimation of number needed to treat (1/absolute risk reduction). However, when compared with ordinal analysis (based on the proportional odds model), sliding dichotomous analysis was less efficient across a number of trials involving participants with traumatic brain injury.<sup>21</sup>

The sliding dichotomous scale needs to be defined before the trial and its calibration will depend on the type of patients who are likely to be recruited, especially their severity. Because there is a significant bias between the National Institutes of Health Stroke Scale scores of left and right strokes for equivalent infarct volumes, the addition of stroke side might also be considered when determining the expected outcome. Other prognostic variables such as infarct size on brain imaging might also be used to determine the cutoffs. Several acute stroke trials have used a sliding dichotomous approach, although they calibrated the dichotomy in differing ways (Table 2).

### Ordinal Analysis

Because the mRS is an ordered categorical outcome, it makes sense to describe and analyze it using ordinal approaches. Summary statistics may be presented as median and interquartile range and the size of treatment effect described as the difference in median mRS scores; for example, alteplase improved the median mRS by an estimated 1 point in the National Institute of Neurological Disorders and Stroke trial (Table 1). Appropriate analytical methods include the Cochran-Armitage trend test,<sup>9,27</sup> Mann-Whitney *U*/Wilcoxon test,<sup>4,9,27</sup> and robust ranks test (a version of the Mann-Whitney *U* test).<sup>9,27</sup> More complex analyses include ordinal logistical regression,<sup>9,28</sup> Cochran-Mantel-Haenszel test (with use of modified Rdit scores),<sup>29,30</sup> and bootstrap analysis.<sup>7,9,31</sup>

**Table 2. Calibration Factors for Acute Stroke Trials Using a Sliding Dichotomous Approach**

Trial	Intervention	Calibration Factors
STICH <sup>22</sup>	Surgery for intracerebral hemorrhage	"Good" vs "bad" outcome dependent on prognostic factor based on age, Glasgow Coma Scale, and hemorrhage lesion volume
AbESTT-II <sup>15,23,24</sup>	Abciximab for acute ischemic stroke	"Good" outcome if final mRS was 0 when starting with a baseline NIHSS of 4–7; mRS 0–1 if baseline NIHSS 8–14; and mRS 0–2 if NIHSS 15–22
PAIS <sup>25,26</sup>	Paracetamol for acute ischemic stroke	Improvement beyond expectation was defined as a score on the mRS lower than the median grade of patients with a similar prognostic index estimated from age, sex, NIHSS, previous stroke, stroke type (hemorrhagic, ischemic), and diabetes mellitus

STICH indicates Surgical Trial in Intracerebral Haemorrhage; AbESTT, Abciximab in Emergency Treatment of Stroke Trial; PAIS, Paracetamol (Acetaminophen) In Stroke; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale.

Ordinal analyses tend to be more efficient statistically than dichotomous approaches; they result in a smaller  $\alpha$  ( $P$ ) value for a fixed sample size and power ( $1-\beta$ ) or smaller sample size for fixed  $\alpha$  and  $\beta$ . This has been shown using data modeling (with analysis of randomly generated data, data shaped to reflect projected treatment effects, or data from completed neutral trials with an artificial treatment effect added<sup>15,17,28</sup>) and empirical analyses using individual patient data from positive or negative (but not neutral) treatments.<sup>9,18,32–34</sup> The benefit arises because there are more transitions for an individual to move across with an ordinal as compared with a dichotomous outcome. Binary outcomes, by focusing only on a single health state transition, ignore important effects of treatments occurring at other health state transitions that are highly valued by patients, physicians, and society. When each step change in an ordinal measure is clinically relevant, ordinal outcomes much better reflect a patient-centered perspective on health status.

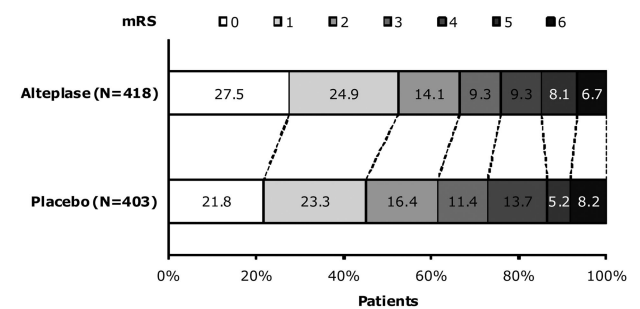
Other advantages exist for ordinal analysis. First, they combine safety and efficacy in a single analysis, whereas it is possible to have a single transition in a dichotomized analysis that demonstrates benefit at 1 end of the outcome distribution at the same time as disregarding harm at the other end. Second, ordinal analyses are more interpretable than using the global statistic.<sup>19,20</sup> The global statistic (Wald test), as derived from a general linear model with logit-link function, simultaneously tests for effect in multiple dichotomous outcomes; as such, the power of the global test is equal or greater than that of any of the single dichotomous outcomes.<sup>35</sup> With ordinal analysis, number needed to treat (NNT) values can be derived for capturing how many patients improve or worsen by  $\geq 1$  steps on the outcome scale; in contrast, the global statistic test reflects change for a latent trait that is not fully captured or articulated by any explicit scale. Last, ordinal analyses are recommended for other neurological diseases as assessed in the International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) study for traumatic brain injury with analysis of Glasgow Outcome Scale data.<sup>21,36–38</sup>

For comparisons between 2 treatment groups, the Mann Whitney  $U$  test, adjusted for ties, provides a suitable test of significance. The Kruskal-Wallis test generalizes this to  $>2$  groups. The only assumption required is that the data can be ordered. The Cochran-Armitage  $\chi^2$  test for trend provides an alternative, which may be more powerful, but only at the cost of a very strong assumption about the spacing of mRS

categories. Neither approach offers an estimate of the size of the effect or enables adjustment for covariates.

The usual way to produce an estimate and to allow for covariates is to use ordered logistic regression. Ordinal logistic regression assumes that the treatment effect is consistent across the whole mRS spectrum, that is, that the OR of being below a given mRS score is the same for each point of the mRS scale. It is this OR that provides the treatment estimate. This may not be the case for some treatments. A formal test of this "proportionality" hypothesis should be made (eg, using the likelihood ratio test for a random effects model or score test for a fixed effects model), although these tests may be oversensitive. Nevertheless, when examining individual data from previous acute trials, most (85%) data sets did not depart significantly from the proportionality assumption.<sup>9</sup> (Similarly, ordinalization of vascular prevention trial data showed that most trials adhered to proportionality of odds.<sup>3</sup>)

The use of ordinal approaches may be inappropriate for some interventions, that is, they may reduce statistical power. In the first scenario, the treatment benefit is similar across the scale (symmetry) but a hazard is only present at 1 end of the scale (asymmetry).<sup>32</sup> For example, alteplase improves outcome independent of severity but can increase intracerebral hemorrhage, especially in those with severe stroke.<sup>32</sup> In this example, if the risk of hazard (hemorrhage) is relatively low, then a shift analysis may still be efficient, as reported in ECASS-3 (Figure 1).<sup>39</sup> Alternatively, the intervention's benefits may be clustered at a single or only limited number of health state transitions; potential examples are early recanalization and hemicraniectomy, which, respectively, move many patients to an excellent or only fair outcome across a wide range of starting severity.<sup>18,40,41</sup> In these situations,



**Figure 1.** Distribution of modified Rankin Scale in the ECASS-3 study; note the shift to a good outcome with alteplase.<sup>39</sup> ECASS indicates European Cooperative Acute Stroke Study.

dichotomous approaches centered on the highly informative transition are more efficient than ordinal analysis.

### Continuous Analysis

Multilevel scales such as the 7-level mRS may be considered to be a continuous variable so that parametric descriptors and analyses can be used. Summary statistics may be presented as mean and SD and the size of treatment effect described as the difference in mean mRS scores, for example, alteplase improved the mRS by an average of 0.53 points in the National Institute of Neurological Disorders and Stroke trial (Table 1). Appropriate analytical methods include: (1) the *t* test<sup>9</sup>; (2) linear regression; and (3) analysis of variance.

Parametric analyses require that the sample size is large enough for the rules of normal approximation to be assessed, in particular that the sample size is sufficient so that the sample statistics have a normal distribution. This condition may apply for trials involving hundreds (or more) of participants, as occurs in modern Phase III acute stroke trials. However, the mRS exhibits skewness to an extent that depends on patient severity, and there are no recommendations on how to normalize it through transformation; mRS data also exhibit kurtosis, which makes it nonnormal. When compared with ordinal and dichotomous analyses of existing acute stroke trial data, continuous analyses (based on the *t* test) had comparable power to ordinal analyses.<sup>9</sup> However, the statistical assumptions underlying the *t* test were not met for most of the acute data sets.<sup>9</sup> As a result, statisticians differ on whether it is appropriate to use parametric approaches involving ordered categorical data such as the mRS, even when the data set is large.

### Which Is the Best Analytic Approach and How Should the Results Be Reported?

It is improbable that there is a “most efficient” analytic method that will work in all situations, but 3 studies have compared several statistical approaches that suggest that some methods are likely to be better than others. Using individual patient data from acute stroke trials, the OAST Collaboration compared several approaches and found that their ordering (“best” first) was: ordinal logistic regression, *t* test, robust ranks test, bootstrap, Mann-Whitney *U*/Wilcoxon test, Cochran-Armitage ranks test,  $2 \times 2$   $\chi^2$  test, and median test.<sup>9</sup> Hence, ordinal and continuous approaches were superior to fixed dichotomous analyses (Figure 2). Another analysis of multiple acute stroke trials determined that ordinal approaches were more efficient than dichotomous when treatment effects were modest and distributed over the entire outcome range or had a profile that could not be confidently predicted before study completion, whereas dichotomous approaches were more efficient than ordinal when treatment effects clustered at single-state transitions that could be prespecified.<sup>40</sup> The IMPACT Collaboration reported that ordinal approaches were more efficient than a sliding dichotomous method in a range of trials involving participants with traumatic brain injury.<sup>21</sup>

The results of dichotomous analyses are easy to explain by presenting the absolute risk reduction in the outcome between the treatment groups. For example, alteplase reduces death

Test	Mean	No. of	Banding
	Rank	Datasets	
Ordinal logistic regression	6.11	54	
<i>t</i> 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	
$\chi^2$ – death or poor outcome vs good	8.87	55	
$\chi^2$ – death or poor outcome vs excellent	9.24	54	
Median test	9.47	55	
$\chi^2$ – 2x3 test	9.96	51	
$\chi^2$ – death vs alive	9.98	51	
$\chi^2$ – 2x4 test	10.02	50	
Kolmogorov-Smirnov test	11.29	55	

**Figure 2.** Comparison of rank scores for 16 statistical tests; lower ranks imply the test is more efficient. Analysis by 2-way ANOVA and Duncan multiple comparison procedure; tests joined by the same vertical band are not statistically different from each other at  $P=0.05$ .<sup>9</sup> ANOVA indicates analysis of variance; RRT, robust ranks test. Comparison of rank scores for 16 statistical tests.

and dependency (mRS 2–6) by an absolute 16% from 74% to 57% (Table 1).<sup>11</sup>

Depending on the type of ordinal analysis, the effect of treatment may be described in 2 ways. First, the OR (and CI) for the reduction of death or dependency may be reported. An OR of 0.83 (95% CI, 0.70–0.99, positive) was found in the Stroke-Acute Ischemic NXY Treatment (SAINT)-1 trial of NXY-059 (analysis by Cochran-Mantel-Haenszel test).<sup>29</sup> Similarly, the Scandinavian Candesartan Acute Stroke Trial (SCAST) trial reported an OR of 1.17 (95% CI, 1.00–1.38, negative trend) for candesartan (analyzed using ordinal logistic regression).<sup>28</sup> Second, the reduction of death or dependency may be described by the difference in medians, for example, alteplase reduced the mRS by a median 1 point in the National Institute of Neurological Disorders and Stroke study of alteplase (analysis by Mann-Whitney *U* test or robust ranks test; Table 1).<sup>9</sup>

When performing a continuous analysis, the reduction of death or dependency may be described by the difference in means, for example, alteplase reduced the mRS by a mean 0.53 mRS points in the National Institute of Neurological Disorders and Stroke study of alteplase (analysis by *t* test; Table 1).<sup>9</sup> A similar approach also applies to the reporting of analyses based on multiple regression analysis.

Alternatively, the likelihood of having a better or worse outcome may be described using NNTs.<sup>33,42</sup>

Practically, it is most important to identify an efficient statistical approach that increases the chance of obtaining a positive (or negative rather than neutral) trial for an intervention. Reporting the results will then follow using the appropriate point estimate (absolute risk reduction, common odds ratio, mean difference) with CI. Secondary analyses of the primary outcome should also be given to show internal

consistency, for example, a fixed dichotomous analysis could follow the primary ordinal analysis. An explanation of the meaning of the results, in particular its clinical relevance, can then follow, perhaps using the NNT.

## Other Issues in Analysis

### Covariate Adjustment

Adjustment for predictive baseline covariates improves the power of trials or reduces their sample size. The principal explanation is that adjustment with covariates that are predictive of the outcome will reduce the variability in the outcome and thus improve the estimate of the treatment effect. The improvement may be apparent as a change in the point estimate, a narrowing of the CIs, or both. An example of this is the false-neutral result for aspirin in the IST megatrial, which disappeared after covariate adjustment.<sup>8</sup>

A lesser reason for covariate adjustment is to address any imbalances in predictive factors between the treatment groups. In stroke trials, the impact of covariates such as age and severity on outcome is typically much larger than the treatment effect that is being measured. If no adjustment is performed, chance imbalances in the population across the treatment arms can result in either false-positive/negative (Type I error) or false-neutral (Type II error) findings. An example is the trend to an increase in death in the Postural Assessment Scale for Stroke Patients (PASS) trial of piracetam, which was related to an imbalance in severity at baseline.<sup>43</sup>

Like with binary analyses, shift analyses are more powerful statistically if adjusted for baseline prognostic covariates.<sup>33</sup> The key covariates in acute stroke trials are the 2 most powerful prognostic factors: severity (eg, baseline National Institutes of Health Stroke Scale) and age; it may also be relevant to add sex as a covariate in view of its central biological role.<sup>33</sup> Due to the bias in the National Institutes of Health Stroke Scale measure of severity between stroke sides, an interaction term, "National Institutes of Health Stroke Scale $\times$ side of lesion," may also be included as a covariate. Alternatively, infarct size (obtained from measurement on the brain neuroimaging scan) could replace severity. Covariate adjustment is also recommended for other neurological diseases, for example, traumatic brain injury, as assessed in the IMPACT study.<sup>21,36–38</sup>

### Sample Size Calculations

Standard methods exist for the calculation of sample size required for a given significance and power.<sup>44</sup> These are well described for both dichotomous and continuous analyses. The sample size for a trial designed to use a sliding dichotomy will need to identify the prognostic bands, the proportion of patients in each, and summate the sample sizes using a binary calculation for each band.

Trials designed to use an ordinal or continuous analysis on their primary outcome (mRS) may use 1 of 2 approaches for estimating the necessary sample size. First, they can use a specific method designed for ordered categorical data, for example, the comparison of medians model of Payne,<sup>45</sup> Mann-Whitney *U* model of Noether,<sup>46</sup> or proportional odds model of Whitehead.<sup>47</sup> The comparison of means model is

**Table 3. Calculation of Total Trial Sample Size for Fixed Dichotomous, Ordinal, and Continuous Analyses for a Potential Neuroprotective Intervention Assuming a Population of Patients With Moderate to Severe Stroke (mRS [0–6] Proportions in the Control Group of 7.8%, 11.7%, 14.1%, 20.8%, 18.9%, 7.6%, and 19.1%), Alpha=0.05, Power=0.90, and Randomization 1:1**

mRS	Test to Be Used	Parameter	Total Sample Size
Fixed dichotomous			
0–1/2–6	Logistic regression	ARR=5.0%	2882
0–2/3–6	Logistic regression	ARR=6.8%	2116
0–3/4–6	Logistic regression	ARR=7.2%	1972
0–4/5–6	Logistic regression	ARR=5.6%	2448
Ordinal			
Whitehead <sup>47</sup>	Ordinal logistic regression	Common odds=0.75	1566
Noether <sup>46</sup>	Mann-Whitney <i>U</i>	Delta=0.547	1533
Continuous			
Linear	<i>t</i> test	Difference=0.307	1536

mRS indicates modified Rankin Scale; ARR, absolute risk reduction.

appropriate for continuous data. When comparing some of these approaches using parameters from completed trials, Whitehead's approach resulted in the smallest sample size.<sup>32</sup> Walters provides guidance on choosing the optimal sample size calculation for differing scenarios when using ordinal outcomes.<sup>48</sup>

Alternatively, sample size may be calculated as though the mRS will be analyzed using fixed dichotomization, that is, by comparing 2 proportions. The sample size can then be reduced by approximately 20%, close to the lower CI for the benefit of ordinal versus binary analyses in the OAST Collaboration.<sup>32</sup> This approach has been recommended by some statisticians because it is likely to be conservative and may be easier to explain to investigators and trial reviewers.

The advantage of ordinal or continuous over fixed dichotomous analysis is seen in an example related to a hypothetical neuroprotective intervention in which the proportional odds model is likely to be relevant, that is, the intervention will shift participants throughout the range of the mRS. Table 3 shows sample size calculations for dichotomous, ordinal, and continuous models. The importance of choosing the correct mRS transition for the fixed dichotomous is apparent with sample sizes varying between 1972 and 2882. The ordinal and continuous calculations give sample sizes between 1500 and 1600.

The resulting sample size needs to be increased somewhat to deal with participants who are lost to follow-up or crossover treatment groups and reduced significantly if adaptive randomization (minimization), covariate adjustment, and/or population enrichment are used.<sup>33,49,50</sup> Taking account of these factors, the example trial might only need to enrol approximately 1200 participants.

### Numbers Needed to Treat

The NNT is a measure of an intervention's efficacy and, for a dichotomous outcome, is equivalent to the reciprocal of the absolute risk reduction between the intervention groups. NNT

**Table 4. No. of Patients Who Benefit From 3 Interventions**

	Patients per Annum (UK; a)	Eligibility, % (b)	Patients Receiving Treatment (c)	Cost per patient, £ (d)	Total Cost, £000 (e)	NNT (Ordinal) <sup>42</sup> (f)	Patients Who Benefit (g)
Aspirin	110 000	80	88 000	1	88	42	2095
Alteplase	110 000	15	16 500	650	10 725	4.5	3667
Hemicraniectomy	110 000	0.1	110	7000	770	2.4	46

NNT indicates number needed to treat.

$c = a \times b$ ;  $e = c \times d$ ;  $g = c/f$ .

may be estimated for ordinal and continuous outcomes using expert generation of joint distribution outcome tables<sup>51,52</sup> or derived mathematically from trial data.<sup>42,53,54</sup> NNTs are smaller with ordinal versus fixed dichotomous approaches, as seen for alteplase in which an NNT of approximately 3 to 4 was present when analyzed using all 7 levels of the mRS versus approximately 8 when assessed for mRS 2 to 6 in a fixed dichotomous analysis.<sup>42,51</sup>

### Minimally Important Difference

The minimally important difference is the difference in outcomes between treatment groups that is declared to be clinically important. The minimally important difference may be defined statistically, for example, as “half a SD,”<sup>55</sup> or on the basis of the outcome under test (mRS in the present discussion) and clinical factors that will depend, in part, on the outcome frequency, safety, availability, and cost of the intervention. So, an expensive intervention will need to demonstrate a larger treatment effect (ie, larger minimally important difference) than an inexpensive intervention. Econometric studies have indicated that effects as low as 2% of the patients moving from mRS 5 to 2 would save significant costs.<sup>56</sup> Although suitable minimally important differences have not been clearly established for stroke, treatment effects are available for existing effective interventions (Table 1) that may guide the definition of minimally important differences for future trials. The variation in absolute risk reduction (1.3%–43%) and mean difference in mRS (0.03–1.10) illustrate clearly that minimally important differences will vary substantially for different interventions and populations of patients. The Stroke Therapy Academic Industry Roundtable has suggested that for neuroprotective therapies, absolute effect sizes of 2% to 8% (dichotomous) would be acceptable.<sup>57</sup>

### What Magnitude of Effect Is Relevant to Clinicians, Regulators, and Healthcare Funders?

This is a function of the point estimate for treatment effect (encompassing both efficacy and safety), availability, and cost of the intervention and includes health economic assessment. The key drivers per patient are the NNT (used as a measure of efficacy) and cost of the intervention; the total healthcare costs will reflect unit costs and the proportion of patients who can receive an intervention among the total stroke population (Table 4).

Therefore, if alteplase had its current eligibility and cost but the NNT of aspirin, it would not be used, being considered too expensive for modest efficacy. Conversely, aspirin

with a NNT of approximately 42 would be irrelevant if it had the cost of alteplase.

The absolute health value of each of the levels of the mRS has been derived using World Health Organization’s Global Burden of Disease project disability weight system. This approach allows calculation of disability-adjusted life-years gained per health dollar expended.<sup>58</sup> A similar approach has been used to relate mRS levels to quality-adjusted life-years.<sup>59</sup>

### What Analysis Is Important to Patients?

It is important to consider the implications of improving the analysis of trials on patients and their understanding of trials. Clearly, a trial reporting a fixed dichotomous outcome such as alteplase reduces the rate of death or dependency by “X”%, as seen in the National Institute of Neurological Disorders and Stroke study and ECASS-3, is easy to understand. The same is true for trials based on sliding dichotomous outcomes. Trials using ordinal or continuous approaches can still report median or mean differences in outcome so patients can be told that the average outcome is better by “X.Y” points in the mRS with a new intervention. Most patients (and many healthcare staff) do not understand the concept of ORs, but this concern applies whether logistic (binary) or ordinal logistic regression is used. In reality, most people, whether patients or healthcare staff, would rather know that a stroke trial was optimally designed so the chance of throwing away an effective treatment was minimized than worry how the results were going to be presented.

### Conclusions

The use of fixed dichotomous analysis of ordered categorical outcomes after stroke (such as the mRS) can rarely be recommended because it is statistically inefficient and 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,<sup>44,56</sup> 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.



## Acknowledgments

The workshop was sponsored by the European Stroke Organisation (ESO) with administrative assistance supplied by Marisa Kretschki on behalf of the ESO.

## Disclosures

The authors and members of the working group are employed by a range of academic and commercial organizations involved in stroke research, including pharmaceuticals and devices. The article was drafted by P.M.W.B. with assistance of K.R.L., P.D.S., and coauthors and approved by all members of the working group. No commercial organization was involved in the decision to publish.

## References

- Lees KR, Bath PMW, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke*. 2012;43:1163–1170
- Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat*. 2009;8:50–61.
- Kennedy J, Hill MD, Ryckborst K, Eliasziw M, Demchuk AM, Buchan AM, for the FASTER Investigators. Fast Assessment of Stroke and Transient ischaemic attack to prevent Early Recurrence (FASTER): a randomised controlled pilot trial. *Lancet Neurology*. 2007;6:961–969.
- Bath PMW, Geeganage CM, Gray LJ, Collier T, Pocock S. Use of ordinal outcomes in vascular prevention trials: comparison with binary outcomes in published stroke trials. *Stroke*. 2008;39:2817–2823.
- Hacke W, Kaste M, Fieschi C, Toni D, Lesaffre E, von Kummer R, Boysen G, Bluhmki E, Hoxter G, Mahagne M-H, Hennerici M, Group. ES. Intravenous thrombolysis with tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Trial (ECASS). *JAMA*. 1995;274:1017–1025.
- Hacke W, Bluhmki E, Steiner T, Tatlisumak T, Mahagne M-H, Sacchetti M-L, Meier D. Dichotomized efficacy end points and global end-point analysis applied to the ECASS intention to treat data set. Post hoc analysis of ECASS 1. *Stroke*. 1998;29:2073–2075.
- Hacke W, Markku K, Fieschi C, von Kummer R, Davalos A, Meier D, Larue 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.
- Stingele R, Bluhmki E, Hacke W. Bootstrap statistics of ECASS II data: just another post hoc analysis of a negative stroke trial? *Cerebrovasc Dis*. 2001;11:30–33.
- International Stroke Trial Collaborative Group. The International Stroke Trial (IST); a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *Lancet*. 1997;349:1569–1581.
- The Optimising Analysis of Stroke Trials (OAST) Collaboration. Can we improve the statistical analysis of stroke trials? Statistical re-analysis of functional outcomes in stroke trials. *Stroke*. 2007;38:1911–1915.
- The STIPAS Investigators. Safety study of tirilazad mesylate in patients with acute ischemic stroke (STIPAS). *Stroke*. 1994;25:418–423.
- The National Institute of Neurological Disorders and Stroke rtPA Stroke Study Group. Tissue plasminogen activator for acute stroke. *N Engl J Med*. 1995;333:1581–1587.
- 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;365:764–772.
- Vahedi K, Hofmeijer J, Vacaute E, George B, Algra A, Amelink GJ, Schmiedeck P, Schwab S, Rothwell PM, Bousser MG, van der Worp HB, Hacke W, for the DECIMAL, DESTINY, and HAMLET investigators. Early decompressive surgery in malignant infarction of the middle cerebral artery: a pooled analysis of three randomised controlled trials. *Lancet Neurol*. 2007;6:215–222.
- Mayer SA, Brun NC, Begtrup K, Broderick J, Davis S, Diringer MN, Skolnick BE, Steiner T, FAST Trial Investigators. Efficacy and safety of recombinant activated factor VII for acute intracerebral hemorrhage. *N Engl J Med*. 2008;358:2127–2137.
- Young FB, Lees KR, Weir CJ, for the GAIN International Steering Committee and Investigators. Improving trial power through use of prognosis-adjusted end points. *Stroke*. 2005;36:597–601.
- Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AIR, Marmarou A, Mendelow AD, Steyerberg EW, Taylor GS, Teasdale GM, Weir CJ. Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy. *J Neurotrauma*. 2005;22:511–517.
- Bolland K, Weeks A, Whitehead J, Lees KR, GAIN International Steering Committee and the Investigators. How a sequential design would have affected the GAIN international study of gavestinel in stroke. *Cerebrovasc Dis*. 2004;17:111–117.
- Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rtPA stroke trial. *Stat Med*. 2009;29:546–557.
- Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, Lyden P, Grotta J. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke tPA stroke trial. *Stroke*. 1996;27:2136–2142.
- Saver JL. Novel-end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. *Stroke*. 2007;38:3055–3062.
- McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT project. *Clin Trials*. 2010;7:44–57.
- Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope TD, Karimi A, Shaw MDM, Barer DH, for the STICH investigators. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the international Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *Lancet*. 2005;365:387–397.
- Adams HP, Leclerc JR, Bluhmki E, Clarke WR, Hansen MD, Hacke W. Measuring outcomes as a function of baseline severity of ischemic stroke. *Cerebrovasc Dis*. 2004;18:124–129.
- Adams HP, Effron MB, Torner J, D'Avalos A, Frayne J, Teal P, Leclerc J, Oemar B, Padgett L, Barnathan ES, Hacke W, for the AbESTT-II investigators. Emergency administration of abciximab for the treatment of patients with acute ischemic stroke: results of an international phase III trial. Abciximab in Emergency Treatment of Stroke Trial (ABESTT-II). *Stroke*. 2008;39:87–99.
- den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, Koudstaal PJ, Dippel DW. The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial. *Lancet Neurol*. 2009;8:434–440.
- den Hertog HM, van der Worp HB, van Gemert HMA, Algra A, Japp KL, van Gijn J, Koudstaal PJ, Dippel DWJ, for the PAIS investigators. Paracetamol (acetaminophen) in stroke; protocol for a randomised, double blind clinical trial. *BMC Cardiovasc Disord*. 2008;8:29–33.
- Siegel S, Castellan NJ. *Nonparametric Statistics for the Behavioral Sciences*. Singapore: McGraw-Hill International Editions; 1988.
- Sandset EC, Bath PM, Boysen G, Jatuzis D, Korv J, Luders S, Murray GD, Richter PS, Roine RO, Terent A, Thijs V, Berge E, on behalf of the SCAST Study Group. The angiotensin-receptor blocker candesartan for treatment of acute stroke (SCAST): a randomised, placebo-controlled double-blind trial. *Lancet*. 2011;377:741–750.
- Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener H-C, Grotta J, Lyden P, Shuaib A, Hardemark H-G, Wasiewski WW, for the Stroke-Acute Ischemic NXY Treatment (SAINT I) Trial Investigators. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006;354:588–599.
- Shuaib A, Lees KR, Lyden P, Grotta J, Davalos A, Davis SMD, H-C., Ashwood T, Wasiewski WW, Emeribe U, for the SAINT II Trial Investigators. NXY-059 for the treatment of acute ischemic stroke. *N Engl J Med*. 2007;357:562–571.
- Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.
- The Optimising Analysis of Stroke Trials (OAST) Collaboration. Calculation of sample size for stroke trials assessing functional outcome: comparison of binary and ordinal approaches. *Int J Stroke*. 2008;3:78–84.
- The Optimising Analysis of Stroke Trials (OAST) Collaboration. Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke*. 2009;40:888–894.
- Savitz SI, Lew R, Bluhmki E, Hacke W, Fisher M. Shift analysis versus dichotomization of the modified Rankin Scale outcome scores in the NINDS and ECASS-II trials. *Stroke*. 2007;38:3205–3212.
- Pocock S, Geller N, Tsiatis A. The analysis of multiple endpoints in clinical trials. *Biometrics*. 1987;43:487–498.

36. Roozenbeek B, Maas AI, Lingsma HF, Butcher I, Lu J, Marmarou A, McHugh GS, Weir J, Murray GD, Steyerberg EW. Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? *Crit Care Med*. 2009;37:2683–2690.
37. Roozenbeek B, Maas AI, Marmarou A, Butcher I, Lingsma HF, Lu J, McHugh GS, Murray GD, Steyerberg EW. The influence of enrollment criteria on recruitment and outcome distribution in traumatic brain injury studies: results from the IMPACT study. *J Neurotrauma*. 2009;26:1069–1075.
38. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, Lu J, Weir J, Roozenbeek B, Murray GD. IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. *Neurotherapeutics*. 2010;7:127–134.
39. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, Larrue V, Lees KR, Medeghri Z, Machnig T, Schneider D, on Kummer R, Wahlgren N, Toni D, ECASS Investigators. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med*. 2008;359:1317–1329.
40. Saver J, Gornbein J. Treatment effects for which shift or binary analyses are advantageous in acute stroke trials. *Neurology*. 2008;72:1310–1315.
41. Guan X LD, Martin N, Gonzalez N, Vespa P, Saver JL. Number needed to treat to benefit and to harm estimates for early decompressive craniectomy in malignant middle cerebral artery infarction *Stroke*. In press.
42. Optimising the Analysis of Stroke Trials (OAST) Collaboration. Calculation of numbers-needed-to-treat (NNT) in parallel group trials assessing ordinal outcomes: case examples from acute stroke and stroke prevention. *Int J Stroke*. 2011;6:472–479.
43. De Deyn P, De Reuck J, Deberdt W, Vlietinck R, Orgogozo JM. Treatment of acute ischemic stroke with piracetam. *Stroke*. 1997;28:2347–2352.
44. 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.
45. Payne R. *Genstat Reference Manual*. Hemel Hempstead, UK; VSN International; 1993.
46. Noether G. Sample size determination for some common nonparametric tests. *J Am Stat Assoc*. 1987;82, No 398:645–647.
47. Whitehead J. Sample-size calculations for ordered categorical-data. *Stat Med*. 1993;12:2257–2271.
48. Walters SJ. Sample size and power estimation for studies with health related quality of life outcomes: a comparison of four methods using the SF-36. *Health Qual Life Outcomes*. 2004;2:26–41.
49. Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Stat Med*. 2003;22:705–726.
50. Weir CJ, Kaste M, Lees KR, for the Glycine Antagonist in Neuroprotection (GAIN) International Trial Steering Committee and Investigators. Targeting neuroprotection clinical trials to ischemic stroke patients with potential to benefit from therapy. *Stroke*. 2004;35:2111–2116.
51. Saver JL. Number needed to treat estimates incorporating effects over the entire range of clinical outcomes. *Arch Neurol*. 2004;61:1066–1070.
52. Saver JL. Clinical impact of NXY-059 demonstrated in the SAINT I trial: derivation of number needed to treat for benefit over entire range of functional disability. *Stroke*. 2007;38:1515–1518.
53. Walter SD. Number needed to treat (NNT): estimation of a measure of clinical benefit. *Stat Med*. 2001;20:3947–3962.
54. Lansberg MG, Schrooten M, Bluhmki E, Thijs VN, Saver JL. Treatment time-specific number needed to treat estimates for tissue plasminogen activator therapy in acute stroke based on shifts over the entire range of the modified Rankin Scale. *Stroke*. 2009;40:2079–2084.
55. Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41:582–592.
56. Samsa GP, Matchar DB. Have randomized controlled trials of neuroprotective drugs been underpowered? An illustration of three statistical principles. *Stroke*. 2001;32:669–674.
57. Fisher M, for the Stroke Therapy Academic Industry Roundtable IV. Enhancing the development and approval of acute stroke therapies. *Stroke*. 2005;36:1808–1811.
58. Hong KS, Saver JL. Quantifying the value of stroke disability outcomes: WHO global burden of disease project disability weights for each level of the modified Rankin Scale. *Stroke*. 2009;40:3828–3833.
59. Samsa GP, Reutter RA, Parmigiani G, Ancukiewicz M, Abrahamse P, Lipscomb J, Matchar DB. Performing cost-effectiveness analysis by integrating randomized trial data with a comprehensive decision model: application to treatment of acute ischemic stroke. *J Clin Epidemiol*. 1999;52:259–271.


 Heart Association | Stroke Association

# Stroke

JOURNAL OF THE AMERICAN HEART ASSOCIATION