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

Deposited on: 30 April 2012
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
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 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, 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), are not included in this document, although ordinal or continuous approaches may be relevant to these trials as well.3

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...
acute ischemic stroke, Phase III trial) was neutral when analyzed using a comparison of medians (the primary analysis) but positive with a dichotomous analysis (comparison of mRS 0, 1 versus 2–6) or analysis of a global outcome; (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) but positive with an analysis using a different dichotomization (mRS 0–2 versus 3–6) or with a bootstrap analysis; (3) International Stroke Trial (IST; aspirin for acute ischemic stroke, Phase III megatrial) was neutral with a dichotomous analysis (primary analysis) but positive using a variety of different ordinal analyses (data from IST); 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 but negative on ordinal analysis (data from the Optimising Analysis of Stroke Trials [OAST] Collaboration).

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); mRS 0 to 3 versus 3 to 6 for dysphagic patients or others with severe stroke (as used for assessment of feeding route); and mRS 0 to 4 versus 5, 6 for patients with very severe stroke (as used for assessment of hemicraniectomy or factor VIIa). 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. 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 ≤4; however, those with a mRS of 4 might still require institutional care in contrast to those with mRS ≤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; or the use of a sequential design, repeated measures, or global statistic. 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.

### 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. This results in improved statistical power as compared with simple dichotomous analysis. 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 χ² 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.

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, Mann-Whitney *U* test, and robust ranks test (a version of the Mann-Whitney *U* test). More complex analyses include ordinal logistical regression, Cochran-Mantel-Haenszel test (with use of modified Ridit scores), and bootstrap analysis.

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

<table>
<thead>
<tr>
<th>mRS</th>
<th>Binary Absolute Risk Reduction</th>
<th>Median Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>1.3 (IST, 4, 3 vs 2, 1)</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Alteplase</td>
<td>16 (mRS 0, 1 vs 2–6)</td>
<td>1</td>
<td>0.53</td>
</tr>
<tr>
<td>Hemicraniectomy</td>
<td>43 (mRS 0–4 vs 5, 6)</td>
<td>2</td>
<td>1.10</td>
</tr>
</tbody>
</table>

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

*Four-level version of mRS.
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\(^\text{15,17,28}\) and empirical analyses using individual patient data from positive or negative (but not neutral) treatments;\(^\text{9,18,32–34}\) 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.\(^\text{19,20}\) 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.\(^\text{35}\) 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.\(^\text{21,36–38}\)

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.\(^\text{9}\) (Similarly, ordinalization of vascular prevention trial data showed that most trials adhered to proportionality of odds.)\(^\text{3}\)

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).\(^\text{32}\) For example, alteplase improves outcome independent of severity but can increase intracerebral hemorrhage, especially in those with severe stroke.\(^\text{32}\) 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).\(^\text{39}\) 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.\(^\text{18,40,41}\) In these situations,

---

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

<table>
<thead>
<tr>
<th>Trial</th>
<th>Intervention</th>
<th>Calibration Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>STICH(^\text{22})</td>
<td>Surgery for intracerebral hemorrhage</td>
<td>“Good” vs “bad” outcome dependent on prognostic factor based on age, Glasgow Coma Scale, and hemorrhage lesion volume</td>
</tr>
<tr>
<td>AbeSTT-II(^\text{15,23,24})</td>
<td>Abciximab for acute ischemic stroke</td>
<td>“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</td>
</tr>
<tr>
<td>PAIS(^\text{25,26})</td>
<td>Paracetamol for acute ischemic stroke</td>
<td>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</td>
</tr>
</tbody>
</table>

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.

---

\*Good\* indicates European Cooperative Acute Stroke Study.

---

![Figure 1. Distribution of modified Rankin Scale in the ECASS-3 study; note the shift to a good outcome with alteplase.\(^\text{30}\) ECASS indicates European Cooperative Acute Stroke Study.](http://stroke.ahajournals.org/)
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; (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. However, the statistical assumptions underlying the t test were not met for most of the acute data sets. 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, 2×2 χ² test, and median test. 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. 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.

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

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). 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). 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).

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). 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. 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.8

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 Elevation Test to Be Used parameter, which was related to an imbalance in severity at baseline.43

Like with binary analyses, shift analyses are more powerful statistically if adjusted for baseline prognostic covariates.33 The key covariates in acute stroke trials are the 2 most powerful prognostic factors: severity (e.g., 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.33 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×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.21,36–38

Sample Size Calculations

Standard methods exist for the calculation of sample size required for a given significance and power.44 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,45 Mann-Whitney U model of Noether,46 or proportional odds model of Whitehead.47 The comparison of means model is appropriate for continuous data. When comparing some of these approaches using parameters from completed trials, Whitehead’s approach resulted in the smallest sample size.32 Walters provides guidance on choosing the optimal sample size calculation for differing scenarios when using ordinal outcomes.48

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.32 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.33,40,50 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 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

<table>
<thead>
<tr>
<th>mRS</th>
<th>Test to Be Used</th>
<th>Parameter</th>
<th>Total Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed dichotomous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1/2–6</td>
<td>Logistic regression</td>
<td>ARR=5.0%</td>
<td>2882</td>
</tr>
<tr>
<td>0–2/3–6</td>
<td>Logistic regression</td>
<td>ARR=6.8%</td>
<td>2116</td>
</tr>
<tr>
<td>0–3/4–6</td>
<td>Logistic regression</td>
<td>ARR=7.2%</td>
<td>1972</td>
</tr>
<tr>
<td>0–4/5–6</td>
<td>Logistic regression</td>
<td>ARR=5.6%</td>
<td>2448</td>
</tr>
<tr>
<td>Ordinal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whitehead47</td>
<td>Ordinal logistic</td>
<td>Common odds=0.75</td>
<td>1566</td>
</tr>
<tr>
<td>Noether46</td>
<td>Mann-Whitney U</td>
<td>Delta=0.547</td>
<td>1533</td>
</tr>
<tr>
<td>Continuous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>t test</td>
<td>Difference=0.307</td>
<td>1536</td>
</tr>
</tbody>
</table>

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

Bath et al Primary Outcome in Acute Stroke Trials 1175

Downloaded from http://stroke.ahajournals.org/ at University of Glasgow Library on April 30, 2012
may be estimated for ordinal and continuous outcomes using expert generation of joint distribution outcome tables or derived mathematically from trial data. 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.

**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,” 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. 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.

**What Magnitude of Effect Is Relevant 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, 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.

---

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

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Patients per Annum (UK; a)</th>
<th>Eligibility, % (b)</th>
<th>Patients Receiving Treatment (c)</th>
<th>Cost per patient, £ (d)</th>
<th>Total Cost, £000 (e)</th>
<th>NNT (Ordinal)42 (f)</th>
<th>Patients Who Benefit (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>110 000</td>
<td>80</td>
<td>88 000</td>
<td>1</td>
<td>88</td>
<td>42</td>
<td>2095</td>
</tr>
<tr>
<td>Hemicraniectomy</td>
<td>110 000</td>
<td>0.1</td>
<td>110</td>
<td>7000</td>
<td>770</td>
<td>2.4</td>
<td>46</td>
</tr>
</tbody>
</table>

NNT indicates number needed to treat. c = a × b; e = c × d; g = c/f.
Acknowledgments
The workshop was sponsored by the European Stroke Organisation (ESO) with administrative assistance supplied by Marisa Kretsch 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
46. Noether G. Sample size determination for some common nonparametric tests.

45. Payne R.


