

Use of Ordinal Outcomes in Vascular Prevention Trials: Comparison With Binary Outcomes in Published Trials

Philip M.W. Bath, Chamila Geeganage, Laura J. Gray, Timothy Collier and Stuart Pocock

Stroke. 2008;39:2817-2823; originally published online July 31, 2008;

doi: 10.1161/STROKEAHA.107.509893

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2008 American Heart Association, Inc. 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/39/10/2817>

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

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

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

Use of Ordinal Outcomes in Vascular Prevention Trials Comparison With Binary Outcomes in Published Trials

Philip M.W. Bath, MD; Chamila Geeganage, MSc; Laura J. Gray, MSc;
Timothy Collier, MSc; Stuart Pocock, PhD

Background and Purpose—Vascular prevention trials mostly count “yes/no” (binary) outcome events, eg, stroke/no stroke. Analysis of ordered categorical vascular events (eg, fatal stroke/nonfatal stroke/no stroke) is clinically relevant and could be more powerful statistically. Although this is not a novel idea in the statistical community, ordinal outcomes have not been applied to stroke prevention trials in the past.

Methods—Summary data on stroke, myocardial infarction, combined vascular events, and bleeding were obtained by treatment group from published vascular prevention trials. Data were analyzed using 10 statistical approaches which allow comparison of 2 ordinal or binary treatment groups. The results for each statistical test for each trial were then compared using Friedman 2-way analysis of variance with multiple comparison procedures.

Results—Across 85 trials (335 305 subjects) the test results differed substantially so that approaches which used the ordinal nature of stroke events (fatal/nonfatal/no stroke) were more efficient than those which combined the data to form 2 groups ($P < 0.0001$). The most efficient tests were bootstrapping the difference in mean rank, Mann–Whitney U test, and ordinal logistic regression; 4- and 5-level data were more efficient still. Similar findings were obtained for myocardial infarction, combined vascular outcomes, and bleeding. The findings were consistent across different types, designs and sizes of trial, and for the different types of intervention.

Conclusions—When analyzing vascular events from prevention trials, statistical tests which use ordered categorical data are more efficient and are more likely to yield reliable results than binary tests. This approach gives additional information on treatment effects by severity of event and will allow trials to be smaller. (*Stroke*. 2008;39:2817-2823.)

Key Words: stroke ■ prevention ■ randomized controlled trial ■ statistical analysis

Major advances have been made in the primary and secondary prevention of stroke with effective strategies based on lifestyle modification, antithrombotic agents, blood pressure and cholesterol lowering, and carotid endarterectomy. In parallel, the absolute risk of recurrence has fallen dramatically over time; in stroke trials, this is apparent as a decrease in the control event rate, eg, 10.8% in the Canadian American Ticlopidine Study (CATS) in 1989² and 3.4% in Perindopril protection against recurrent stroke study (PROGRESS) in 2001.³ This trend is likely to continue as new and effective interventions are added. Because absolute event rates are a key component in sample size calculations for binary (“yes/no” event) outcomes, low rates equate to larger trials.⁴ An additional pressure in performing trials is that their number has increased as new prophylactic strategies are tested, eg, antiplatelets (thromboxane synthase inhibitors), anticoagulants (thrombin/factor Xa inhibitors), and carotid interventions (stenting, treatment of asymptomatic stenosis). The combination of more and larger trials means it is becoming increasingly difficult to find sufficient patients to enroll into new studies.

New strategies are required to bring trial sample sizes down and to maximize the potential to demonstrate benefit. In the past, composite outcomes of vascular death, nonfatal stroke, and nonfatal myocardial infarction (MI) have been used, in part to increase the number of events. This approach can be extended to include further events in the composite such as hospitalization, silent brain infarcts (as identified by MRI), or by counting all vascular events rather than just the first one.⁵ However, the use of composite outcomes has been criticized.⁶ An alternative approach is to analyze vascular prevention trials in a way which does not lose clinically relevant data. Most studies compare binary (stroke/no stroke) event rates between the treatment and control group. However, stroke or MI events may be fatal or nonfatal, so trichotomous outcomes (fatal event/nonfatal event/no event) can be analyzed. This approach can be extended to 4 (fatal stroke/severe nonfatal stroke/mild stroke/no stroke) or 5 (fatal stroke/severe nonfatal stroke/mild stroke/transient ischemic attack [TIA]/no event) levels. Similar ordered categorical outcomes can be developed for MI, composite vascular

Received November 12, 2007; accepted March 19, 2008.

From the Division of Stroke Medicine (P.M.W.B., C.G., L.J.G.), University of Nottingham, and the Medical Statistics Unit (T.C., S.P.), London School of Hygiene and Tropical Medicine, UK.

Correspondence to Professor Philip Bath, Stroke Trials Unit, Division of Stroke Medicine, University of Nottingham, Clinical Sciences Block, City Hospital campus, Hucknall Road, Nottingham NG5 1PB UK. E-mail philip.bath@nottingham.ac.uk

© 2008 American Heart Association, Inc.

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

DOI: 10.1161/STROKEAHA.107.509893

outcomes, and bleeding, as well as other vascular events, such as heart failure. The analysis of such ordered categorical (ordinal) events is usually more efficient statistically (because data on severity are not lost) thereby offering the potential for reducing trial sample size while maximizing the potential to find small clinically relevant treatment benefits.⁷ Such polytomization of events assumes that the ordering of events is meaningful, ie, that fatal vascular events are considered more severe than nonfatal ones. If so, ordinal outcomes may be more informative to patients, carers, healthcare professionals, and government than binary outcomes.

We report a comparison of the relative efficiencies of using and analyzing binary and polytomous outcomes from vascular prophylaxis trials. Although the use of ordinal statistical approaches is well defined in the methodological literature, its use for designing and analyzing vascular prevention trials is entirely novel.

Methods

Identification of Trials

We sought summary patient data from randomized controlled trials assessing primary or secondary vascular prevention, ie, preventing first or recurrent events respectively, 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, an approach which follows our previous study in acute stroke trials.⁷ We included vascular trials involving nonstroke patients and those measuring nonstroke outcomes because stroke patients suffer subsequent nonstroke vascular events, and those with other vascular conditions can go on to have a stroke. Taking this approach means the findings are generalizable across the field of vascular medicine. Published studies fulfilling these criteria were identified from electronic searches of the Cochrane Library and included studies of antithrombotic, BP or lipid lowering therapy, carotid endarterectomy, and hormone replacement therapy. Trials were excluded if they were neutral and related to a neutral intervention (as determined from a published meta-analysis) or did not include adequate ordered categorical information for at least one vascular outcome.

Trial Data

The numbers of subjects at the end of follow-up having a stroke (fatal, nonfatal, severe nonfatal, mild, TIA), MI (fatal, nonfatal), composite vascular event (fatal stroke or MI, nonfatal stroke or MI), and bleeding (major, minor, no bleeding) were obtained, where available, for each treatment group (active, control) from the primary trial publication. In factorial trials or those having more than two treatment groups,⁸ data were analyzed for each active comparison versus control. Data were assessed by intention-to-treat where possible.

Statistical Tests

We compared different statistical tests for assessing treatment effect.^{9–14} Some of these required the ordinal data to be combined into two groups (eg, Pearson's Chi-square test), whereas others used the raw ordered categorical data (eg, Mann-Whitney *U* test, un-pooled *t* test, bootstrapping the mean rank, ordinal logistic regression [also known as the proportional odds regression]). A description of the statistical tests used is given in the supplemental Appendix I, available online at <http://stroke.ahajournals.org>.

Comparison and Ordering of Statistical Tests

Each data set was analyzed using each statistical test. The results were then ordered within each trial and given a rank, with the lowest rank given to the test which produced the smallest probability value within that trial. A 2-way analysis of variance test (Friedman with

Table 1. Assessment of 10 Statistical Approaches for Analyzing Stroke as a 3-Level Event (Fatal/Nonfatal/No Stroke) in 85 Vascular Prevention Trials

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

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

adjustment for ties¹⁵; ANOVA) was then performed to assess which statistical test produced the lowest ranks (ie, the most statistically significant values). Duncan multiple range test was used to assess the ordering of tests and determine where significant differences between tests were present. We also assessed how many statistically significant (at 5%) results each test found.

To assess the validity and reliability of the results found, a number of supplementary analyses were carried out. First, the comparison of statistical tests was repeated within subgroups of trials sharing similar characteristics to assess whether particular types of trials suited different statistical approaches; second, the statistical assumptions of the tests were assessed; and third, the sensitivity (type 1 error) of the tests was assessed. Technical details of these supplementary analyses can be found in the supplemental Appendix II.

Analyses were carried out in SAS (version 8.2) and Stata (version 7); significance was taken at $P<0.05$ for analyses of trials and $P<0.01$ for ANOVA.

Results

Trials

Of 243 identified trials, 101 (416 020 subjects) were included, these comprising 35 primary and 66 secondary prevention studies (supplemental Table I). One hundred forty-two trials were excluded, mostly because their published data did not distinguish between fatal and nonfatal vascular events so that 3-level data could not be calculated (supplemental Table II).

Stroke

The trials variably included intracerebral hemorrhage within the outcome of stroke. The results of the statistical tests differed significantly with 3-level data (fatal stroke/nonfatal stroke/no stroke; 85 trials, 335 305 subjects; ANOVA $P<0.0001$); ordinal analyses ranked above binary approaches (Tables 1 and 2; Figure 1) with the Mann-Whitney *U* test, bootstrapping (difference in mean rank), and ordinal logistic regression significantly better than the other methods (sup-

Table 2. Ranking of Statistical Tests (1 to 10 With 1 “Best”) for Measure of Stroke (3, 4, and 5-Levels), Myocardial Infarction (3-Level), Composite Vascular Outcome (3-Level), and Bleeding (3-Level)

Outcome	Trials	P Value	Ranking of Tests Relative to Each Other									
			MWU	BS	OLR	RRT	CAT	t Test	χ^2 2x3	χ^2 Event	χ^2 Dead	Median Test
Fatal stroke/nonfatal stroke/no stroke	85	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	4	5	6	7	8	9	10
Fatal stroke/severe nonfatal/mild/no stroke	21	<0.0001	<u>2</u>	<u>1</u>	4	3	5	6	8	7	9	10
Fatal stroke/nonfatal stroke/TIA/no stroke	29	<0.0001	<u>2</u>	<u>1</u>	<u>5</u>	<u>6</u>	<u>3</u>	<u>4</u>	7	8	9	10
Fatal stroke/severe nonfatal stroke/mild stroke/TIA/no stroke	11	<0.0001	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>1</u>	<u>2</u>	8	7	9	10
Fatal MI/nonfatal MI/no MI	58	<0.0001	<u>1</u>	<u>3</u>	<u>5</u>	6	<u>2</u>	<u>4</u>	7	8	9	10
Fatal vascular event/nonfatal vascular event/no vascular event	43	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	5	6	7	8	9	10
Severe-major bleeding/minor bleeding/no bleeding	15	<0.0001	<u>3</u>	<u>2</u>	<u>1</u>	6	<u>4</u>	<u>5</u>	8	7	9	10

The most efficient tests are underlined and do not differ from each other statistically.

BS indicates bootstrap; CAT, Cochran–Armitage test; MWU, Mann–Whitney U test; OLR, ordinal logistic regression; RRT, robust ranks test.

plemental Figure I). Similar results were seen for the other stroke outcome assessments: 4-level (fatal stroke/severe nonfatal stroke/mild stroke/no stroke), 4-level including TIA (fatal stroke/nonfatal stroke/TIA/no stroke or TIA), and 5-level (fatal stroke/severe nonfatal stroke/mild stroke/TIA/no stroke or TIA; each ANOVA $P < 0.0001$; Table 2). Although the absolute ordering of the tests varied for these polytomous outcomes, ordinal tests always performed better than binary ones (Table 2). Six trials gave sufficient data to compare qualitatively 3-, 4-, and 5-level stroke data; 4-level data (with TIA included as an event) and 5-level data (including TIA) appeared to be the most efficient approaches. When assessed by how many trials were statistically significant (positive or negative but not neutral), those tests which did not collapse the data into groups again out-performed other approaches; for example the Mann–Whitney U test

gave a statistically significant result in 44% of trials in comparison with the Pearson’s χ^2 2x3 test at 32% (Figure 1).

Myocardial Infarction

Fifty-eight trials (232 515 subjects) were included. The analyses differed significantly for a 3-level outcome (fatal MI/nonfatal MI/no MI; $P < 0.0001$), with ordinal approaches performing better than binary (Table 2).

Composite Vascular Event

Forty-three trials (204 108 subjects) gave data for a 3-level composite vascular outcome (fatal stroke or MI/nonfatal stroke or MI/no stroke or MI). Ordinal tests performed best ($P < 0.0001$) with the Mann–Whitney U test, bootstrapping (the difference in mean rank) and ordinal logistic regression ranking highest (Table 2).

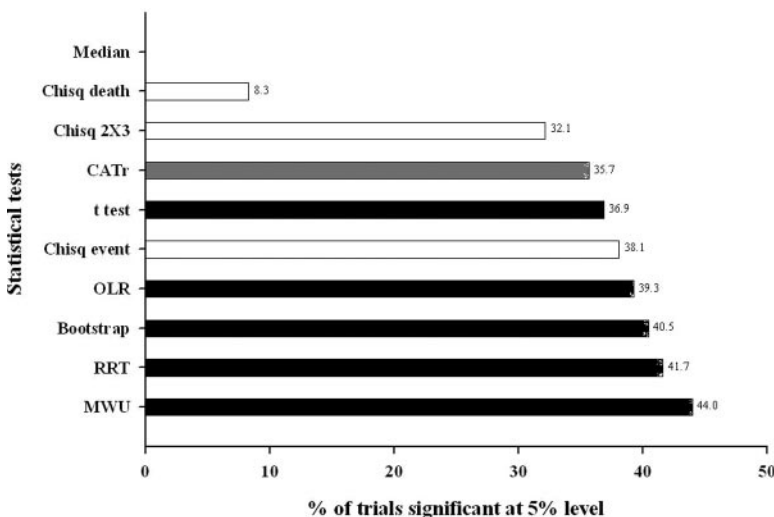


Figure 1. The number of significant trials (positive or negative but not neutral, $P < 0.05$) for each statistical test for 3-level stroke (fatal, nonfatal, no stroke).

Table 3. Ranking of Statistical Tests (1 to 10 With 1 'Best') for 3-Level Stroke (Fatal, Nonfatal, No Stroke) in Subgroups of Vascular Prevention Trials

Outcome	Trials	P Value	Ranking of Tests Relative to Each Other									
			MWU	BS	OLR	RRT	CAT	t Test	χ^2 2×3	χ^2 Event	χ^2 Dead	Median Test
Prevention, primary	29	<0.0001	<u>1</u>	<u>2</u>	5	<u>3</u>	6	4	7	8	9	10
Prevention, secondary	56	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	4	5	6	7	8	9	10
Anticoagulants	12	<0.0001	<u>2</u>	<u>1</u>	<u>7</u>	<u>6</u>	<u>4</u>	<u>3</u>	<u>5</u>	8	9	10
Antiplatelets	33	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	5	<u>4</u>	6	7	8	9	10
Antihypertensives	23	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	<u>4</u>	5	6	7	8	9	10
Lipid lowering	10	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	<u>4</u>	<u>5</u>	6	8	7	9	10
Carotid endarterectomy	4	<0.0001	<u>2</u>	<u>1</u>	<u>5</u>	<u>3</u>	6	<u>4</u>	8	7	9	10
Hormone replacement therapy	2	0.86
Age <65 years	34	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	4	5	6	7	8	9	10
Age >65 years	31	<0.0001	<u>2</u>	<u>1</u>	4	3	6	5	7	8	9	10
Trial, small (n<2,520)	42	<0.0001	<u>1</u>	<u>2</u>	5	6	<u>3</u>	<u>4</u>	7	8	9	10
Trials, large (n>2,520)	42	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	<u>4</u>	5	6	8	7	9	10
Follow-up, short term (<36 months)	45	<0.0001	<u>1</u>	<u>2</u>	4	5	3	6	7	8	9	10
Follow-up, long term (>36 months)	39	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	<u>4</u>	5	6	7	8	9	10
Risk of death in control, low (<0.2% per month)	43	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	4	5	6	7	8	9	10
Risk of death in control, high (>0.2% per month)	41	<0.0001	<u>2</u>	<u>1</u>	3	4	5	6	7	8	9	10
Risk of stroke in control, low (<0.17% per month)	40	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	4	5	6	7	8	9	10
Risk of stroke in control, high (>0.17% per month)	41	<0.0001	<u>2</u>	<u>1</u>	5	6	3	4	7	8	9	10
Time from index event, short (<87 days)	22	<0.0001	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	5	6	8	7	9	10
Time from index event, long (>87 days)	22	<0.0001	<u>2</u>	<u>1</u>	<u>3</u>	6	<u>4</u>	<u>5</u>	7	8	9	10

The most efficient tests are underlined and do not differ from each other statistically.

BS indicates bootstrap; CAT, Cochran–Armitage test; MWU, Mann–Whitney *U* test; OLR, ordinal logistic regression; RRT, robust ranks test.

Bleeding

Fifteen trials (26 215 patients) were identified as including information on bleeding at three levels: major bleeding, minor bleeding, no bleeding. Definitions of bleeding differed between trials. Once again, ordinal analytic approaches ranked highest (Table 2).

Sensitivity Analysis and Test Assumptions

The ordering of statistical tests, with ordinal more efficient than binary, was maintained for all subgroups of trials irrespective of type of prevention and treatment, average age of patients, trial size and length of follow-up, risk of death or stroke, and time from index event (Table 3). When considering the 19 trials (27 datasets) with a high event rate (>10%

overall), ordinal tests remained most efficient. Published hazard ratios (which take into account the time to event, as derived from the Cox proportional hazards model) for stroke were available for 36 trials; a comparison of the 11 statistical tests, including Cox results, revealed bootstrapping, Mann–Whitney *U*, and ordinal logistic regression to be as good if not slightly superior to the Cox model (Duncan multiple range test).

The statistical assumptions for ordinal logistic regression were not violated ($P>0.05$) in 79 of 85 trials with 3-level stroke data; no violations were present for 11 trials with 5-level stroke data (supplemental Appendix III). The sensitivity analysis showed that the top performing statistical tests (ordinal logistic regression, Mann–Whitney *U* test) were not

overly sensitive, and statistically significant treatment effects were only found where they are likely to be present (supplemental Appendix III). Using ordinal logistic regression, the odds ratios were similar for different strata of severity for 3-level, 4-level, and 5-level data (supplemental Table III).

Discussion

Improvements in secondary prevention are leading to falling event rates in clinical trials. This means that future vascular prevention trials will need to be longer and, with an increasing number of new interventions, the availability of subjects is becoming limited. Thus, new approaches to trial design and analysis are needed to help reduce sample size. This study has shown that it is feasible to create 3-level ordered categorical outcomes for stroke, MI, a composite vascular event (fatal stroke and MI/nonfatal stroke and MI), and bleeding. Analysis reveals that, in general, statistical approaches which use ordinal data are more efficient than conventional binary tests based on “event/no event.” A further increase in efficiency comes from using 4-level or 5-level data for stroke (with or without TIA). Ordering vascular events by severity has both biological and clinical meaning. Fatal events are clearly the most extreme health state whereas a severe stroke (normally defined as a stroke resulting in dependency on others) is a disaster for the patient, their career, and society, for both clinical and economic reasons. A mild stroke leaves the patient independent, even if residual impairment remains, and those who are younger can often return to work.

The most efficient statistical tests were those which examined ordinal data, including ordinal logistic regression, the Mann–Whitney *U* test, and bootstrapping the mean rank. In addition to improving statistical efficiency, the use of ordered categorical outcomes gives information on the ability of an intervention to reduce the severity of an event, not just the number of events. Ordinal logistic regression allows both estimation (with confidence intervals) and inclusion of baseline prognostic covariates in analyses. However, it assumes that any treatment effect is similar across outcome levels, ie, the odds of moving a treated patient from fatal to severe nonfatal stroke are similar to those for moving from TIA to no event (“proportionality of odds”). This assumption requires justification because it is neither widely recognized nor obvious in most published vascular trial data. First, it is biologically plausible to suggest that prophylactic interventions will reduce severity as well as the total number of events. Since the development of atherosclerosis and increases in thrombosis, coagulation and inflammation are not binary events in nature, and their magnitude is a determinant of the severity of clinical vascular events, it is reasonable to expect that interventions will move patients from fatal to severe, severe to mild, and mild to no events. If this assumption (of proportional odds) is not met, an alternative ordinal model could be considered.¹⁶

Second, there is existing published evidence that interventions do alter severity: simvastatin reduced the risk of stroke of different severities by similar risk reductions in the Heart Protection Study (HPS),¹⁷ hormone replacement therapy increased both stroke and its severity in the Women’s Estrogen for Stroke Trial (WEST),¹⁸ and antiplatelet agents reduced

both fatal and nonfatal vascular events in the Antithrombotic Trialists’ (ATT) Collaboration meta analysis.¹⁹ The apparent failure of most vascular prevention trials to show individual effects on death or severe events is largely because they were not powered to assess these specific and, therefore, relatively uncommon events. Third, the odds reduction at each outcome level appeared to be relatively constant when individual trials were assessed (Figure 2); formal statistical assessment using the likelihood ratio test indicated that “proportionality of odds” was present in most cases (although this test is known to be conservative; Appendix 6). Last, using ordinal statistical tests was more powerful than binary approaches, the central finding of this study. Although this is not a novel idea in the statistical community,²⁰ ordinal outcomes have not been applied to vascular prevention trials in the past. In this context, it is worth noting that ordinal logistic regression is relatively robust to deviations in its assumptions even if they are not met in a particular trial. Another efficient ordinal test is the Mann–Whitney *U* test, which is widely available in statistical packages and can produce a point estimate (median difference between groups) with confidence intervals. The major assumption of the test is that the treatment groups should be independent, and this is met here. The final efficient statistical approach was bootstrapping the mean rank; this approach is computer intensive¹³ and its application and the interpretation of results are not well appreciated by clinicians, although it is free of assumptions.

The conventional approach to analyzing vascular prevention trials is to perform time to event analyses, as visualized using Kaplan–Meier curves and analyzed with Cox regression. When the frequency of events is high, analyses based on time-to-event are more efficient than those using frequencies (as analyzed using logistic regression). However, the frequency of vascular events in most primary and secondary prevention trials running over 3 to 5 years is relatively low; recent vascular prevention trials have tended to report annualized stroke rates of 2% to 4%.^{21,22} Logistic and Cox models give similar results when the overall event frequency is less than 10%.^{23,24} Where the frequency of events is higher, ordinal data may be analyzed by time to event.^{25,26} In the current dataset, the Cox model was slightly less efficient than bootstrapping, Mann–Whitney *U*, and ordinal logistic regression.

In this study, we have focused on assessing stroke as the primary outcome rather than using a composite vascular outcome (fatal vascular event, nonfatal stroke, and MI). Stroke was of interest since it has been used in several prevention trials, eg, the European Stroke Prevention Study-II (ESPS-II) and PROGRESS,^{3,27} and 4- or 5-level data (including TIA) may be created. Nevertheless, ordered categorical outcomes may also be created for composite outcomes (fatal stroke or MI/nonfatal stroke or MI/no event) as well as other events such as MI or bleeding. Our results suggest that the use and analysis of polytomous outcomes would benefit trials assessing any of these vascular outcomes, and it is likely that the approach would work for others such as heart failure and venous thromboembolism; we are currently assessing this.

Using ordered categorical data will mean that results will need to be reported differently. The results of binary tests are

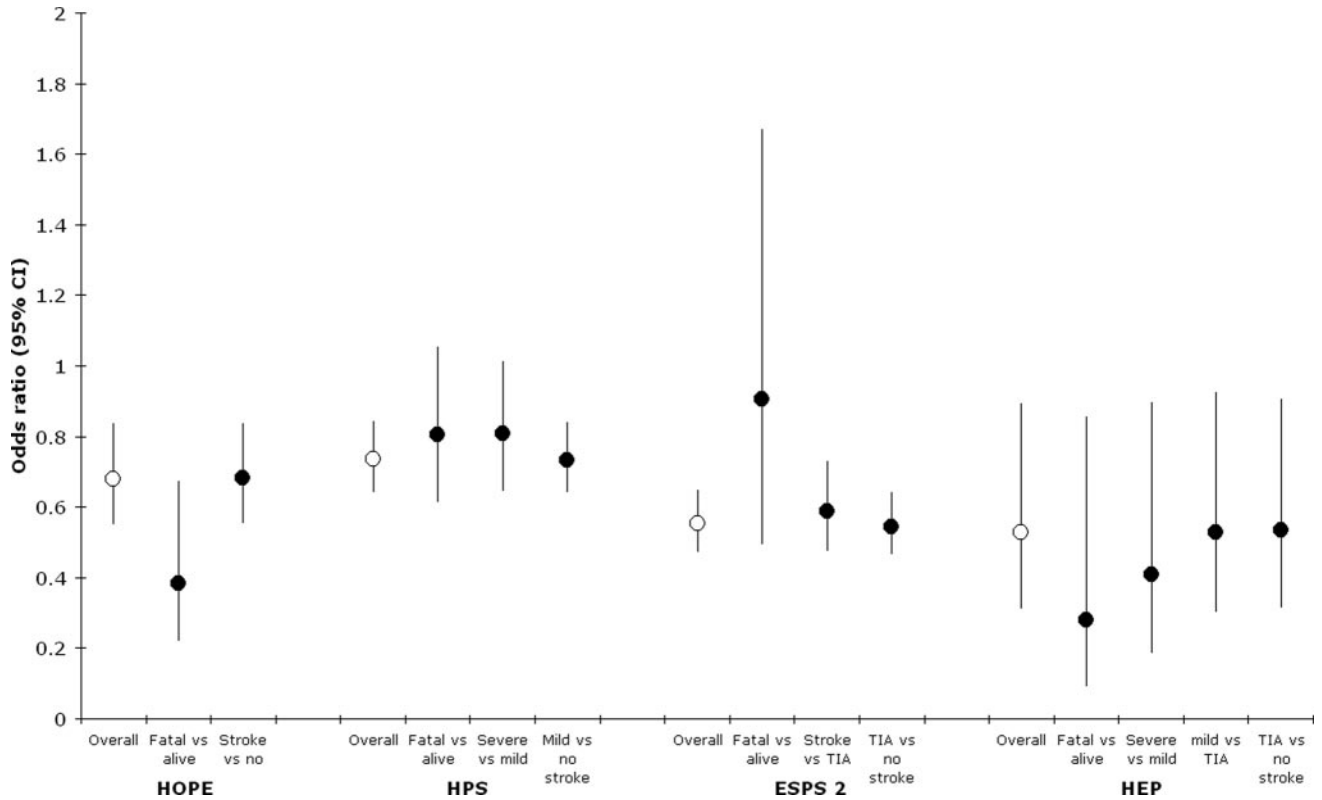


Figure 2. Odds ratios across trial (by ordinal logistic regression) and by individual outcome levels for 4 trials to illustrate the assumption of proportionality of odds.

summarized easily as the proportion of patients who benefit (or suffer) with a treatment, ie, oral anticoagulation reduced absolute stroke recurrence by 1.46% (odds ratio 0.75, $P=0.036$) in the Anticoagulants in the Secondary Prevention of Events in Coronary Thrombosis (ASPECT) trial.²⁸ In contrast, ordinal tests will need to be presented as the average absolute improvement in outcome, eg, anticoagulation reduced stroke recurrence and its severity with an odds ratio of 0.60 (or reduced the mean severity by 0.5 points, $P=0.013$) on a 5-level scale.²⁸ In this respect, health consumers will need to decide what odds ratio or difference in events is worthwhile, both clinically and in terms of health economics. In reality, it is reasonable to present the primary result using the odds ratio (or median change in event severity) and to

give the absolute percentage change calculated from the binary outcome as a secondary measure. Further, a visual presentation of the data can be displayed as the percentage of patients within each category by treatment group (data from the North American Symptomatic Carotid Endarterectomy Trial [NASCET], Figure 3).

Just as sample size calculations exist for trials using dichotomised analyses,⁴ analogous approaches exist for ordinal tests.²⁹ Because ordinal analyses are more powerful statistically, trial size may be reduced for a given power of say 90%; eg, sample size falls by 15% to 24% as the number of outcome categories increases from 3 to 7.²⁹ This reduction is worthwhile and would reduce competition between trials for patients, and lower trial costs and complexity. Taking the

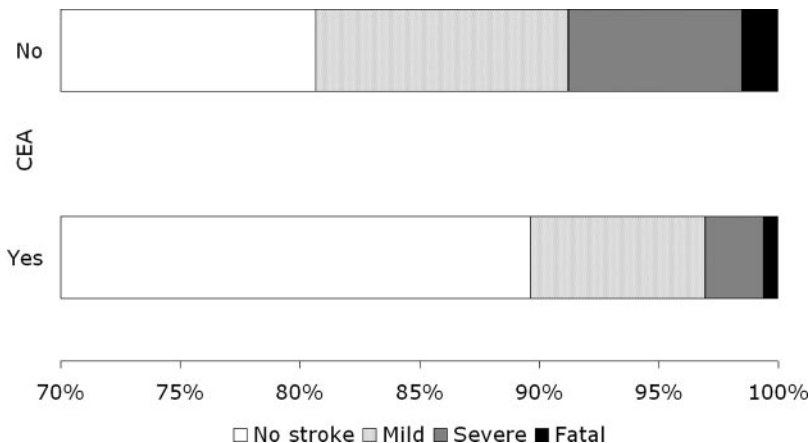


Figure 3. Example 4-level ordinal data from NASCET¹ of carotid endarterectomy (CEA).

Hypertension in Elderly Patients (HEP) trial³⁰ as an example (and assuming significance=0.05 and power=0.9), the sample size is reduced by 48% from 1556 for a binary outcome of stroke/no stroke to 810 for a 3-level stroke outcome as calculated using the method of Whitehead;²⁹ this is further reduced to 772 with a 5 level stroke outcome.

A number of caveats must be made about this study. First, a majority of identified trials could not be included because they did not publish adequate information on vascular events. As data were missing for a variety of trial types (primary, secondary prevention), sizes, and outcome measures (stroke/MI/vascular/bleeding) it is unlikely that a systematic bias was introduced into the findings; however, the precision of the results will have been attenuated by the missing data. Future trial publications should give this information, including vital status for the main vascular outcomes, so that ordered outcome categories can be calculated. Second, we did not use all possible statistical tests relevant to the problem of analyzing ordered categorical data; instead, we focused on those approaches which are readily available in statistical textbooks¹¹ and computer packages.

In summary, we suggest that vascular prevention trials should consider using statistical approaches, which use the inherent ordered categorical data present within vascular outcome events. The resulting trials could be smaller (with savings in patient numbers, numbers of centers, and study cost and complexity) and would allow appreciation of the effect of interventions on severity, as well as absolute number of events, to be highlighted. Appropriate tests include ordinal logistic regression, the Mann–Whitney *U* test, and bootstrapping the mean rank.

Sources of Funding

The Division of Stroke Medicine (University of Nottingham) receives core funding from The Stroke Association (UK). P.B. is Stroke Association Professor of Stroke Medicine. L.G. receives funding from the Medical Research Council. The funding sources had no involvement in the project.

Disclosures

P.W.B. participated in some of the included trials.

References

- North Am Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med.* 1991;325:445–453.
- Gent M, Blakely JA, Easton JD, Ellis DJ, Hachinski VC, Harbison JW, Panak E, Roberts RS, Sicurella J, Turpie AGG, Group TC. The Canadian American Ticlopidine study (CATS) in thromboembolic stroke. *Lancet.* 1989;1215–1220.
- PROGRESS Collaborative Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *Lancet.* 2001; 358:1033–1041.
- 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.
- Schrader J, Luders S, Kulschewski A, Hammersen F, Plate K, Berger J, Zidek W, Dominiak P, Diener H-C, for the MOSES study group. Morbidity and mortality after stroke, eprosartan compared with nitrendipine for secondary prevention. Principal results of a prospective randomised controlled study (MOSES). *Stroke.* 2005;36:1218–1226.
- Ferreira-Gonzalez I, Permanyer-Miralda G, Domingo-Salvany A, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Pacheco-Huergo V, Wu P, Mills EJ, Guyatt GH. Problems with the use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ.* 2007;334:786.
- 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.
- UK-TIA Study Group, 2025. The United Kingdom Transient Ischaemic Attack (UK-TIA) aspirin trial: Final results. *J Neurol Neurosurg Psych.* 1991;54:1044–1054.
- Altman DG. *Practical Statistics for Medical Research.* London: Chapman & Hall; 1991.
- Conover WJ. *Practical Nonparametric Statistics.* New York: John Wiley & Sons; 1971.
- Siegel S, Castellan NJ. *Nonparametric Statistics for the Behavioral Sciences.* Singapore: McGraw-Hill; 1988.
- Fligner MA, Policello GE. Robust rank procedures for the Behrens-Fisher problem. *J Am Stat Assoc.* 1981;76:162–168.
- Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* New York: Chapman & Hall; 1993.
- 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.
- Hollander M, Wolfe DA. *Nonparametric Statistical Methods.* New York: John Wiley & Sons inc; 1999.
- Stokes ME, Davis CS, Koch GG. *Categorical Data Analysis Using SAS.* Cary, NC: SAS Institute; 1995.
- Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: A randomised placebo-controlled trial. *Lancet.* 2002;360:7–22.
- Viscoli CM, Brass LM, Kernan WN, Sarrel PM, Suissa S, Horwitz RI. A clinical trial of estrogen-replacement therapy after ischemic stroke. *N Engl J Med.* 2001;345:1243–1249.
- Antithrombotic Trialists Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ.* 2002;324:71–86.
- Valenta Z, Pitha J, Pioledne R. Proportional odds logistic regression - effective means of dealing with limited uncertainty in dichotomizing clinical outcomes. *Stat Med.* 2006;25:4227–4234.
- Bhatt DL, Fox KAA, Hacke W, Berger PB, Black HR, Boden WE, Cacoub P, Cohen EA, Creager MA, Easton JD, Flather MD, Haffner SM, Hamm CW, Hankey GJ, Johnston SC, Mak K-H, Mas J-L, Montalescot G, Pearson TA, Steg PG, D, Steinhubl SR, Weber MA, Brennan DM, Fabry-Ribaud L, Booth J, Topol EJ, for the CHARISMA Investigators. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med.* 2006;354:1706–1717.
- The Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) Investigators. High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med.* 2006;355:549–559.
- Ingram DD, Kleinman JC. Empirical comparisons of proportional hazards and logistic regression models. *Stat Med.* 1989;8:525–538.
- Annesi I, Moreau T, Lellouch J. Efficiency of the logistic regression and Cox proportional hazards models in longitudinal studies. *Stat Med.* 1989; 8:1515–1521.
- Schatzkin AR, Cupples LA, Heeren T, Morelock S, Kannel WB. Sudden death in the Framingham heart study. Differences in the incidence and risk factors by sex and coronary disease status. *Am J Epidemiol.* 1984; 120:888–899.
- Berridge DM, Whitehead J. Analysis of failure time data with ordinal categories of response. *Stat Med.* 1991;10:1703–1710.
- Diener HC, Cunha L, Forbes C, Sivenius J, Smets P, Lowenthal A. European stroke prevention study 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *J Neurol Sci.* 1996;143:1–13.
- Anticoagulants in the Secondary Prevention of Events in Coronary Thrombosis (ASPECT) Research Group. Effect of long-term oral anticoagulant treatment on mortality and cardiovascular morbidity after myocardial infarction. *Lancet.* 1994;343:499–503.
- Whitehead J. Sample-size calculations for ordered categorical-data. *Stat Med.* 1993;12:2257–2271.
- Coope J, Warrender TS. Randomised trial of treatment of hypertension in elderly patients in primary care. *BMJ.* 1986;293:1145–1151.