Novel End Point Analytic Techniques and Interpreting Shifts Across the Entire Range of Outcome Scales in Acute Stroke Trials

Jeffrey L. Saver, MD

Background and Purpose—Stroke treatments are generally not curative, but rather alter patient outcome over the entire range of functional measures. Dichotomizing outcome scales reduces computational complexity, but discards substantial outcome information, artificially privileges only a single health state transition as clinically meaningful, and often reduces study power. Newer approaches to endpoint analysis have several advantageous properties.

Summary of Review—The global statistic assesses treatment effects on multiple outcome measures simultaneously. However, translating the global statistic multidimensional vector effect at the population level into benefit or harm expected in the individual patient is problematic. Responder analysis adjusts outcome thresholds to patient stroke severity at study entry, identifying achievable goals for each patient. However, responder analysis still discards substantial outcome information. Shift analysis gauges change in outcome distributions over the full range of ascertained outcomes, incorporating benefit and harm at all health state transitions valued by patients and clinicians, and often increasing study power. Translation of findings of shift analyses into clinically accessible terms may be accomplished using the recently developed joint outcome table specification technique, which yields the following values for the number needed to treat for 1 patient to improve in a clinically important manner: nimodipine in subarachnoid hemorrhage, 6.8; coiling over clipping, 5.9; intra-arterial pro-urokinase in acute cerebral ischemia, 4.8; intravenous tissue plasminogen activator, 3.3.

Conclusions—Dichotomized, global statistic, responder, and shift analyses each offer distinctive benefits and drawbacks. Choice of primary end point analytic technique should be tailored to the study population, expected treatment response, and study purpose. Shift analysis generally provides the most comprehensive index of a treatment's clinical impact. (***Stroke.*** **2007;38:3055-3062.**)

Key Words: cerebral hemorrhage ■ cerebral infarction ■ cerebrovascular accident ■ clinical trials ■ research design

In approximately the year 2000, the field of acute stroke clinical trials experienced a scientific crisis. Over the course of the 20th century, at least 75 promising agents had emerged from animal testing, rich with promise, to undergo clinical study. Although at least 178 randomized clinical trials enrolling >73 000 patients were conducted of these agents, only 3 trials reported positive findings and only 1 agent had been approved by the Food and Drug Administration for use in acute cerebral ischemia.[1] This remarkable record of failure led to a fundamental reassessment of the methodological framework of preclinical and clinical development of pharmacologic agents for acute stroke.[2–6]

In the sphere of clinical trial design, 6 common defects were noted: (1) treatment was initiated too late; (2) dosages of neuroprotective class agents were too low; (3) patients were enrolled who were unable to respond to the study agent mechanism of action; (4) patients with uninformative deficits were enrolled; (5) studies were underpowered; and (6) the endpoints and statistical endpoint analytic techniques discarded substantial outcome information. These designs led to studies that examined infarcts that were already largely completed; failed to attain adequate neuroprotective tissue levels of the study drug; and enrolled inappropriate patients, such as those with pure white matter infarcts for studies of drugs active only at gray matter synaptic terminals or those with deficits too mild or too severe to permit demonstration of a treatment effect. The use of small sample sizes yielded studies insensitive to modest but clinically important benefits of tested agents. Discarding relevant outcomes in primary endpoint analysis reduced study power.

These insights have informed the design many 21st century trials, which have made substantial efforts to accelerate time of treatment start, attain neuroprotective blood levels, enroll only patients with informative conditions, increase samples sizes, and use novel statistical techniques to increase study power.

Received March 19, 2007; final revision received April 26, 2007; accepted April 27, 2007.
From Stroke Center and Department of Neurology, David Geffen School of Medicine at the University of California, Los Angeles, Calif.
Correspondence to Jeffrey L. Saver, MD, UCLA Stroke Center, 710 Westwood Plaza, Los Angeles, CA 90095. E-mail jsaver@ucla.edu

Stroke **is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.107.488536**

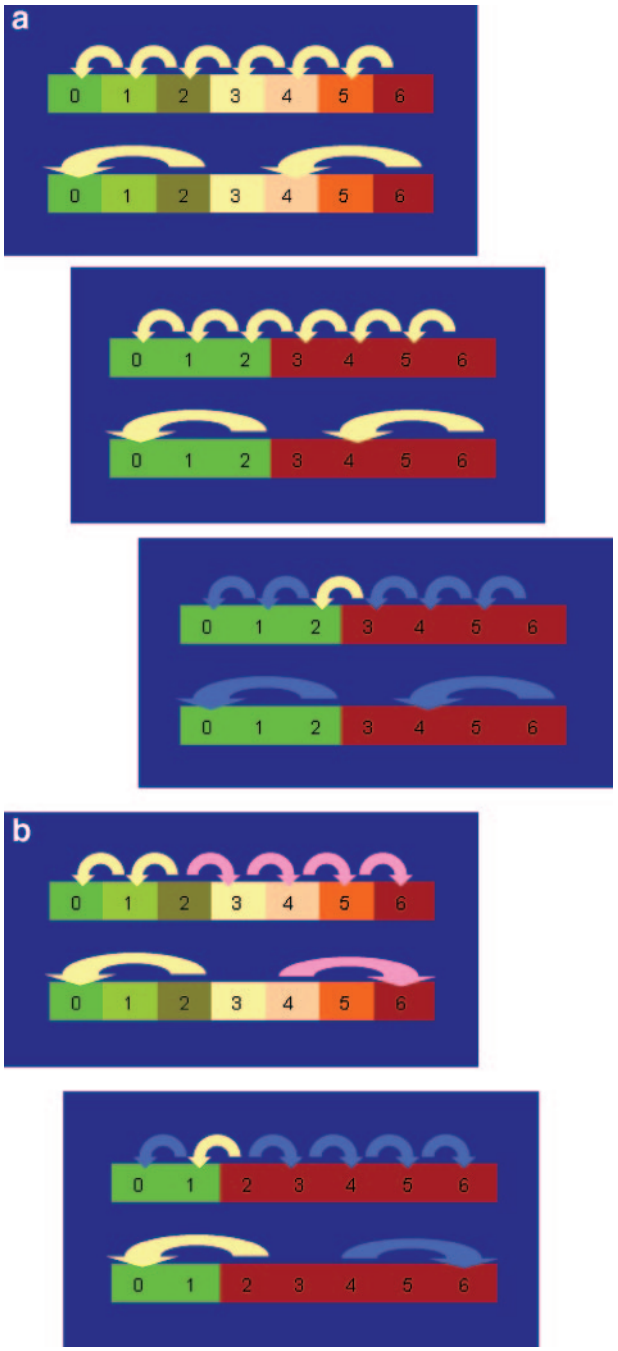

Figure 1. a, Discarding of outcome information in dichotomized analysis. Top panel, Schematic illustrations of 7-level outcome scale. Top row, A range of potential small step improvements in response to therapy. Lower row, A range of moderate step improvements. Middle panel, 7-level scale converted to a binary measure by dichotomization at 0 to 2 vs 3 to 6. Lower panel, 3 to 2 health state transition improvements measured in the binary analysis (white arrow) and improvements that go unmeasured in the primary analysis (light blue arrows) (©UCLA Stroke Program). b, Worsening at unanalyzed transitions in dichotomized analysis. Top panel, Schematic illustration of a response pattern on a 7-level outcome scale to an agent that benefits patients with mild stroke severities at trial entry but harms patients with moderate to severe strokes at entry (eg, fibrinolytic drug with high hemorrhagic propensity). Top row, Small step changes. Lower row, Moderate step changes. Bottom panel, After conversion of the 7-level scale to a binary measure by dichotomization at 0 to 1 vs 2 to 6, only at improvements at the 2 to 1 health state transition are counted (white arrows). The binary analysis suggests the agent is beneficial, even though it causes harm to more patients (light blue arrows) than it benefits (©UCLA Stroke Program).

Perhaps the most exotic of these innovations to clinicians are novel statistical analysis approaches used in recent and current trials. Simple dichotomized analysis of a single end point has virtually disappeared as a primary end point. Nearly all acute stroke clinical trials now use 1 of 3 new approaches to end point statistical assessment: (1) the global statistic; (2) responder analysis; and (3) shift analysis. Hence, the need to understand the advantages, disadvantages, and interpretation of these prevalent statistical analytic techniques takes on great urgency. This review surveys these novel statistical techniques, with special emphasis on shift analysis, as the least familiar of these approaches. Understanding these innovative outcome analysis strategies is facilitated by placing them in perspective against the traditional dichotomized approach to end point analysis.

Traditional Dichotomized End Point Analysis

Stroke treatments are generally not curative, but rather confer improvement to patients along an entire range of functional outcomes. A patient's final health state may range from full recovery, with no residual symptoms; through mild symptoms without alteration in vocational capacity; varying levels of dependence in activities of daily living; permanent vegetative state; to death. To capture this range of health states, the outcome scales commonly used in stroke trials rate patients across multiple ranks.[7] The modified Rankin Scale (mRS) divides global disability into 7 strata, the Barthel Index rates functional activities of daily living among 20 levels, and the NIHSS assigns patients according to 42 ranks of neurologic deficit severity. Stroke-related quality-of-life scales have even finer gradations, further reducing ceiling and floor effects.

Dichotomized end point analyses convert these ordinal scales, which have a spectrum of outcomes, into binary measures. The dichotomized statistical approach can be thought of as asking, "Does the treatment make more patients achieve a good outcome, defined as better than threshold X?" Binary analysis has the advantage of computational simplicity and straightforward clinical interpretation. Net number needed to treat to benefit (NNT) values are easily derived from dichotomized end points. This derivation allows clinicians to clearly convert trial results into how many patients are needed to achieve one additional net good outcome (where "good" means exceeding the dichotomization threshold).

Binary analysis also has several disadvantages. It requires that trialists privilege as valuable only 1 health state transition among the many that matter to patients and caregivers. The choice regarding which health state transition to value above all others is often made arbitrarily. Moreover, by inspecting only a single health state transition, binary analysis forces trialists to discard substantial amounts of outcome information, which will generally lead to underestimation of treatment benefit or treatment harm or both (Figure 1a).

Table 1. Modified Rankin Scale

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to perform all usual duties and activities
2	Slight disability; unable to perform all previous activities, but able to take care of self without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent, and requiring constant nursing care and attention
6	Dead

Consider, for example, the most commonly used primary end point in stroke clinical trials, the mRS of global disability (Table 1).[10] When the 7-level mRS is dichotomized as 0 to 2 vs 3 to 6, a traditional breakpoint in binary trial analysis, the resulting analysis examines only 1 important transition in health state, from vocationally impaired, but able to live independently, to requiring assistance in daily living. However, this analysis places absolutely no value on other health state transitions that are pertinent to patients. For example, going from vocationally impaired (mRS=2) to no symptoms at all (mRS=0) is not counted as a clinically meaningful improvement, nor is going from dead (mRS=6) to moderately disabled and able to walk on one's own (mRS=3). Binary outcome analyses prioritize only a single health state transition as clinically worthwhile, whereas patients naturally place great value on several health state transitions.[8–10]

The focus of binary analysis on a single transition in health state is of particular concern when a treatment has beneficial and harmful effects differentially distributed over the range of disease severities. For example, a treatment such as a fibrinolytic drug with a strong hemorrhagic propensity might be beneficial for patients with milder strokes, but harmful for patients with more severe stroke. When a trial of this type of drug is analyzed with a binary primary end point focused on a health state transition to very good outcomes, the statistical analysis might be formally positive for a beneficial effect. Yet the agent could actually be harming more patients in the unanalyzed severe end of the outcome spectrum than it is helping in the analyzed milder end (Figure 1b).

Newer Approaches to End Point Analysis

Because of the drawbacks of dichotomization, several new approaches to end point analysis have recently been adopted in acute stroke clinical trials (Figure 2), including the global statistic, responder analysis, and shift analysis.

Global Statistic

The global statistic simultaneously looks at multiple outcome measures, each related to an underlying unitary health state, and examines whether treatment is increasing the proportion of patients with good outcomes on each measure.[11] The global statistic can be thought of as asking the question, "Does the treatment make the treated population better as assessed in these several ways?"

The global statistic has several advantageous features. It reflects the reality that many measurement instruments capture desired health states only partially. For example, patients and physicians consider good stroke recovery to include all of the following: having few neurologic deficits, being independent, having a high level of cognitive function, and having a high quality of life. Generalized estimating equations and other mathematical techniques can combine individual statistics, each measuring the association between a treatment and 1 aspect of stroke recovery, into a global statistic that provides a more complete map of outcome. Furthermore, use of the global statistic often improves study power. By combining several different correlated measurements of the same trial population and the same underlying health outcome state, study power is enhanced through the reduction of noise generated by random fluctuations along any single outcome scale.

Nevertheless, the global statistic has drawbacks. Statistical methods for applying the global statistic over all levels of several ordinal scales, rather than only dichotomized end points, are still in development.[12] As a result, acute stroke trialists have combined only multiple dichotomized end points in global statistic analyses, from the landmark NINDS recombinant tissue plasminogen activator trials to the recent IMAGES trial. The resulting hyperdichotomized end point has many of the disadvantages of standard dichotomized end points, including discarding much of the outcome data and failing to reflect beneficial or harmful transitions over the

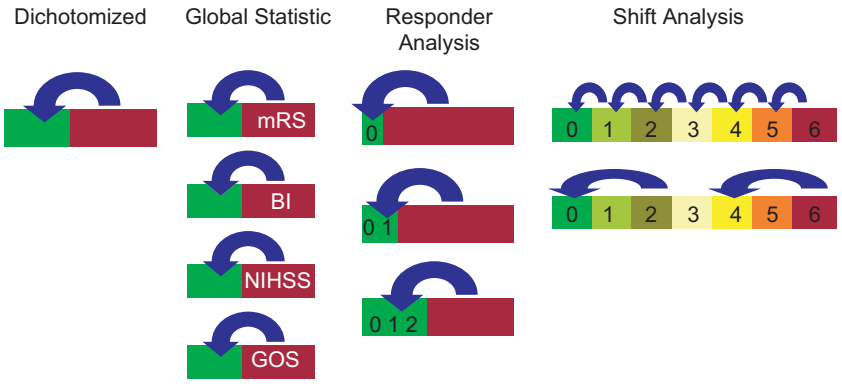

Figure 2. Four approaches to end point analysis. Column 1 depicts standard dichotomized analysis, counting only transitions across a single scale threshold. Column 2 shows global statistic analysis, counting transitions on several different outcome scales, eg, mRS, Barthel Index, NIHSS, and Glasgow Outcome Scale. Column 3 depicts sliding dichotomy responder analysis. Patients with mild deficits at study entry must attain level 0 to count as responders; patients with moderate deficits ≤level 1, and patients with severe deficits ≤level 2. Column 4 shows shift analysis, with transitions across all levels of the scale counted (©UCLA Stroke Program).

entire outcome range. When component scales are dichotomized at cut points identifying extreme good outcomes, trial assessment may be unduly driven by responses seen only in mildly affected patients.[13] Perhaps the greatest disadvantage of the global statistic is that it is difficult to express the results in terms intuitively meaningful to the patient and clinician at bedside. Techniques for translating multidimensional vector effects at the population level into benefit or harm to be expected in the individual patient remain undeveloped. For this reason, in the past the Food and Drug Administration and other regulatory agencies have discouraged the use of the global statistic as the primary end point in pivotal acute trials.

Responder Analysis

Responder analysis, also known as baseline severity-adjusted analysis, adjusts outcome thresholds according to stroke severity at study entry.[14,15] Although responder analysis can be applied across several ranks, to date only a sliding dichotomy analysis strategy has been deployed in acute stroke trials.[16–18] If a patient has a mild deficit at study entry, then only an excellent final outcome is considered a positive treatment response; for a patient with a moderate deficit at entry, a good outcome is judged as positive; and for a patient with a severe deficit at entry, a fair final outcome is considered positive. Responder analysis asks the question, "Does the treatment make patients better, considering where they started?"

Responder analysis has the advantage of analyzing an achievable goal for each patient rather than a fixed uniform outcome that would be inappropriate for very mild or severe stroke patients. As a result, it tends to increase detection of true signals occurring in the data set without increasing noise, thereby improving study power. However, like standard dichotomized assessments, responder analysis collapses outcome categories into binary or other reduced categories, discarding much outcome information and ignoring important disease state transitions. Another substantial drawback is that for optimal calibration of the "slide" (demarcating the entry severity categories and outcome thresholds), responder analysis requires that the investigators have foreknowledge of where along the outcome continuum the study treatment is most likely to alter patient status.

Shift Analysis

Shift analysis gauges change in outcome distributions over the full (or nearly full) range of ascertained outcomes. Also known as analysis of distributions, rank analysis, and analysis over levels, shift analysis asks the question, "Does the treatment make the patient better to some degree?" Several recent National Institute of Health, academic, and industry acute stroke clinical trials have used shift analysis.[19–21]

Shift analysis has several distinct advantages compared with simple dichotomization. It is irreducibly polychotomous, and consequently incorporates benefits and harm seen at all disease state transitions. Like responder analysis, shift analysis evaluates achievable goals for every patient. In addition, shift analysis inspects a greater number of achievable goals than sliding dichotomy implementations of responder analysis and does not require investigator foreknowledge of where in the spectrum of stroke severities an experimental treatment is most likely to work. A general advantage of shift analysis compared with other approaches is that it makes the least assumptions about the type of patients who will end up enrolled in a trial, the type of outcomes they will experience, and the treatment effect pattern the study agent will exert. Because it detects all important disease state transitions, shift analysis heightens the recognition of true signal and, therefore, in many situations, can increase study power. Shift analysis will especially enhance trial efficiency compared with collapsed analyses when treatment effects occur at all severities or at unpredictable state transitions.

Shift analysis does have disadvantages, primarily its computational complexity. Although modern statistical software can readily handle the appropriate ranked calculations, investigators may not be fully adept at projecting the population distributions of outcomes required for sample size and power calculations using shift analysis. If polychotomization is carried to the extreme, unimportant differences in outcome may enter into the analysis, resulting in detection of treatment effects so small as to be clinically insignificant[22,23] or in nondifferential misclassification that may reduce analytical power.[24] These problems likely arise when the full 42-level NIHSS is used as an outcome measure; they are less of an issue when using scales of 7 levels or fewer levels, such as the mRS or the Glasgow Outcome Scale. Shift analysis may be less efficient than collapsed analyses when treatment effects cluster at one state transition. Another historical drawback of shift analysis is difficulty expressing the treatment effect measured in terms meaningful to patients and clinicians.

Extensive evidence from both model and actual stroke trial populations show that shift analysis often improves study power to detect a true treatment effect. In the extreme case of a model population in which a drug was postulated to improve the outcome of every enrolled patient to the same small degree, Berge and Barer[25] showed that, compared with dichotomized analysis, shift analysis reduced by >60% the sample size required to demonstrate the true treatment effect. The effect of shift analysis in actual stroke trial populations has been analyzed by the Optimizing Acute Stroke Trials Collaboration.[26] The Optimizing Acute Stroke Trials group collected data from 55 treatment comparison in trials of acute stroke or stroke rehabilitation interventions that the investigators felt likely exerted a biological effect, either beneficial or harmful. The collaborators compared various modes of analysis of final end points, including standard dichotomized analysis, trichotomized analysis, and a variety of forms of ranked analysis. Ordinal shift analyses consistently outperformed dichotomized and other collapsed approaches. These findings confirm empirically the theoretical expectation that for a disease like stroke, which causes a range of functional impairments to develop, shift analysis will often be a more powerful mode of end point analysis than standard dichotomization. Stroke trialists have indeed long been handicapping their studies by use of an inefficient statistical mode of primary end point analysis.

Clinical Interpretation of Shift Analyses

The perceived difficulty in translating ranked analyses into meaningful terms at the bedside was the primary reason for

Table 2. PROACT 2 Trial Number

Rankin Transition	NNT, net	
Dichotomized		
0 vs 1–5	100	
0–1 vs 2–5	11.1	
0–2 vs 3–5	6.7	
0–3 vs 4–6	20	
0–4 vs 5–6	Infinity	
0–5 vs 6	50	
	NNT	NNH
Across all transitions of 7-level Rankin scale		
Minimum possible	2.6	25
Maximum possible	5.9	25
Biologically most plausible	4.8	28.4

Needed-to-treat values for different 6-level Rankin transitions.

accepting this statistical handicap in previous studies. Translating standard dichotomized end points into NNT terms is statistically straightforward. The NNT is the inverse of the absolute risk difference in the rate of dichotomized outcome between treated and control groups.[27,28] It is important to note, however, that this straightforward derivation is somewhat illusory. It provides only a "net" NNT, conflating in one number patients who improve and patients who worsen as a result of therapy, rather than providing disambiguated values for number needed to treat for benefit and number needed to treat for harm (NNH). Even more importantly, it provides a net NNT for only 1, somewhat arbitrarily selected, privileged health state transition.

Standard approaches to assessing the clinical relevance of the outcome scale treatment effect observed in acute stroke clinical trials include stating the mean difference in outcome scale scores and providing net NNT values for selected dichotomized cut points in the outcome scale. Both are problematic. Simply stating the treatment effect in terms of mean scale value differences does not provide intuitive understanding of the magnitude of benefit. For example, consider the PROACT 2 trial. In PROACT 2, treatment with intra-arterial prourokinase improved the mean mRS score from 3.59 to 3.26. If a physician is able to state, "This drug will improve your outcome on the mRS by an average of 0.33 points," how are the patient, family, the payor, the clinician, and society to use this knowledge in treatment decisions and health policy?

Table 2 shows the NNTs for the PROACT 2 Trial for different dichotomizations of the mRS. Each of the dichotomized NNTs provides only an incomplete index of treatment impact and the values range widely depending on the breakpoint used (6.7 to infinity). Proponents of a therapy may highlight only the most favorable NNTs[21] and opponents only the least favorable, a reporting emphasis that might be termed "impact bias." More importantly, all of these dichotomized NNTs underestimate the treatment effect of most relevance to decision-makers, which is benefit over the full range of outcomes rather than just at a single, arbitrarily selected, health state transition.

Because of the need to express a shift treatment effects distributed over a range of outcomes in a manner intuitively meaningful to patients and clinicians, single NNT and NNH values that incorporate health state transitions across the entire range of outcomes are highly desirable. However, deriving NNT values over the entire range of outcomes from randomized clinical trials requires a novel approach to NNT calculation that overcomes 2 statistical barriers.

The first barrier is scale nonlinearity. Nonlinearity refers to the fact that transitions at different points on ordinal scales like the mRS are not necessarily equal in value. A 1-point change from 6 to 5 (dead to persistent vegetative state) is qualitatively different and almost certainly quantitatively different from the 1-point change from 1 to 0 (symptoms without functional impairment to no symptoms at all). However, a useful NNT measure can be derived from noncomparable health state transitions, as long as each transition that is being counted is clinically worthwhile for the patient to attain.

Some transitions are not considered desirable by many individuals. For example, among healthy, elderly Americans, 47% consider a stroke outcome of surviving alive but severely disabled as worse, not better, than death.[29] For these individuals, transitioning from an mRS of 6 to 5 is not a better outcome, it is worse. For this reason, it sometimes will make sense to collapse mRS 5 and 6 categories into a single worst outcome group and not count transitions from 6 to 5 as improvements. In contrast, virtually all other single-point transitions on the mRS appear to be clinically valuable. Once nonvaluable transitions are discarded from the analysis, a value can be calculated for the NNT for 1 patient to improve by $\geq$1 clinically worthwhile steps.

The second statistical barrier, stated technically, is that derivation of an NNT across an ordinal scale requires knowledge of the within-patient correlation, and these data are not specified by parallel group clinical trials, but only by crossover design trials.[10,30,31] In nontechnical terms (Figure 3), the challenge is that trial data do not specify if group differences in the distribution of outcomes occur because many individuals each improve a little bit, a few individuals each improve a lot, or a mixture of these outcomes.

This problem can be overcome, however, by using the method of joint outcome table specification.[10] This method permits derivation of: (1) the trial data specified minimum possible value for the NNT; (2) the trial data specified maximum possible value for the NNT; (3) individual expert judgment of the biologically most plausible NNT; and (4) individual expert judgment of the biologically most plausible NNH.

Trial data fully specify the minimum and maximum possible values for the NNT across an ordinal scale, if a treatment does not cause harm or causes a known pattern of harm. By completing joint outcome tables under the extreme assumption that every patient who improves does so by only one step, the lowest possible NNT compatible with the trial data can be derived. By completing joint outcome tables under the extreme assumption that every patient who improves does so by the largest number of steps compatible with

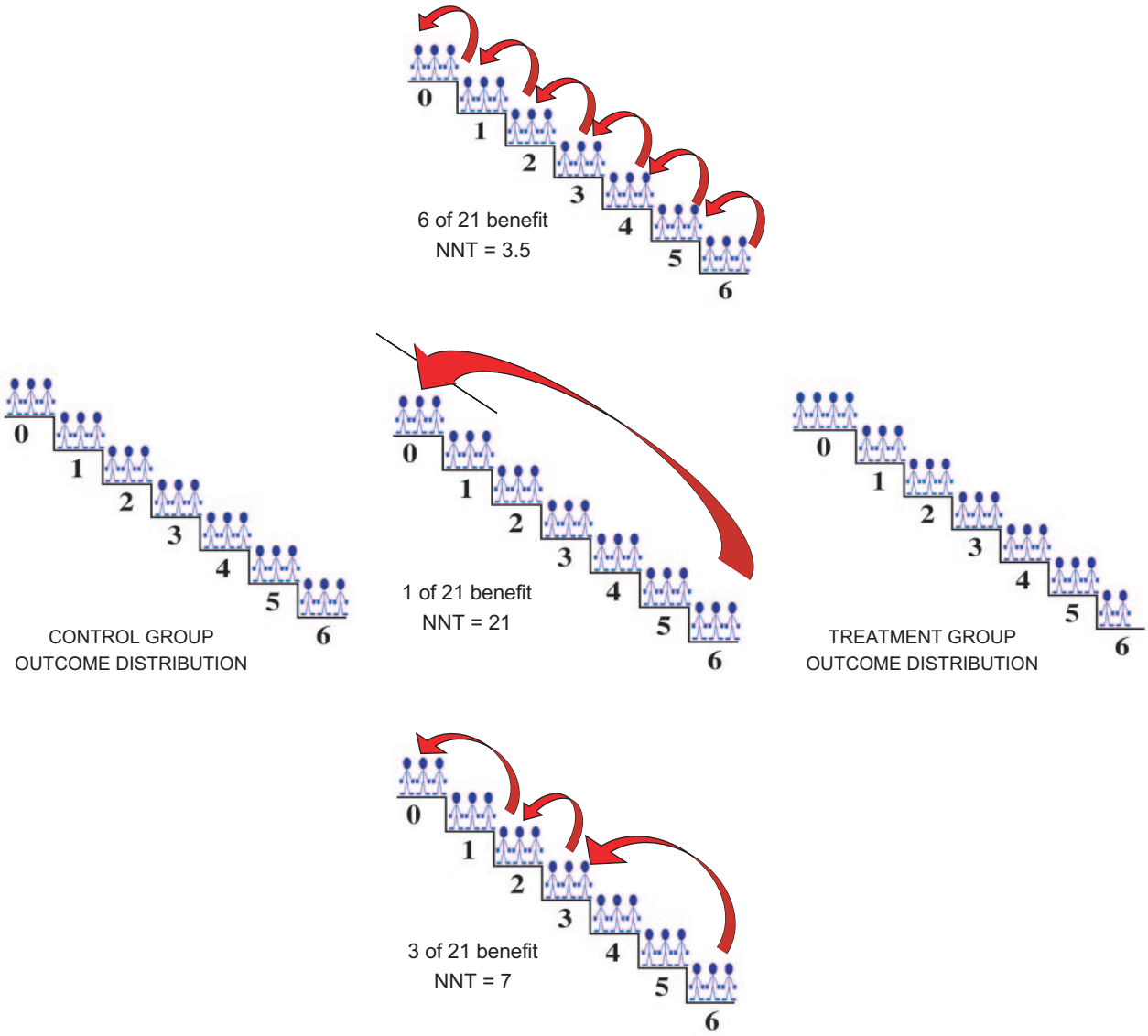

Figure 3. NNT not fully specified by parallel group trial data. Leftmost image row shows results from a hypothetical randomized controlled trial in which 21 patients in the control group are equally distributed at each of the 7 levels of the primary end point scale, whereas rightmost image row shows that in the treatment group 4 patients are at level 0, 2 at level 6, and 3 at each other level. These group data do fully specify the NNT, because there are several possible patterns of treatment response that could yield the final distribution, as depicted in the 3 central images. The maximum possible NNT compatible with the data occurs when 1 patient at each level improves by 1 step (many patients each improving by a little bit, center top image), the minimum possible NNT when 1 patient at level 6 improves to level 0 (1 patient improving by a lot, center image), and a biologically more likely NNT when a blend of small and large improvements occurs (center bottom image) (© UCLA Stroke Program).

the final group distribution, the highest possible NNT compatible with the trial data can be derived.[32,33]

Moreover, having disease experts specify the biologically most reasonable distribution of responses specifies within-patient correlation and permits derivation of the biologically most plausible NNT and NNH, reflecting benefits and adverse effects accrued across the entire outcome range. Averaging the NNTs and NNHs independently derived from multiple experts avoids idiosyncratic perspectives and permits derivation of benefit and harm values that represent the best insight from the field.[10,32] This method allows derivation of the number needed to treat for 1 additional patient to benefit by improving by ≥1 grades of global disability on the mRS.

Tables 2 and 3 demonstrate the results that arise from this analytic method. Table 2 shows that the PROACT 2 trial data indicate the lowest possible NNT for 1 patient to benefit by improving by ≥1 grades of global disability on the mRS is 2.6, and the highest possible is 5.9. The expert-derived most plausible NNT is 4.8 (95% CI, 4.1 to 5.5). For every 100 middle cerebral artery occlusion patients treated within 6 hours of onset with intra-arterial pro-urokinase, 21 improve by ≥1 grades on the mRS global disability scale and 4 worsen by ≥1 grades. Table 3 shows the NNT and NNH across the 6-level version of the mRS for 5 acute stroke interventions with positive phase 3 trials, recombinant tissue plasminogen activator, intra-arterial pro-urokinase, and NXY-059 for acute cerebral

Table 3. Treatment Effect Summary Measures for 5 Positive Acute Stroke Studies

Treatment	Source Trials	Mean mRS Difference	Net NNT 0–1 vs 2–6	Net NNT 0–2 vs 3–6	NNH Across 6-Level Scale	NNT Across 6-Level Scale
Intravenous TPA	NINDS Trials 1 and 2	0.53	6.2	8.4	56.6	3.3
Intra-arterial ProUK	PROACT 2	0.32	11.1	6.7	28.4	4.8
Coiling in SAH	ISAT	0.26	16.7	14.3	NH	5.9
NXY-059	SAINT 1	0.13	41.7	47.6	NH	9.8
		Mean GOS Difference	Net NNT 5 vs 1–4	Net NNT 4–5 vs 1–3	NNH Across Full 5-Level Scale	NNT Across Full 5-Level Scale
Nimodipine in SAH	British Nimodipine Trial	0.37	9.6	7.6	NH	6.8

GOS indicates Glasgow Outcome Scale; ProUK, pro-urokinase; TPA, tissue plasminogen activator; SAH, subarachnoid hemorrhage; NH, no harm estimated.

ischemia, and nimodipine and coiling (versus clipping) for aneurysmal subarachnoid hemorrhage.

Good judgment comes from experience. Experience comes from bad judgment. One consolation of the many nonpositive stroke trials performed to date is the hope that acute stroke trialists have at last failed often enough to gain sufficient wisdom to devise trials that will be successful. One of the most important lessons we have learned is that we can design and clinically interpret trials that employ primary end point analyses sensitive to treatment effects over the entire range of outcomes. For any individual trial, the selection of primary end point analysis technique should be guided by the study population, expected treatment response, and study purpose; often, though not always, shift analysis will be advantageous. Shift analysis is an important advance in the interpretation, as well the design, of clinical trials, and generally provides the most comprehensive single value summary of a treatment's clinical impact. As clinicians and as researchers, we seek to help patients afflicted by a disease that cripples as well as kills, and we are deeply interested in detecting all treatment benefits that matter to our patients and their families.

Acknowledgments

The author thanks fellow members of the pro-urokinase NNT and NNH derivation expert panel: Gary Duckwiler, MD, Reza Jahan, MD, David Liebeskind, MD, Sidney Starkman, MD, and Fernando Vinuela, MD.

Sources of Funding

This work was supported in part by NIH-NINDS Awards U01 NS 44364 and P50 NS044378.

Disclosures

Scientific Advisory Boards and Scientific Consulting from: Ambit Biosciences, Quintiles, Nuvelo, Co-Axia, Astellas, Novo Nordisk, Astra Zeneca, FibroGen, Boehringer Ingelheim (secondary prevention), Talacris. Trial site subinvestigator: Takeda, Eli Lilly, Astra Zeneca, Neurobiological Technologies. Trial investigator in NIH supported studies with drug or devices supplied by the following-companies: CLEAR Trial, IMS 1–3 Trials, MR RESCUE Trial. Speaker's Bureau or Speaking Honoraria received from Boehringer Ingelheim (secondary prevention), Concentric Medical.

References

1. Kidwell CS, Liebeskind DS, Starkman S, Saver JL. Trends in acute ischemic stroke trials through the 20th century. *Stroke*. 2001;32:1349–1359.
2. Grotta J. Neuroprotection is unlikely to be effective in humans using current trial designs. *Stroke*. 2002;33:306–307.
3. Ovbiagele B, Kidwell CS, Starkman S, Saver JL. Neuroprotective agents for the treatment of acute ischemic stroke. *Curr Neurol Neurosci Rep*. 2003;3:9–20.
4. DeGraba TJ, Pettigrew LC. Why do neuroprotective drugs work in animals but not humans? *Neurol Clin*. 2000;18:475–493.
5. Gorelick PB. Neuroprotection in acute ischaemic stroke: a tale of for whom the bell tolls? *Lancet*. 2000;355:1925–1926.
6. Stroke Therapy Academic Industry Roundtable II (STAIR-II). Recommendations for clinical trial evaluation of acute stroke therapies. *Stroke*. 2001;32:1598–1606.
7. Kasner SE. Clinical interpretation and use of stroke scales. *Lancet Neurol*. 2006;5:603–612.
8. Duncan PW, Samsa GP, Weinberger M, Goldstein LB, Bonito A, Witter DM, Enarson C, Matchar D. Health status of individuals with mild stroke. *Stroke*. 1997;28:740–745.
9. Lai SM, Duncan PW. Stroke recovery profile and the modified Rankin assessment. *Neuroepidemiology*. 2001;20:26–30.
10. Saver JL. Number needed to treat estimates incorporating effects over the entire range of clinical outcomes: novel derivation method and application to thrombolytic therapy for acute stroke. *Arch Neurol*. 2004;61:1066–1070.
11. Lu M, Tilley BC. Use of odds ratio or relative risk to measure a treatment effect in clinical trials with multiple correlated binary outcomes: data from the NINDS t-PA stroke trial. *Stat Med*. 2001;20:1891–1901.
12. Huang G-H, Bandeen-Roche K, Rubin GS. Building marginal models for multiple ordinal measurements. *Appl Statist*. 2002;51:37–57.
13. Thomassen L, Waje-Andreassen U, Naess H, Elvik MK, Russell D. Long-term effect of intravenous thrombolytic therapy in acute stroke: Responder analysis versus uniform analysis of excellent outcome. *Cerebrovasc Dis*. 2005;20:470–474.
14. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, 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.
15. Young FB, Lees KR, Weir CJ. Improving trial power through use of prognosis-adjusted end points. *Stroke*. 2005;36:597–601.
16. Adams HP Jr, Leclerc JR, Bluhmki E, Clarke W, Hansen MD, Hacke W. Measuring outcomes as a function of baseline severity of ischemic stroke. *Cerebrovasc Dis*. 2004;18:124–129.
17. Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope DT, Karimi A, Shaw MD, Barer DH. 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.
18. Saver JL, Yafeh B. Confirmation of TPA treatment effect by baseline severity-adjusted end point reanalysis of the NINDS-TPA stroke trials. *Stroke*. 2007;38:414–416.
19. Ehrenreich H, Hasselblatt M, Dembowski C, Cepek L, Lewczuk P, Stiefel M, Rustenbeck HH, Breiter N, Jacob S, Knerlich F, Bohn M, Poser W, Ruther E, Kochen M, Gefeller O, Gleiter C, Wessel TC, De Ryck M, Itri L, Prange H, Cerami A, Brines M, Siren AL. Erythropoietin therapy for acute stroke is both safe and beneficial. *Mol Med*. 2002;8:495–505.

20. The FAST-MAG Trialists. Field Administration of Stroke Therapy—Magnesium (FAST-MAG) phase 3 trial. Available at: www.fastmag.info. Accessed 2007.
21. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, Grotta J, Lyden P, Shuaib A, Hardemark HG, Wasiewski WW. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006;354:588–600.
22. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371–383.
23. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. *Med Decis Making*. 2005;25:250–261.
24. Rothman KJ, Greenland S. Precision and validity of studies. In: Rothman KJ, Greenland S, eds. *Modern epidemiology.* Philadelphia, PA: Lippincott Williams & Wilkins; 1998:115–134.
25. Berge E, Barer D. Could stroke trials be missing important treatment effects? *Cerebrovasc Dis*. 2002;13:73–75.
26. The Optimising Analysis of Stroke Trials (OAST) Collaboration. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. *Stroke*. 2007;38:1911–1915.
27. Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. *BMJ*. 1995;310:452–454.
28. McAlister FA, Straus SE, Guyatt GH, Haynes RB. Users' guides to the medical literature: Xx. Integrating research evidence with the care of the individual patient. Evidence-based medicine working group. *JAMA*. 2000;283:2829–2836.
29. Samsa G, Matchar D, Goldstein L, Bonito A, Duncan P, Lipscomb J, Enarson C, Witter D, Venus P, Paul J, Weinberger M. Utilities for major stroke: Results from a survey of preferences among persons at increased risk for stroke. *Am Heart J*. 1998;136:703–713.
30. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ*. 1998;316:690–693.
31. Walter SD. Number needed to treat (NNT): Estimation of a measure of clinical benefit. *Stat Med*. 2001;20:3947–3962.
32. Rajajee V, Vespa PM, Duckwiler G, Jahan R, Frazee J, JL S. Benefit from coiling in the ISAT trial: Reanalysis of the number needed to treat. *130th Annual Meeting of the American Neurological Association Abstract Book.* 2005:14.
33. Saver JL. Clinical impact of NXY-059 demonstrated in the SAINT 1 trial: derivation of number needed to treat for benefit over entire range of functional disability. *Stroke.* 2007;38:1515–1518.