Lees, K.R., Davalos, A., Davis, S.M., Diener, H.-C., Grotta, J., Lyden, P., Shuaib, A., Ashwood, T., Hardemark, H.G., Wasiewski, W., Emeribe, U. and Zivin, J.A. (2006) *Additional outcomes and subgroup analyses of NXY-059 for acute ischemic stroke in the SAINT I trial.* Stroke, 37 (12). pp. 2970-2978. ISSN 0039-2499

http://eprints.gla.ac.uk/21525/

Deposited on: 23 January 2012

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/37/12/2970

Additional Outcomes and Subgroup Analyses of NXY-059 for Acute Ischemic Stroke in the SAINT I Trial

Kennedy R. Lees, MD, FRCP; Antonio Davalos, MD; Stephen M. Davis, MD, FRACP;
Hans-Christoph Diener, MD; James Grotta, MD; Patrick Lyden, MD; Ashfaq Shuaib, MD;
Tim Ashwood, PhD; Hans-Goran Hardemark; Warren Wasiewski, MD; Ugochi Emeribe, PhD;
Justin A. Zivin; for the SAINT I Investigators

Background and Purpose—NXY-059 is a free radical-trapping neuroprotectant demonstrated to reduce disability from ischemic stroke. We conducted analyses on additional end points and sensitivity analyses to confirm our findings.

Methods—We randomized 1722 patients with acute ischemic stroke to a 72-hour infusion of placebo or intravenous NXY-059 within 6 hours of stroke onset. The primary outcome was disability at 90 days, as measured by the modified Rankin Scale (mRS), a 6-point scale ranging from 0 (no residual symptoms) to 5 (bed-bound, requiring constant care). Additional and exploratory analyses included mRS at 7 and 30 days; subgroup interactions with final mRS; assessments of activities of daily living by Barthel index; and National Institutes of Health Stroke Scale (NIHSS) neurological scores at 7 and 90 days.

Results—NXY-059 significantly improved the distribution of the mRS disability score compared with placebo at 7, 30, and 90 days (Cochran-Mantel-Haenszel test $P=0.002$, 0.004, 0.038, respectively; 90-day common odds ratio 1.20; 95% CI, 1.01 to 1.42). The benefit was not attributable to any specific baseline characteristic, stratification variable or subgroup interaction. Neurological scores were improved at 7 days (odds ratio [OR], 1.46; 95% CI, 1.13, 1.89; $P=0.003$) and the Barthel index was improved at 7 and 30 days (OR, 1.55; 95% CI, 1.22, 1.98; $P<0.0001$; OR, 1.27; 95% CI, 1.01, 1.59; $P=0.02$).

Conclusions—NXY-059 within 6 hours of acute ischemic stroke significantly reduced disability. Benefit on neurological scores and activities of daily living was detectable early but not significant at 90 days; however, our trial was underpowered to measure effects on the neurological examination. The benefit on disability is not confounded by interactions and is supported by other outcome measures. (***Stroke***. 2006;37:2970-2978.)

Key Words: acute care ■ drug trials ■ free radicals ■ ischemia ■ neuroprotection ■ NXY-059

Treatment for acute ischemic stroke improves outcome. Specialist stroke unit care reduces mortality and disability and is generally applicable. Aspirin is also generally applicable and offers a small benefit. Reperfusion strategies such as administration of alteplase (recombinant tissue plasminogen activator [rt-PA]) can offer substantial benefit to improve neurological function and reduce disability but are at present delivered to only a small proportion of stroke patients, and their use is associated with practical difficulties and a modest risk of hemorrhage.[1] We need a treatment that can protect brain tissue from ischemic damage until spontaneous or therapeutic reperfusion occurs, and that can be adminis-

tered safely and easily.[2] It should be suitable for use in a wide range of patients, because even modest benefits that can be delivered to a majority of patients will have greater public health impact than substantial benefit derived by a few. The search for such a neuroprotectant has been arduous and so far no therapy has been generally approved.

We have recently reported reduction in disability associated with use of the free radical-trapping neuroprotectant, NXY-059, in acute ischemic stroke: the Stroke-Acute Ischemic NXY Treatment (SAINT I) trial.[3] NXY-059 is a well-tolerated agent that is believed to act through interruption of the ischemic cascade at the level of free radical induced injury.[4–6] This is a

late stage in the process of both permanent and transient ischemia that occurs in brain and vascular endothelium cells.

The SAINT I trial demonstrated that NXY-059 treatment was associated with a significant reduction in global disability, achieved without affecting mortality or causing obvious safety problems.[3] Disability was first in a prespecified hierarchy of end points; the trial was not statistically significant on the end points lower in the hierarchy. A closer examination is justified of the effects on disability and neurological status, and on possible sources of interaction or confounding influence. Here we report detailed results of exploratory analyses of the trial because these demonstrate that the secondary outcome measures (scores of neurological function and activities of daily living) were strongly positive when measured within weeks of study treatment, that they are supportive of the disability findings, and that the disability findings themselves show consistency across a range of subgroups.

Methods

The study design and patient population have been described elsewhere.[3] Briefly, this was a double-blind, randomized, and placebo-controlled trial that assessed the effect of a 72-hour infusion of NXY-059 on disability and neurological outcomes in patients treated within 6 hours of an acute ischemic stroke. Patients otherwise received standard care for acute stroke. We stratified randomization for country, National Institutes of Health Stroke Scale[7] (NIHSS) score at baseline, side of cerebral infarct (ie, left or right hemisphere), and intention to cotreat with alteplase using a dynamic algorithm to maintain balance of prognostic factors between treatment groups.[8] The randomization scheme also forced allocation of patients so that the average time to treatment at each center was not >4 hours at any time during enrolment. The trial had appropriate regulatory and ethical committee approval in all countries and patients or their representatives gave informed consent to participation. The steering committee developed the trial protocol, approved the statistical plan, had full access to the data, wrote the manuscript, and was responsible for decisions regarding publication. The academic authors vouch for the veracity and completeness of the data and data analyses.

Outcome Measures

We assessed patients at enrolment and recorded outcomes after 7, 30, and 90 days, taking the last rating for our primary outcome, which was based on the modified Rankin Scale (mRS).[9] The mRS is a disability scale that awards scores in 6 categories ranging from 0 (no residual symptoms) through 5 (bed-bound, requiring constant nursing care). Investigators were trained, tested, and certified in use of mRS, using a DVD method developed specifically in preparation for this trial. Assessments of stroke severity at baseline, 7 days, and 90 days were based on a neurological rating scale, the NIHSS, which quantifies the neurological examination, and were conducted by observers trained and certified in its use.[7] NIHSS scores range from 0 (normal) to 42 (most severe). A hierarchy of additional outcome measures was prespecified after the mRS,[3] starting with the NIHSS and then Barthel index (BI).[10] The BI is an activities of daily living scale that awards scores in increments of 5 ranging from 0 (totally dependent) through 100 (no help required for activities of daily living).

Statistical Analysis

The analyses were prespecified and followed the intention-to-treat principle: all patients are included within this analysis who were randomized, received any study infusion, and had any follow-up assessment performed. Our primary end point was mRS at 90 days or last rating, analyzed across the whole distribution of scores by the

Cochran-Mantel-Haenszel test (CMH), adjusted for stratification variables (NIHSS, side of infarct and alteplase use), and using modified ridit scores, ie, midrank/(number of observations+1), to account for ordered categories.[11] A significant benefit on this end point was considered sufficient for us to declare a positive trial.[3,12] Odds ratios were calculated by proportional odds logistic regression, which assumes a common odds ratio across all cut-points of the mRS, but were used only as an estimate of treatment effect. The odds ratios were adjusted in the same way as for the primary end point. Whereas the odds ratio gives an estimate of treatment effect, the CMH test gives a more accurate probability value. For all analyses, deaths were included within the worst outcome category (mRS, 5; NIHSS, 42; BI, 0) and last available rating was used to replace any missing values in survivors. Patients who had recurrent stroke during follow-up were included in our analysis, even though disability from the index stroke will thus have been overestimated. The sample size was chosen to have 90% power to detect a common odds ratio of 1.3.

The change from baseline in the NIHSS score at 90 days ("change score") was second in our hierarchy of tests and termed coprimary in the study protocol; joint testing with mRS was not proposed.[3] It was analyzed using analysis of covariance adjusted for baseline variables of NIHSS, side of infarct and use of alteplase; in calculating the change from baseline, deaths were awarded the worst score, 42. Because the NIHSS analysis by ANCOVA was compromised through failure to meet statistical assumptions over distributions of changes in score (essentially, the distribution was not Gaussian, being both skewed and bimodal; Figure 1), we have here used exploratory analyses of final NIHSS score using the approach prespecified for our primary end point and the same method of interpretation as reported in other recent acute stroke trials.[13,14] For overall assessment of effect on NIHSS, we have used the CMH test, adjusted in the same way as for mRS; for estimation of odds ratios, we divided the final NIHSS scores into 2 groups: one with scores of 0 to 1, representing patients with excellent recovery, and one with scores of 2 to 42 (with deaths awarded 42), representing patients with a worse outcome.

The third outcome in the hierarchy was BI (dichotomized into 0 to 90 versus 95 to 100) at 90 days.[3] Because the planned analysis of NIHSS was not significant, analysis of the BI was exploratory. CMH and logistic regression testing were used and all probability values are nominal.

Power Calculations Based on NIHSS

The NIHSS was included within the SAINT I trial protocol as an end point without first examining statistical properties or undertaking formal power calculations. The planned analysis of NIHSS was nonsignificant and the approach clearly required review to inform
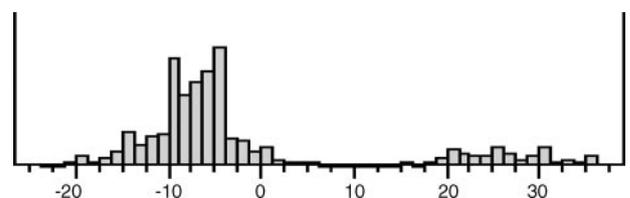
Figure 1. Distribution of change scores on the NIHSS, based on the prespecified approach to analysis, final NIHSS score recorded at 90 days, or last rating was subtracted from the baseline score to generate a "change score." For patients who had died, a final score of 42 (worst category) was substituted. Thus, for example, if a typical patient may have entered the trial with an NIHSS score of 12, and improved by 90 days to a score of 5, then the "change score" would then be −7. However, if that patient had died before 90 days, the assigned final score would be 42, and the "change score" would be +30. Ninety-day mortality in the trial was 17%.[3] A basic assumption for the planned analysis of covariance is that the change scores would be normally distributed; instead, the Figure reveals that these change scores follow a bimodal and skewed distribution. Nonparametric analysis is essential (see text).

future trials, such as the sister trial, SAINT II. As part of our exploratory analysis of the trial results, we have therefore considered the power of various methods of analyzing NIHSS under various sample sizes.

Because the NIHSS data fail to meet assumptions required for parametric analysis, a power assessment based on ANCOVA would be inappropriate and power calculations reported here are based on a nonparametric approach using the CMH test. The model on which the CMH test was calculated included the following variables for analysis of NIHSS final and change scores: treatment (2 levels), side of infarct (2 levels), stratified NIHSS baseline score (4 levels), and intent-to-treat or treatment with alteplase (2 levels). Using our trial data, a bootstrap method was used randomly to draw 5000 times with replacement "n" patients from the NXY-059 and placebo groups. The resulting CMH test for NIHSS final and change scores and for the dichotomization approach were calculated for each random sample. The power is defined as the percentage of CMH tests, which resulted in $P \leq 0.05$, ie, as the sensitivity of the test to deliver a result that reaches statistical significance at $P < 0.05$ when there exists a genuine difference of a defined magnitude between treatments.

Results

We randomized 1722 patients, and mRS outcome data were available for 1699 of the 1705 patients (99.6%) who received any treatment: 850 were allocated to NXY-059 and 849 to placebo. Mean age was 68.4 years and 947 (55.5%) were male. Mean time from onset of stroke to treatment was 3 hours 46 minutes and 28.7% received alteplase. The treatment groups were well balanced for baseline prognostic variables.[3]

Outcome Measures

Outcomes according to mRS, NIHSS, and BI showed clear benefit from NXY-059 treatment at 7 days with benefit still present on BI and mRS at 30 days and on mRS by 90 days (Figure 2). The benefits of treatment with NXY-059 demonstrated on mRS at 90 days/last follow-up were consistently noted across subgroups. Alternative methods of analysis for mRS, dichotomized at the various possible thresholds (0 versus 1 to 5; 0 to 1 versus 2 to 5, etc) or categorized into 3 groups (0 to 1 versus 2 to 3 versus 4 to 5) were internally consistent, reaching significance for complete cure (mRS, 0; $P = 0.003$), recovery of ambulation (mRS, ≤ 3; $P = 0.02$), and for the trichotomy ($P = 0.03$). Neither the mortality nor the proportion of patients who were severely disabled (mRS 5/death) was altered by treatment with NXY-059.[3] There was no significant effect on BI at last rating either on dichotomization (95 to 100 versus 0 to 90; 60 to 100 versus 0 to 55) or by trichotomy (95 to 100 versus 60 to 90 versus 0 to 55). Early outcomes were strongly positive, however. Point estimates of odds ratios for efficacy according to all analyses of functional measures showed similar trends toward benefit and their CIs overlap widely: all domains of outcome improved to a similar extent.

Stratification Variables

As expected, outcomes varied among strata (not shown), but there was no interaction between the treatment effect of NXY-059 and the stratification variables (Figure 3).

Baseline Characteristics

None of the baseline or medical history variables evaluated showed apparent evidence of interaction with the treatment

effect of NXY-059 (Figures 4 and 5). Nominal probability values were <0.05 for diastolic blood pressure and ischemic heart disease; however, there was no adjustment for multiplicity, although at least 25 such tests were conducted, and no biological rationale had been proposed for any interaction between treatment and these effects.

Power Calculations Based on NIHSS

When the generalized CMH test with modified ridit scores was applied we found that a trial of 1700 patients would have power of only 16% to detect the difference between the treatment groups using the change in NIHSS from baseline score versus power of 35% to detect the observed difference in final NIHSS. Using dichotomization of final NIHSS scores at 0 to 1 versus 2 to 42, power lies between these values, at 24% (Table).

Discussion

SAINT I is the first trial of a putative neuroprotectant in acute ischemic stroke to report a positive result on its primary end point. Interpretation of these findings will be greatly assisted when the twin trial, SAINT II, reports next year. Meantime, we have closely examined our results seeking evidence of confounding factors or subgroups that could be responsible for a false-positive finding, and/or that may explain the apparent discrepancy between the primary end point (mRS at 90 days) and the prespecified additional end points (NIHSS and BI).

In essence, we have not identified any interaction between treatment effect and either baseline variables or medical history that could undermine our result. Not only was our trial large enough for randomization to deliver reasonable balance on prognostic factors, duly confirmed by our finding that the variables that we measured were evenly distributed between treatment groups,[3] but also the formal tests for interaction between treatment effect and these variables were uniformly reassuring, ie, they have raised no evidence that any residual imbalance in baseline variables influenced the study findings. To assist with interpretation of interaction tests around an overall odds ratio of 1.2 for treatment effect, it should be noted that a study of this size will have 80% power to detect an interaction when the odds ratio in one subgroup is at least 50% higher than the other.

Second, we found no subgroup interaction related to our stratification variables that either would undermine our result or suggest that we had included an unreasonably heterogeneous population. The treatment benefit is present regardless of use of alteplase, side of stroke, or initial stroke severity. It was also independent of time from stroke onset to treatment, which we had rigorously controlled. The lack of interaction with either rt-PA use or time to treatment deserves comment. Certainly the experimental data support efficacy of NXY-059 across the 6-hour time window.[6] Less time-dependency than reported for alteplase[1] may therefore be expected in the SAINT trial. The limited power for the time-dependency analysis and the preponderance of patients (two-thirds) being treated within the narrow 3- to 5-hour delay also need to be considered. It would be premature to conclude that the benefits of treatment with NXY-059 are not time-dependent
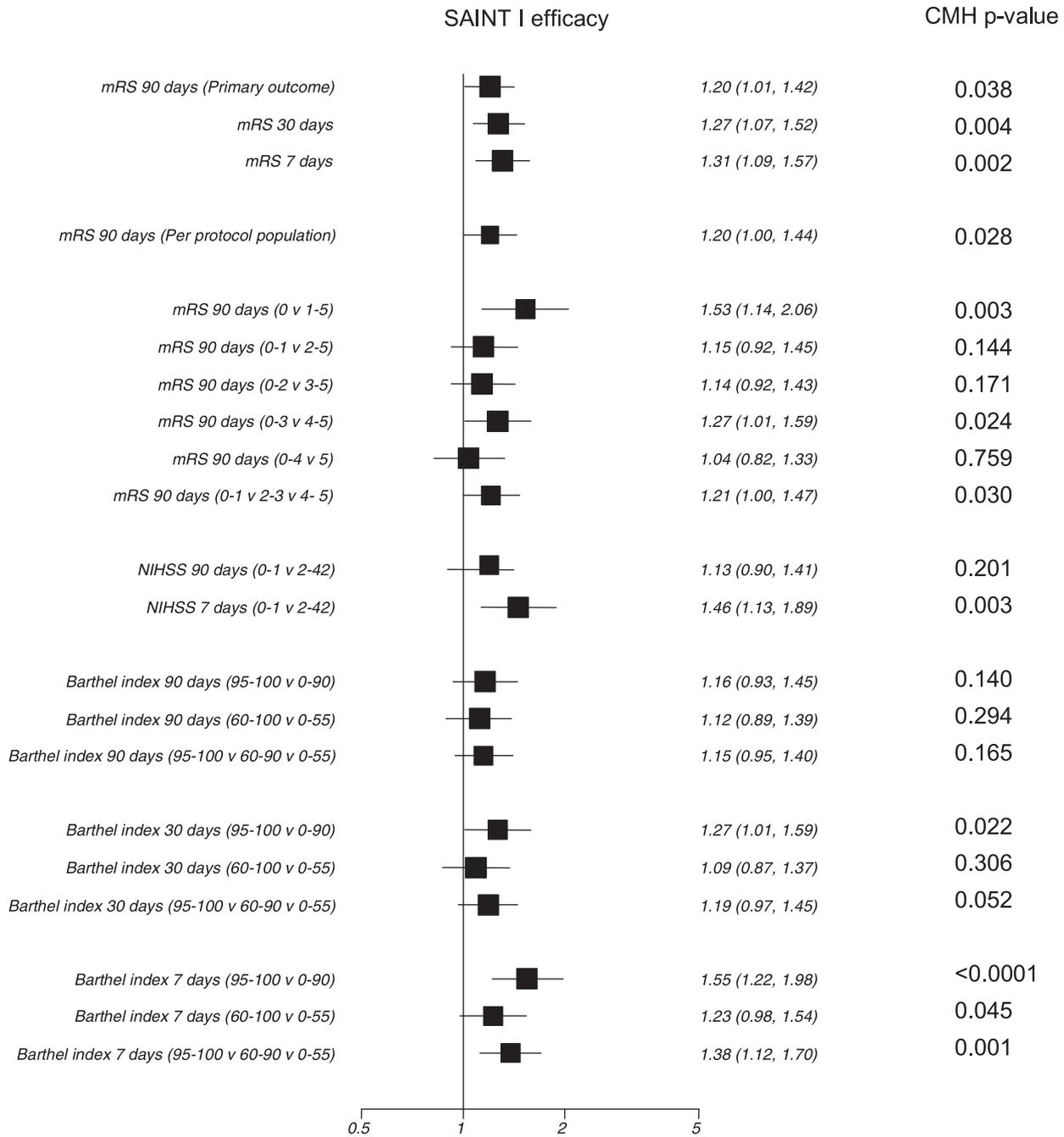
Figure 2. Efficacy outcomes for the SAINT I trial. The primary outcome measure was mRS at 90 days or last rating, as analyzed by the CMH test with modified ridit score, adjusted for baseline stratification variables of stroke severity (NIHSS in 4 categories), side of infarct (right or left hemisphere), and use or intended use of tissue plasminogen activator. The CMH test considers the full range of the mRS, with deaths included at mRS 5. Odds ratios were estimated by logistic regression, adjusting in the same way as for the primary end point, and are shown with their estimated 95% CIs, along with the probability value from the CMH test; however, because logistic regression uses a different algorithm to the CMH test, it is possible for minor discrepancies to occur between the CMH probability value and the 95% CIs. The CMH probability value is the primary measure of significance. Only 4 of the analyses of mRS by dichotomization were prespecified as secondary measures (excluding 0 to 4 versus 5); trichotomization 0 to 1 versus 2 to 3 versus 4 to 5 was also prespecified. Analysis of mRS at 7 and 30 days was prespecified. The analyses of final NIHSS shown in the Figure were conducted post hoc, after the planned approach to examine change of NIHSS from baseline was considered inappropriate. NIHSS was not recorded at 30 days. The prespecified analysis of BI was to examine the dichotomization at 95 to 100 versus 0 to 90 at 90 days or last rating; however, because the planned NIHSS analysis that lay higher in the statistical hierarchy of end points was not significant, all analyses of BI are considered exploratory. For purposes of all analyses shown, patients who died were awarded the worst score on the relevant scale. All analyses are based on the intent-to-treat principle.

within the 6-hour window studied. Regarding use alongside alteplase, the experimental data indicate that NXY-059 can provide at least as good efficacy in the absence or presence of reperfusion, ie, in a permanent focal ischemia model as in a

transient ischemia model.[6] The subgroup analyses in SAINT I are fully compatible with similar benefits being achieved with NXY-059 treatment whether administered to patients also treated with alteplase or given as monotherapy. Firmer
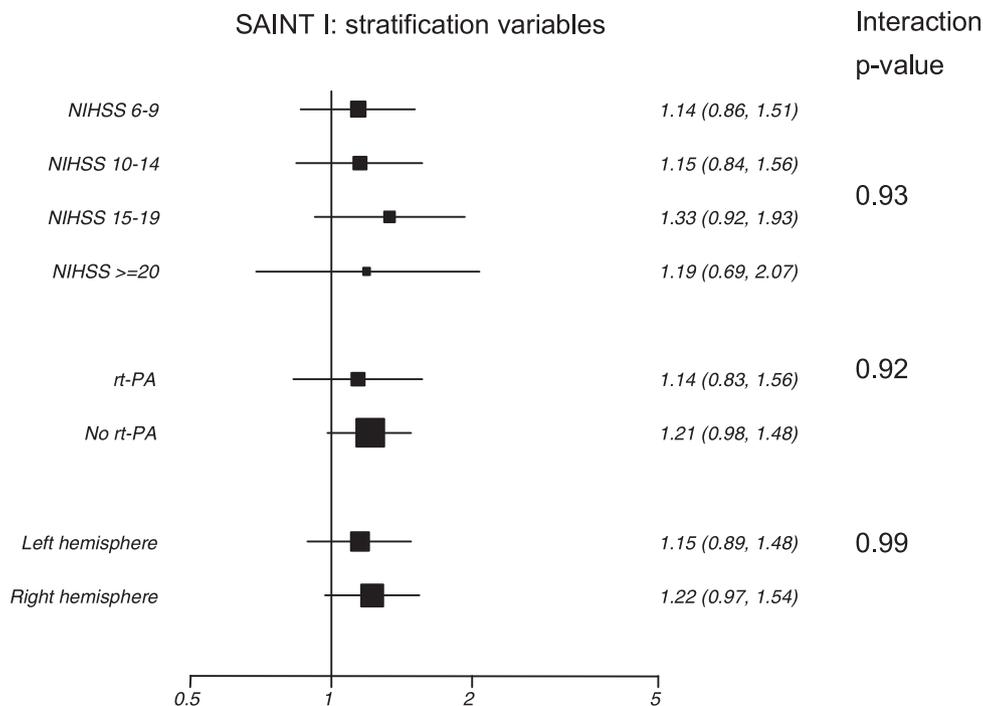
SAINT I: stratification variables

Interaction

p-value

NIHSS 6-9	1.14 (0.86, 1.51)
NIHSS 10-14	1.15 (0.84, 1.56)
NIHSS 15-19	1.33 (0.92, 1.93)
NIHSS >=20	1.19 (0.69, 2.07)

0.93

rt-PA	1.14 (0.83, 1.56)
No rt-PA	1.21 (0.98, 1.48)

0.92

Left hemisphere	1.15 (0.89, 1.48)
Right hemisphere	1.22 (0.97, 1.54)

0.99

0.5   1   2   5

Figure 3. Efficacy outcomes of the SAINT I trial according to stratification variables. The primary outcome measure was mRS at 90 days or last rating, as analyzed by the CMH test with modified ridit score, adjusted for baseline stratification variables of stroke severity (NIHSS in 4 categories), side of infarct (right or left hemisphere), and use or intended use of tissue plasminogen activator. The CMH test considers the full range of the mRS, with deaths included at mRS 5. Odds ratios were estimated by logistic regression, adjusting in the same way as for the primary end point, and are shown with their estimated 95% CIs, along with the probability value from the CMH test; however, because logistic regression uses a different algorithm to the CMH test, it is possible for minor discrepancies to occur between the CMH probability value and the 95% CIs. The CMH probability value is the primary measure of significance. Interactions between the prognostic variables used in the stratification algorithm and the treatment effect were tested using the CMH test, and the probability value for each of these is shown in the Figure. Because none of these was significant, the individual odds ratios and their 95% confidence limits are provided only for illustration. A more reliable estimate of treatment effect is obtained from the primary outcome in Figure 2. For purposes of all analyses shown, patients who died were awarded the worst score on the relevant scale. All analyses are based on the intent-to-treat principle.

conclusions need to await the planned pooled analyses of the SAINT trials after the completion of SAINT II, which will have much greater power to detect differential response to NXY-059 dependent on time to treatment or alteplase usage.

Third, we see consistency in effect sizes and directions among 3 measures of stroke outcome: disability, neurological status, and activities of daily living. We also see consistency across the follow-up period, with greater effects at 7 and 30 days than at 90 days. The more profound time-related effect on NIHSS than mRS is consistent both with an effect on neurological status that later translates into a benefit on disability, and with the expected dilution of treatment effects by unrelated conditions that occur after stroke. Taken to an extreme, if outcome were tested 10 years after acute ischemic stroke, the incidence of unrelated ischemic heart disease, cancer, and death would be so high in both groups that any beneficial effect of NXY-059 on stroke disability would likely be diluted to undetectable levels. As expected, the mRS was consistently the most sensitive efficacy variable.[15] The lesser sensitivity of NIHSS and BI may contribute to their lack of statistical significance at some of the later assessments. The BI considers only the need for help with activities of daily living and not the presence of symptoms or loss of social activities. It is known to suffer ceiling effects, ie, it fails to detect milder forms of disability, and will therefore lose

sensitivity as recovery progresses across time from stroke.[16] A 90-day follow-up period has long been a standard within the field to account appropriately for the time course of functional recovery.[16,17] Assessment at earlier time points may, however, as evidenced by the present trial, be more sensitive in determining the existence of a treatment response.[16]

The main potential concern raised by the SAINT I trial results was the apparent lack of correspondence between the mRS and NIHSS. Here, we see that our failure to affect NIHSS is primarily a statistical fallacy, because the odds ratios, CIs, and probability values of the exploratory NIHSS tests correspond well with those for mRS and BI: the prespecified test was inappropriate and in retrospect was chosen in error. There is accumulating evidence to suggest that mRS may be the optimal end point for phase III stroke trials based on statistical, clinical, and regulatory grounds, in preference to grading a neurological examination or to analyses of the BI.[15–19] The SAINT I trial was designed to deliver 90% power to detect an effect on mRS equivalent to a common odds ratio of 1.3; we actually observed a common odds ratio of 1.2. We have also examined the power of a trial of 1700 patients to detect a benefit on NIHSS under these circumstances and found that a trial of that size will have <20% power to declare significance with the observed effect on change in baseline in NIHSS but 35% for detecting the
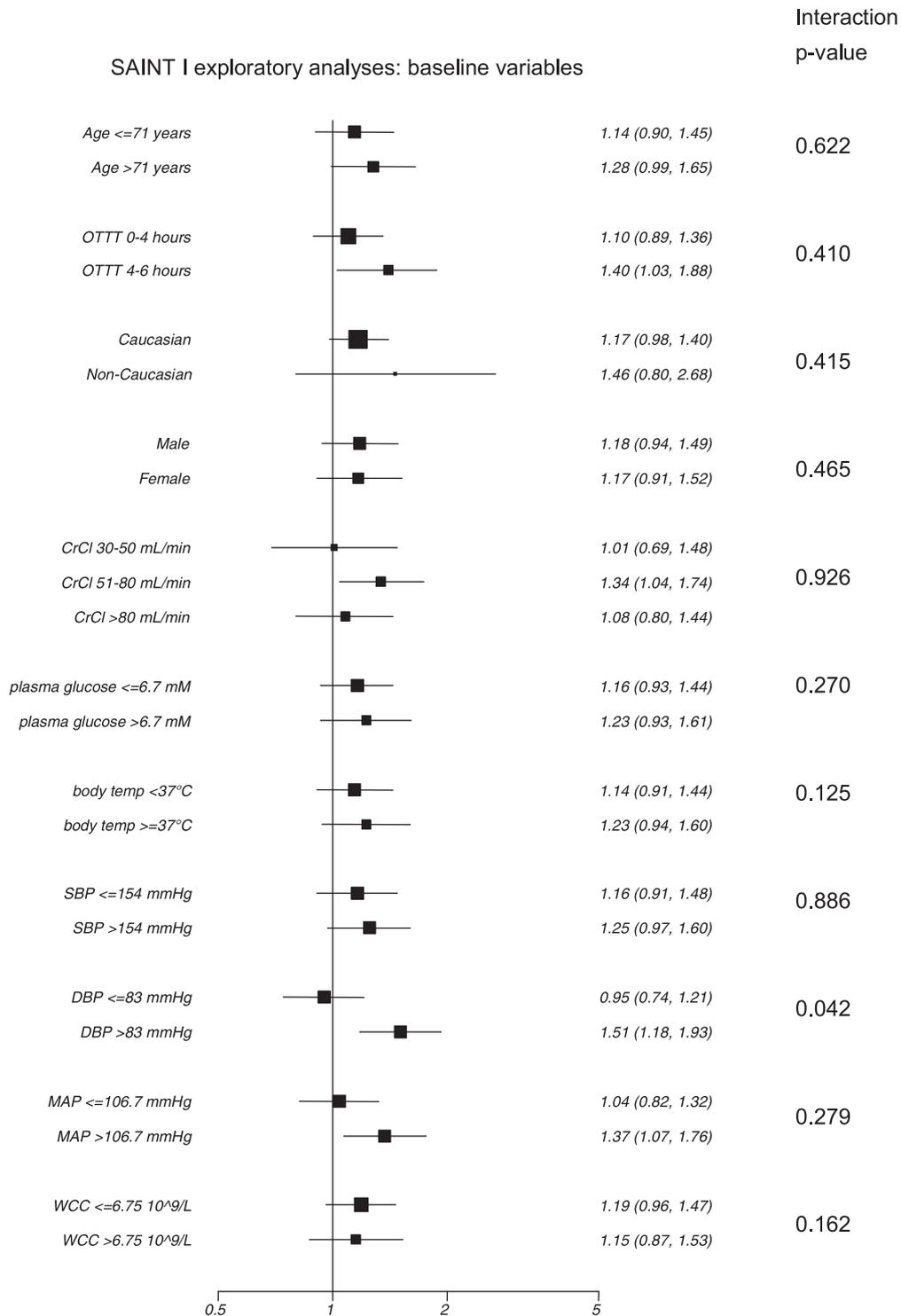
Figure 4. Efficacy in SAINT I trial according to baseline variables. The primary outcome measure was mRS at 90 days or last rating, as analyzed by the CMH test with modified ridit score, adjusted for baseline stratification variables of stroke severity (NIHSS in 4 categories), side of infarct (right or left hemisphere), and use or intended use of tissue plasminogen activator. The CMH test considers the full range of the mRS, with deaths included at mRS 5. Odds ratios were estimated by logistic regression, adjusting in the same way as for the primary end point, and are shown with their estimated 95% CIs, along with the probability value from the CMH test; however, because logistic regression uses a different algorithm to the CMH test, it is possible for minor discrepancies to occur between the CMH probability value and the 95% CIs. The CMH probability value is the primary measure of significance. Interactions between the baseline variables recorded and the treatment effect were tested using the CMH test, and the probability value for each of these is shown in the Figure. No adjustment has been made for multiplicity. Interpretation should be cautious in light of the large number of such exploratory tests: as an approximate guide, Bonferroni correction would suggest that a probability value of 0.002 would indicate a "significant" interaction. Because none of these was significant, the individual odds ratios and their 95% confidence limits are provided only for illustration: a more reliable estimate of treatment effect is obtained from the primary outcome in Figure 2. For purposes of all analyses shown, patients who died were awarded the worst score on the relevant scale. All analyses are based on the intent-to-treat principle.
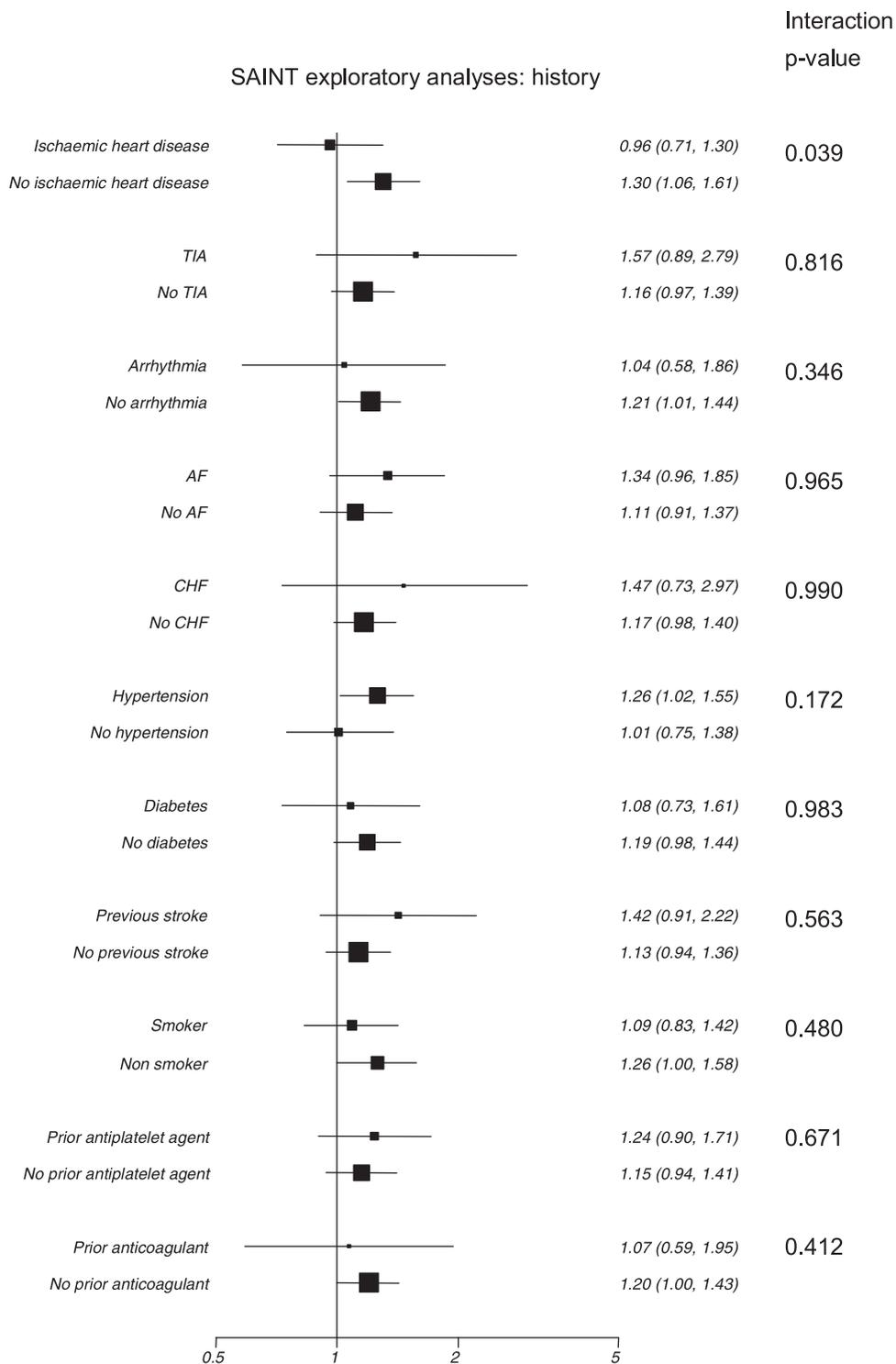
Interaction
p-value

SAINT exploratory analyses: history

Subgroup	OR (95% CI)	Interaction p-value
Ischaemic heart disease	0.96 (0.71, 1.30)	0.039
No ischaemic heart disease	1.30 (1.06, 1.61)	
TIA	1.57 (0.89, 2.79)	0.816
No TIA	1.16 (0.97, 1.39)	
Arrhythmia	1.04 (0.58, 1.86)	0.346
No arrhythmia	1.21 (1.01, 1.44)	
AF	1.34 (0.96, 1.85)	0.965
No AF	1.11 (0.91, 1.37)	
CHF	1.47 (0.73, 2.97)	0.990
No CHF	1.17 (0.98, 1.40)	
Hypertension	1.26 (1.02, 1.55)	0.172
No hypertension	1.01 (0.75, 1.38)	
Diabetes	1.08 (0.73, 1.61)	0.983
No diabetes	1.19 (0.98, 1.44)	
Previous stroke	1.42 (0.91, 2.22)	0.563
No previous stroke	1.13 (0.94, 1.36)	
Smoker	1.09 (0.83, 1.42)	0.480
Non smoker	1.26 (1.00, 1.58)	
Prior antiplatelet agent	1.24 (0.90, 1.71)	0.671
No prior antiplatelet agent	1.15 (0.94, 1.41)	
Prior anticoagulant	1.07 (0.59, 1.95)	0.412
No prior anticoagulant	1.20 (1.00, 1.43)	

0.5   1   2   5

Figure 5. Efficacy in SAINT I trial according to medical history. The primary outcome measure was mRS at 90 days or last rating, as analyzed by the CMH test with modified ridit score, adjusted for baseline stratification variables of stroke severity (NIHSS in 4 categories), side of infarct (right or left hemisphere), and use or intended use of tissue plasminogen activator. The CMH test considers the full range of the mRS, with deaths included at mRS 5. Odds ratios were estimated by logistic regression, adjusting in the same way as for the primary end point, and are shown with their estimated 95% CIs, along with the probability value from the CMH test; however, because logistic regression uses a different algorithm to the CMH test, it is possible for minor discrepancies to occur between the CMH probability value and the 95% CIs. The CMH probability value is the primary measure of significance. Interactions between the medical history variables recorded and the treatment effect were tested using the CMH test, and the probability value for each of these is shown in the Figure. No adjustment has been made for multiplicity. Interpretation should be cautious in light of the large number of such exploratory tests: as a rough guide, Bonferroni correction would suggest that a probability value of 0.002 would indicate a "significant" interaction. As none of these was significant, the individual odds ratios and their 95% confidence limits are provided only for illustration: a more reliable estimate of treatment effect is obtained from the primary outcome in Figure 2. For purposes of all analyses shown, patients who died were awarded the worst score on the relevant scale. All analyses are based on the intent to treat principle.

Power to Detect Observed Difference Between Treatment Groups on NIHSS According to Sample Size and Selection of End Point, Based on CMH Test With Modified Ridit to Adjust for Side of Stroke, NIHSS at Entry, and Treatment With Alteplase

Sample Size	1700	2200	2700	3200	4700	7000
Total NIHSS at last rating	35	44	51	59	75	90
Change from baseline in NIHSS at last rating	16	19	23	26	36	48
Dichotomized 0–1 vs >1 at last rating	24	29	35	40	55	NA

This assumes analysis of NIHSS at 90 days or last rating, analysed across the whole distribution of scores by the Cochran-Mantel-Haenszel χ^2 test (CMH), adjusted for stratification variables (NIHSS, side of infarct, and alteplase use), and using modified ridit scores, i.e., midrank/(N of observations + 1), to account for ordered categories. The model on which the CMH test was calculated included the following variables for analysis of NIHSS total and change score: treatment (2 levels), side of infarct (2 levels), stratified NIHSS baseline score (4 levels), and intent-to-treat or treatment with rt-PA (2 levels). Using our trial data, a bootstrap method was used randomly to draw 5000 times with replacement "n" patients from the NXY-059 and placebo groups. The resulting CMH test for NIHSS final and change scores and for the dichotomization approach were calculated for each random sample. The power is defined as the percentage of CMH tests, which resulted in $P \le 0.05$.

NA indicates not assessed.

observed difference in final NIHSS score. The CMH test is a nonparametric rank-based test across categorized data. The ranks across values of total NIHSS are different and less variable than the ranks across the change from baseline, both being adjusted for the stratification variables. In our trial, we observed values for individual patients' final score that ranged from 0 to 42 (43 categories), and for their change score that ranged from −23 to 36 (50 categories) (Figure 1). This increase in number of categories weakens the test. The SAINT II trial (NCT 0006101022) has now been expanded from 1700 patients to 3200 to enhance its power (to 80%) to confirm our result on mRS. The sample size change was based entirely on mRS considerations. This revision has been reviewed and approved by the relevant ethics committees and regulatory authorities. The secondary analysis has been changed to examine total NIHSS at last rating using the CMH test across all categories; however, even these combined changes will deliver under 60% power to confirm a similar benefit as in SAINT I on NIHSS (Table). NIHSS therefore remains a secondary end point of SAINT II, after analysis of the mRS as primary end point.

Discussion of statistical power should not detract from the important question of clinical utility. The greatest advances in health derive from modest benefits that can be applied to populations rather than life-saving surgery for a few. On present estimates, a one-grade improvement in mRS appears likely to occur once among every 7 or 8 patients treated, and the need for a wheelchair avoided once for every 27 patients treated.[3] Six million new strokes occur per annum worldwide. To individual patients, clinicians could explain that NXY-059 may improve their odds of avoiding more severe forms of disability by 20% and, that left untreated, most patients will have some degree of long-term symptoms or disability, but that NXY-059 treatment could bring an extra 1 in 8 patients

out of long-term symptoms or confinement to a wheelchair. The "cure rate" will increase from 1 in 9 to 1 in 6 or 7.

There are risks associated with exploratory analyses of trials, especially when the main treatment effect has been only modestly positive. Our analyses presented here make it neither more nor less likely that the initial finding was caused entirely by chance; the probability of a false-positive result remains 0.038. We remain cautious in our use and interpretation of these exploratory analyses. However, they are based on the intention-to-treat principle; where exploratory analyses have been undertaken we have reported all rather than a subset and have done so without adjustment for multiplicity; our exploratory analyses are based only on predefined variables, ie, those that were prospectively collected at study entry; and in instances in which statistical methodology diverges from the protocol-defined approach, we have used the method from our protocol or from the literature that most closely matches the original intention. For example, the analysis of NIHSS outcomes using dichotomization between 0 and 1 and 2 and 42 is not the most sensitive to reveal a treatment effect in our dataset but as the method most widely reported elsewhere has been reported here.[13,14] We have simultaneously reported the results using the CMH test, which matches our primary analysis.

In summary, we report reassuring findings from exploratory analyses of interactions with baseline variables and of subgroups in the SAINT I trial. Although only our primary end point was positive on prespecified analysis, examination of additional analyses of each of the main outcome measures at various intervals after trial entry reveals supportive trends toward benefit. The power of a study of the size of SAINT I to detect useful clinical benefit on additional outcomes measured by NIHSS or BI is low, and only a larger confirmatory trial such as SAINT II will provide conclusive evidence of whether NXY-059 is effective in limiting disability after acute ischemic stroke.

Appendix

Acknowledgments

including collecting and storing the data and performing analysis according to the approved plan.

Disclosures

K.R.L., J.A.Z., J.G., and S.M.D. have received fees and expenses from AstraZeneca for steering committee work and lectures but have no financial or related interest in AstraZeneca, Renovis or NXY-059. JAZ has received payments from Ambit, AstraZeneca, CytRx, diaDexus, Halozyme Therapeutics, HealthCare Research and Consulting, Merck Research Laboratories, Johnson and Johnson, Photo-Thera, PhRMA, Pfizer, Remedy Pharmaceuticals; he and his institution (UCSD) have applied for a use patent for the combination of rt-PA plus a neuroprotective agent, for which a ruling is awaited from the patent office. A.D. is or has been a consultant or speaker for AstraZeneca, Boehringer Ingelheim, Pfizer, MSD, Sanofi-Synthelabo, BMS, Bayer, Paion, Forest, Daiichi Asubio, Lilly, Fujisawa, Novonordisk, and Ferrer International. H.C.D. is or has been a consultant or speaker for AstraZeneca, GlaxoSmithKline, Boehringer Ingelheim, BASF, Abbott, Novartis, Parke-Davis, MSD, Servier, Sanofi-Synthelabo, Bayer, Fresenius, and Janssen Cilag. P.L. is or has been a consultant or speaker for AstraZeneca, Bayer, Mitsubishi, Pfizer, Lilly, Merck; and holds research contracts with AstraZeneca and Bayer. J.G. has research suppport from AstraZeneca, NovoNordisk and Boehringer Ingleheim. A.S. is or has been a consultant or speaker for AstraZeneca, GlaxoSmithKline, Boehringer Ingelheim, Pfizer, Roche, Merck, and Sanofi-Synthelabo. T.A., H.G.H., and W.W. are employees of AstraZeneca and hold stock in AstraZeneca. The SAINT trials are sponsored by AstraZeneca. NXY-059 is subject to a partnership agreement between AstraZeneca and Renovis. Renovis had no influence on the conduct, analysis, or interpretation of the study.

References

1. Hacke W, Donnan G, Fieschi C, Kaste M, von Kummer R, Broderick JP, Brott T, Frankel M, Grotta JC, Haley EC Jr., Kwiatkowski T, Levine SR, Lewandowski C, Lu M, Lyden P, Marler JR, Patel S, Tilley BC, Albers G, Bluhmki E, Wilhelm M, Hamilton S; ATLANTIS Trials Investigators; ECASS Trials Investigators; NINDS rt-PA Study Group Investigators. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet.* 2004 363(9411):768–74.
2. IMAGES Study Group. Magnesium for acute stroke (Intravenous Magnesium Efficacy in Stroke trial): randomised controlled trial. *Lancet.* 2004;363:439–445.
3. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, Grotta J, Lyden P, Shuaib A, Hardemark HG, Wasiewski WW; Stroke-Acute Ischemic NXY Treatment (SAINT I) Trial Investigators. NXY-059 for acute ischemic stroke. *N Engl J Med.* 2006;354:588–600.
4. Lees KR, Sharma AK, Barer D, Ford GA, Kostulas V, Cheng YF, Odergren T. Tolerability and pharmacokinetics of the nitrone NXY-059 in patients with acute stroke. *Stroke.* 2001;32:675–680.
5. Lees KR, Barer D, Ford GA, Hacke W, Kostulas V, Sharma AK, Odergren T; SA-NXY-0004 Investigators. Tolerability of NXY-059 at higher target concentrations in patients with acute stroke. *Stroke.* 2003; 34:482–487.
6. Green AR, Ashwood T, Odergren T, Jackson DM. Nitrones as neuroprotective agents in cerebral ischemia with particular reference to NXY-059. *Pharmacol Ther.* 2003;100:195–214.
7. Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke.* 1989;20: 864–870.
8. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975;31: 103–115.
9. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, Van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988;19:604–607.
10. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md Med J.* 1965;14:61–65.
11. Stokes ME, Davis CS, Koch GG. Categorical Data Analysis using the SAS system, 2nd edition, Cary, NC: SAS Institute Inc; 2000.
12. Chi GYH. Multiple testing: multiple comparisons and multiple endpoints. *Drug Information J.* 1998;32: 1347S–1362S
13. Lees KR, Asplund K, Carolei A, Davis SM, Diener HC, Kaste M, Orgogozo JM, Whitehead J. Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. GAIN International Investigators. *Lancet.* 2000;355: 1949–1954.
14. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* 1995;333:1581–1587.
15. Young FB, Lees KR, Weir CJ. Strengthening acute stroke trials through optimal use of disability endpoints. *Stroke.* 2003;34:2676–2680.
16. Fisher M; for the Stroke Therapy Academic Industry Roundtable. Enhancing the development and approval of acute stroke therapies: Stroke Therapy Academic Industry Roundtable. *Stroke.* 2005;36: 1808–1813.
17. Committee for Proprietary Medicinal Products (CPMP). Points to consider on clinical investigation of medicinal products for the treatment of acute stroke. The European Agency for the Evaluation of Medicinal Products. *London.* 2001 CPMP/EWP/560/98.
18. Young FB, Weir CJ, Lees KR; GAIN International Trial Steering Committee and Investigators. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. *Stroke.* 2005;36:2187–2192.
19. Weimar C, Kurth T, Kraywinkel K, Wagner M, Busse O, Haberl RL, Diener HC; German Stroke Data Bank Collaborators. Assessment of functioning and disability after ischemic stroke. *Stroke.* 2002;33:2053–2059.