



University  
of Glasgow

Whitehead, J., Bolland, K., Valdes-Marquez, E., Lihic, A., Ali, M. ,  
and Lees, K.R. (2009) *Using historical lesion volume data in the design  
of a new phase II clinical trial in acute stroke.* Stroke, 40 (4). pp. 1347-  
1352. ISSN 0039-2499

<http://eprints.gla.ac.uk/18111/>

Deposited on: 18 January 2012

# Stroke

American Stroke  
Association<sup>SM</sup>

JOURNAL OF THE AMERICAN HEART ASSOCIATION

A Division of American  
Heart Association



## Using Historical Lesion Volume Data in the Design of a New Phase II Clinical Trial in Acute Stroke

John Whitehead, Kim Bolland, Elsa Valdès-Márquez, Anela Lihic, Myzoon Ali and Kennedy Lees

*Stroke* 2009, 40:1347-1352: originally published online February 19, 2009

doi: 10.1161/STROKEAHA.108.531442

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2009 American Heart Association. All rights reserved. Print ISSN: 0039-2499. Online ISSN: 1524-4628

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

<http://stroke.ahajournals.org/content/40/4/1347>

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

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

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

# Using Historical Lesion Volume Data in the Design of a New Phase II Clinical Trial in Acute Stroke

John Whitehead, PhD; Kim Bolland, PhD; Elsa Valdès-Márquez, PhD; Anela Lihic, MD; Myzoon Ali, MRes; Kennedy Lees, MD, FRCP; for the VISTA Collaborators

**Background and Purpose**—Clinical research into the treatment of acute stroke is complicated, is costly, and has often been unsuccessful. Developments in imaging technology based on computed tomography and magnetic resonance imaging scans offer opportunities for screening experimental therapies during phase II testing so as to deliver only the most promising interventions to phase III. We discuss the design and the appropriate sample size for phase II studies in stroke based on lesion volume.

**Methods**—Determination of the relation between analyses of lesion volumes and of neurologic outcomes is illustrated using data from placebo trial patients from the Virtual International Stroke Trials Archive. The size of an effect on lesion volume that would lead to a clinically relevant treatment effect in terms of a measure, such as modified Rankin score (mRS), is found. The sample size to detect that magnitude of effect on lesion volume is then calculated. Simulation is used to evaluate different criteria for proceeding from phase II to phase III.

**Results**—The odds ratios for mRS correspond roughly to the square root of odds ratios for lesion volume, implying that for equivalent power specifications, sample sizes based on lesion volumes should be about one fourth of those based on mRS. Relaxation of power requirements, appropriate for phase II, lead to further sample size reductions. For example, a phase III trial comparing a novel treatment with placebo with a total sample size of 1518 patients might be motivated from a phase II trial of 126 patients comparing the same 2 treatment arms.

**Discussion**—Definitive phase III trials in stroke should aim to demonstrate significant effects of treatment on clinical outcomes. However, more direct outcomes such as lesion volume can be useful in phase II for determining whether such phase III trials should be undertaken in the first place. (*Stroke*. 2009;40:1347-1352.)

**Key Words:** magnetic resonance imaging scan ■ phase II trial ■ sample size ■ stroke database

Clinical research into new interventions for patients with acute stroke is complicated, costly, and too often unsuccessful. Whereas the NINDS study of tPA<sup>1</sup> found significant evidence of a worthwhile treatment effect and the first SAINT trial<sup>2</sup> demonstrated a modestly beneficial effect, other substantial phase III trials, including the GAIN study<sup>3</sup> and the second SAINT trial,<sup>4</sup> have failed to find an advantage of treatment. Such negative experiences have underlined the need for careful clinical research before the launch of definitive studies and in particular for high-quality phase II studies. At the same time, the development of reliable imaging techniques based on computed tomography (CT) and magnetic resonance imaging (MRI) scans has allowed researchers to form a clearer picture of the direct effects of treatment on lesion volume.<sup>5-7</sup> Although demonstration of a treatment effect on lesion volume may not by itself be clinically convincing nor sufficient for licensing purposes, it can provide compelling motivation for the phase III study of

an agent intended to act through limiting infarction size. It is thus appropriate to conduct phase II studies of such treatments primarily to establish a convincing effect in terms of lesion volume.

The purpose of this article is to present design methodology, in particular the computation of sample size. Investigators have experience with deciding the size of neurologic effect that a phase III trial should be powered to detect. We consider the magnitude of effect on lesion volume that is consistent with the specified effect on functional outcome. Usually the effect on lesion volume will be greater and easier to demonstrate with small to moderate samples. The approach to design is illustrated with data from the Virtual International Stroke Trials Archive (VISTA).<sup>8</sup> These data were not collected with our purpose in mind, they are not ideal, and we do not present our findings on the relation between lesion volume and neurologic outcomes as definitive conclusions on the matter. The comparator to our approach is the current

Received July 11, 2008; final revision received September 9, 2008; accepted September 17, 2008.

From the Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, U.K. (J.W.), School of Biological Sciences, The University of Reading, UK (K.B., E.V.-M.), Xigen S.A., Lausanne, Switzerland (A.L.) and Division of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, UK.

Correspondence to John Whitehead, PhD, Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK. E-mail j.whitehead@lancaster.ac.uk

© 2009 American Heart Association, Inc.

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

DOI: 10.1161/STROKEAHA.108.531442

**Table 1. Cross-Tabulation of Lesion Volume (cm<sup>3</sup>) and mRS, Both Assessed at 90 Days in 301 Placebo Patients**

Lesion Volume	mRS, n (Row %)							Total (Col %)
	0	1	2	3	4	5	6 (Death)	
≤5	23 (28.05)	20 (24.39)	13 (15.85)	9 (10.98)	11 (13.41)	2 (2.44)	4 (4.88)	82 (27.24)
5–25	8 (11.11)	21 (29.17)	9 (12.50)	11 (15.28)	7 (9.72)	6 (8.33)	10 (13.89)	72 (23.92)
25–50	2 (5.88)	5 (14.71)	9 (26.47)	8 (23.53)	5 (14.71)	0 (0)	5 (14.71)	34 (11.30)
50–75	0 (0)	1 (5.88)	2 (11.77)	4 (23.53)	7 (41.18)	0 (0)	3 (17.65)	17 (5.65)
75–100	0 (0)	2 (10.00)	2 (10.00)	6 (30.00)	6 (30.00)	2 (10.00)	2 (10.00)	20 (6.65)
100–125	0 (0)	0 (0)	1 (5.88)	3 (17.65)	7 (41.18)	2 (11.77)	4 (23.53)	17 (5.65)
>125	0 (0)	0 (0)	0 (0)	3 (5.09)	14 (23.73)	9 (15.25)	33 (55.93)	59 (19.60)
Total	33 (10.96)	49 (16.28)	36 (11.96)	44 (14.62)	57 (18.94)	21 (6.98)	61 (20.27)	301

strategy of either conducting no phase II study at all, or else designing such a study without formal regard to the magnitude of effect that would be consistent with a positive finding at phase III. Against this standard of comparison, the use of some data, even if imperfect, is far better than the use of none. Were the methodology of this article to be taken up, more satisfactory data might be collected specifically for this purpose.

For simplicity we consider only the comparison of a single experimental treatment and placebo, with patients being randomized in equal numbers between the 2 study arms. The phase II study is conducted to decide whether or not to take the experimental treatment forward to a full phase III trial. In phase II, the primary response will be lesion volume as determined by CT or MRI scan at 90 days after randomization to treatment. A secondary response of the phase II study, which will become the primary response of any subsequent phase III trial, is the functional outcome assessed by the modified Rankin Score (mRS) at 90 days after randomization, or earlier if it is the last observation carried forward.

## Methods

### Establishing the Relation Between Lesion Volume and mRS

Cross-tabulation of 90-day mRSs and lesion volumes as recorded at 90 days was made for 301 patients (Table 1). This was achieved with data from VISTA, which contains records from > 27 500 patients treated within major clinical trials. The identification of some of the studies in the VISTA database when used for particular analyses is constrained by agreements with the contributing parties, and they were used in this investigation without specification of the individual sources. Records of 1300 placebo-treated patients who had acute ischemic stroke and had recorded mRSs at 90 days were extracted. Lesion volumes at 1, 7, or 90 days, all measured by assessors blinded to treatment group with CT scanning stroke, were available for 309 of these patients. The latest of these scores was used in the analysis (last observation carried forward). Eight more patients were removed from the dataset owing to other missing or inconsistent data. The method introduced here requires both lesion volume and mRSs to be classified into a small number of discrete groups. The method is illustrated with lesion volumes classified into the 7 groups: ≤5, >5 and ≤25, >25 and ≤50, >50 and ≤75, >75 and ≤100, >100 and ≤125, and >125 cm<sup>3</sup> and mRSs classified into the 7 groups: 0, 1, 2, 3, 4, 5, and 6 (death), although other classifications could be used.

### Quantifying Treatment Effect

The data presented in Table 1 represent a sample of placebo patients. To build a picture of patients treated with a drug having the desired effect, a statistical model known as the proportional odds model<sup>9</sup> is

used. The model concerns ratios such as that of the probability that the lesion volume is ≤50 to the probability that it is >50: known as the odds that the lesion volume is ≤50. Suppose that these odds are  $\psi_L$  times greater for a patient on active treatment than for a patient on placebo (the symbol  $\psi_L$  represents this multiple and is known as the odds ratio). The larger the value of  $\psi_L$ , the greater the benefit of active treatment, with  $\psi_L=1$  when the 2 treatment groups have identical lesion volume distributions. It is assumed that the same multiple,  $\psi_L$ , applies if we dichotomize the scale at 5, 25, 75, 100, or 125 instead of 50. For any given value of  $\psi_L$ , this model can be used to transform the observed proportions of placebo patients in each of the lesion categories to corresponding proportions of active treatment patients.

If the treatment acts solely through limiting lesion volume, the effect passed on to mRSs can be found from the cross-tabulation of lesion volumes and mRSs. This process is demonstrated numerically in the Results section, and it leads to the construction of the distribution of mRSs for patients on active treatment that would follow from the assumed effect on lesion volume. Comparison of this distribution with that observed for placebo patients leads to an approximate value for the odds ratio for mRSs, denoted by  $\psi_M$ .

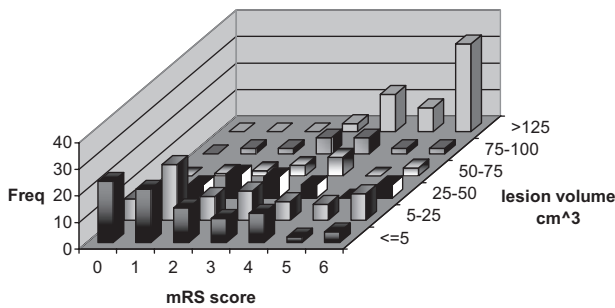
### Sample Size Calculations

The clinically relevant value of the odds ratio  $\psi_M$  for mRSs was deduced by consideration of recent phase III stroke trials and by use of a “number needed to treat” criterion. This is the value that, if present, would be undesirable to miss. The corresponding value of the odds ratio  $\psi_L$  for lesion volumes was found from the cross-tabulation of lesion volumes and mRSs. A standard sample size calculation<sup>10</sup> was then used to find out how many patients to include in the phase II trial.

### Simulations

Four criteria for advancing the active treatment through to phase III testing were explored using simulation. Criterion (a) is that lesion volumes in the active treatment group should be significantly lower than those on placebo at level  $\alpha=0.4$  (2-sided) according to a Mann–Whitney test. Criterion (b) is that mRSs in the active treatment group should be significantly lower than those on placebo at level  $\alpha=0.4$  (2-sided) according to a Mann–Whitney test. Criterion (c) is that criterion (a) is satisfied and there is “a trend” (no matter how small) toward reduction of mRSs. Criterion (d) is the same as criterion (c), but with  $\alpha$  set at 0.48. The choice of these values of  $\alpha$  is justified in the Discussion. For simulations under the null hypothesis, the lesion volume for each patient was generated according to the distribution found from the VISTA data, regardless of whether the patient was on active treatment or placebo. Under the alternative hypothesis, lesion volume data for placebo patients were generated as described, whereas lesion volumes for active treatment patients followed the distribution derived assuming the given odds ratio  $\psi_L$ . Under both hypotheses, the mRS outcome of a patient was generated from the lesion volume already determined, from the

MRS by Lesion volume at 90 days (LOCF)



**Figure.** Diagrammatic representation of the relation between lesion volume (cm<sup>3</sup>) and mRS, both assessed at 90 days in 301 placebo patients.

cross-tabulation of lesion volumes and mRSs. For each scenario 100 000 simulations were run.

**Results**

**Establishing the Relation Between Lesion Volume and mRS**

Of the 301 patients extracted from the VISTA database, 178 were female and 123 male. The age distribution was as follows: 34 younger than 50; 47 in their 50s; 93 in their 60s; 117 between 70 and 84; and 10 age >85. None of the patients received rt-PA. The cross-tabulation of lesion volume and mRS at 90 days for these 301 patients is presented in Table 1, and a graphic representation of the relation is given in the Figure. The Spearman correlation coefficient for this table is 0.586, which is significant with  $P < 0.0001$  (2-sided). (This is consistent with values of 0.54<sup>11</sup> and 0.61<sup>12</sup> reported elsewhere for correlations between lesion volumes and NIHSS scores.) The first row of Table 2 presents the percentages of placebo patients in each of the 7 lesion volume categories and is taken from the final column of Table 1. The second row shows the percentages for active treatment consistent with a proportional odds model with odds ratio  $\psi_L = 1.7070$  (a value that will be justified in due course).

Table 1 is used to find what distribution of mRSs would be anticipated for an agent having the effect on lesion volume shown in Table 2. The percentage of patients on active drug expected to lie in mRS category 0 is found by reading down column 1 of Table 1 and is taken to be 28.05% of those with a lesion volume  $\leq 5$ , 11.11% of those with a lesion volume between 5 and 25, 5.88% of those with a lesion volume between 25 and 50, and so on, to give  $0.2724 \times 28.05 + 0.2392 \times 11.11 +$

**Table 2. Tabulations of Observed Lesion Volumes (cm<sup>3</sup>) at 90 Days for Placebo and Corresponding Constructed Values for Active Treatment Under the Proportional Odds Model With Odds Ratio=1.7070**

Treatment	Lesion Volume, cm <sup>3</sup>						
	$\leq 5$	5-25	25-50	50-75	75-100	100-125	>125
placebo	27.24	23.92	11.30	5.65	6.65	5.65	19.60
active	39.00	25.14	9.82	4.51	5.01	4.03	12.50

Table entries represent the percentage of patients in each category, by treatment.

**Table 3. Tabulations of Observed mRSs at 90 Days for Placebo and Corresponding Constructed Values for Active Treatment Corresponding to an Effect on Lesion Volume With Odds Ratio  $\psi_L = 1.7070$**

Treatment	mRS at 90 Days						
	0	1	2	3	4	5	6 (Death)
Placebo	10.96	16.28	11.96	14.62	18.94	6.98	20.27
Active	14.31	19.06	13.19	14.34	17.10	5.93	16.07

Table entries represent the percentage of patients in each category, by treatment.

$0.1130 \times 5.88 + 0.0565 \times 0 + 0.0665 \times 0 + 0.0565 \times 0 + 0.1960 \times 0 = 10.96\%$  expected to lie in mRS category 0. The proportions with the different lesion volumes are taken from the second row of Table 2, so that  $0.3900 \times 28 + 0.2514 \times 11 + 0.0982 \times 6 + 0.0451 \times 0 + 0.0501 \times 0 + 0.0403 \times 0 + 0.1250 \times 0 = 14.31\%$  of patients on active drug are expected to have an mRS of 0. Proceeding in this way, the distributions for placebo and active drug shown in Table 3 are found. The effect illustrated in Table 3 is not an exact proportional odds model. It can be approximated by such a model in which the odds ratio for mRS is  $\psi_M = 1.3389$ .

**Quantifying Treatment Effect**

One way of expressing the magnitude of the effect is via the “number needed to treat” as expressed by Lees et al.<sup>2</sup> From Table 3 it can be seen that the expected mRS for placebo is  $0 \times 0.1096 + 1 \times 0.1628 + 2 \times 0.1196 + \dots + 6 \times 0.2027 = 3.1628$ , whereas for the active drug a similar calculation gives 2.8295. The difference between the 2 expected mRS values is 0.3333, so that the benefit of active drug amounts to an average improvement of 0.3333 points on the mRS scale per patient, or 1 point per 3 patients. Hence, the odds ratio  $\psi_L = 1.7070$  for lesion volumes, which forms the basis of Table 3, corresponds to the “number needed to treat” equal to 3. Table 4 shows the results of several similar sets of calculation. In each case, computation starts with the odds ratio for the effect of treatment on lesion volume given in the fourth column and uses the transition probabilities shown in Table 1, together with the proportional odds models, to find the corresponding odds ratio for the effect of treatment on mRS and the associated number needed to treat. From a large number of

**Table 4. Odds Ratios for the Effect of Treatment on Lesion Volume at 90 Days Corresponding to Various Odds Ratios for the Effect on mRS at 90 Days**

No. Needed to Treat for an Expected Total Gain of 1 Point on the mRS	Odds Ratio for the Effect of Treatment on mRS ( $\psi_M$ )	Odds Ratio for the Effect of Treatment on Lesion Volume ( $\psi_L$ )
2	1.5518	2.3099
3	1.3389	1.7070
4	1.2444	1.4827
5	1.1911	1.3662
6	1.1569	1.2949
$\infty$ (null)	1.0000	1.0000



**Table 5. Sample Sizes for Phase III and Phase II Trials Powered to Detect Various Magnitudes of Treatment Effect**

NNT	$\psi_M$	$\psi_L$	$n_{III}$	$n_{II}$	$n_{II}$
			(0.05, 0.90)	(0.05, 0.90)	(0.40, 0.80)
2	1.5518	2.3099	670	194	52
3	1.3389	1.7070	1518	468	126
4	1.2444	1.4827	2706	858	232
5	1.1911	1.3662	4230	1364	368
6	1.1569	1.2949	6092	1986	535

NNT indicates number needed to treat.

such calculations made by the authors, those leading to the numbers needed to treat equal to 2, 3, 4, 5, and 6 have been selected for display. Also included is the null situation of no treatment effect. A very rough rule of thumb is that  $\psi_M$  is the square root of  $\psi_L$ .

As odds ratios are larger for lesion volumes, they will be easier to detect, so that trials powered in terms of lesion volume will require smaller sample sizes. It is important to realize the speculative nature of Table 4. If a treatment has a given effect on lesion volume, and if patients on active treatment with a given lesion volume behave just like untreated patients with that same lesion volume, then the treatment effect on mRS will be as shown. The second condition is a suitable assumption to make for planning a large confirmatory phase III trial, and thus Table 4 is appropriate for use in such planning. Table 4 is not in any way intended to replace the subsequent phase III trial and has no basis as proof of any magnitude of treatment effect on neurologic outcome.

**Sample Size Calculations**

We start with a conventional power calculation for a phase III trial based on the mRS outcomes at 90 days, expressed in the ordered categories shown in Tables 1 and 3 and analyzed using the Mann-Whitney test applied to data grouped into categories. (This test is identical to analysis using a proportional odds regression model in the absence of prognostic factors.) Suppose that placebo patients are expected to follow the 90-day mRS distribution shown in Table 3. The trial will be powered to detect a treatment effect with magnitude expressed as an odds ratio of 1.3389. Thus, if the mRS distribution on active treatment is also as shown in Table 3, then there should be a probability of  $(1-\beta)$  of detecting significant treatment effect at level  $\alpha$  (2-sided). The appropriate sample size  $n$  is given by<sup>10</sup>

$$(1) \quad n = \frac{12 (u_{\alpha/2} + u_{\beta})^2}{(\log_e \Psi_M)^2 \left( 1 - \sum_{j=0}^6 \bar{p}_j^3 \right)}$$

Here  $n$  is the total sample size, divided equally between  $1/2n$  on active treatment and  $1/2n$  on placebo, and  $u_{\alpha/2}$  and  $u_{\beta}$  are the upper  $1/2\alpha$  and  $\beta$  percentage points of the standard normal distribution, respectively. The odds ratio  $\psi_M$  is set to its clinically relevant value, and  $\bar{p}_j$  denotes the proportion of

**Table 6. Proportions of Positive Trial Findings for 4 Different Criteria of Being Positive, Estimated From 100 000-Fold Simulations**

$\psi_L$	Criterion			
	(a)	(b)	(c)	(d)
1	0.1999	0.2012	0.1664	0.1948
1.7070	0.8006	0.5394	0.7109	0.7352

patients in the  $j$ th mRS outcome category, averaging over the placebo and active treatment arms.

A conventional power calculation for a phase III trial based on the mRS outcomes at 90 days proceeds as follows. For  $\psi_M=1.3389$ , and the outcome category probabilities  $\bar{p}_j$  taken from Table 3, equation (1) yields  $n=1518$ , ie, 759 patients per treatment arm. This sample size lies at the lower end of the range of phase III sizes used in practice, as they are usually powered to detect more modest treatment effects. A similar calculation can be performed for an analysis based on lesion volumes. Taking  $\psi_L=1.7070$  and the probabilities of the 7 outcome categories for lesion volumes from Table 2, equation (1) yields  $n=468$ , or 234 patients per treatment arm. This remains a large sample size for phase II. Although the ideal policy would be to recruit this number of patients into the phase II trial, this might be unfeasible in practice. A compromise might be possible. The settings of  $\alpha$  and  $1-\beta$  in the power requirement are suitable for the design of a definitive phase III study but are perhaps unnecessarily demanding for phase II. Instead, values such as  $\alpha=0.40$  and  $1-\beta=0.80$  might be considered. The 2-sided significance level of 0.40 corresponds to a 1-sided level of 0.20. A treatment will be taken forward to phase III if it achieves a 1-sided probability value  $\leq 0.20$  in favor of smaller lesion volumes relative to placebo. With this criterion, a totally inactive treatment is allowed a 20% probability of further study, whereas a treatment with  $\psi_L=1.7070$  on the lesion volume outcome, consistent with an important effect on mRS at 90 days, will not be taken forward with probability  $1-0.80=0.20$ . These error rates lead to a sample size of  $n=126$ , or 63 patients per treatment.

Table 5 reworks these calculations for various target odds ratios. It can be seen that, for equivalent error rates, the sample size required to detect a treatment effect on lesion volume is a little more than a quarter of the eventual sample size required for phase III, and for the relaxed power requirement at phase II, a further reduction of almost a quarter is achieved.

**Simulations**

Simulations were conducted to evaluate criteria (a), (b), (c), and (d) for proceeding to phase III, for phase II trials with a total sample size of 126, or 63 patients per treatment; ie, the design in the second row of Table 5, corresponding to a power of 0.8 to detect significance at the level  $\alpha=0.4$  (2-sided) when  $\psi_L=1.7070$ . Table 6 presents the proportion of runs in which the experimental treatment would be advanced to phase III according to each of the 4 criteria. It can be seen that both criteria (a) and (b) lead to a type I error

rate of 0.2, which is  $\frac{1}{2}\alpha$ , as theory predicts. For the double criterion (c), the type I error rate is lower. Raising  $\alpha$  to 0.48 in criterion (d) returns the 1-sided type I error rate to just short of the allowed value of 0.20. Because criterion (d) involves consideration of both lesion volumes and mRSs, it is more difficult to meet: the use of a “nominal” value of  $\alpha=0.48$  (2-sided) achieves an actual type I error of the magnitude specified. The power for criterion (a) is 0.80, as intended, whereas for mRS (criterion b) it is much lower, at 0.54. Criterion (c), through adding a second requirement to (a), reduces power to 0.71, and criterion (d) recovers some of the lost power to reach 0.74.

### Discussion

In therapeutic areas such as stroke, definitive evidence on the efficacy of a novel treatment should be based on clinical outcomes observed after a lengthy period of follow-up. Phase II trials to determine whether or not to proceed to such a phase III study should be smaller and shorter. As a confirmatory study will follow, a larger type I error is permissible: mistakes can later be rectified. Relaxation of power is more troublesome, as discarded treatments cannot easily be restored. One option in phase II is to focus on an intermediate physiologic end point. If the drug is intended to work by influencing this end point, then evidence that it does so is reasonable motivation for taking it further: lack of such evidence should be sufficient to discard the treatment. Phase III can be used to determine whether this physiologic effect is indeed converted into a clinical advantage. In stroke, the physiologic end point in question might be lesion volume, and a drug devised to improve clinical outcomes after stroke by limiting infarct size might be expected to demonstrate a direct effect on lesion volume before being taken forward. Expert groups such as STAIR have recommended a search for suitable surrogate outcomes to be used in phase II.<sup>13</sup> Our approach will be valid only for treatments that directly affect lesion volume: they would be inappropriate for certain neuroprotective or restorative strategies. Outcomes relating to other forms of therapeutic action could be used in place of lesion volumes in a manner similar to that shown here.

The results presented here were based on lesion volumes at 90 days (or earlier, if it is the last observation carried forward) because we were comparing volume with functional outcome at 90 days and because such data were to hand. It could be advantageous to consider a much earlier imaging end point. If confounding effects of edema can be discounted, then earlier assessment may limit losses due to mortality or withdrawal. Disadvantages of imaging end points must also be considered: CT is insensitive to small subcortical and posterior situated infarcts; both CT and MRI may show several lesions, some of which can be old and thus unrelated to the current stroke. Careful patient selection can limit these disadvantages.

To calculate the size of such a phase II trial, the worthwhile reduction in lesion volume (relative to placebo) must be specified. In this article, we have shown how to specify an effect that is consistent with a meaningful effect on neurologic outcome. The phase III trial will then determine whether the potential due to reduction in lesion volumes is indeed passed on to clinical responses. As the advantage gained

through a direct physiologic effect is likely to be diluted by other effects before being passed on to the clinical outcome, the former direct effect is likely to be larger and consequently easier to detect. In turn, this will justify smaller sample sizes. In the context of stroke, we have found that further measures, such as a large relaxation in the limit on type I error and a smaller reduction in power, are needed to produce phase II sample sizes that might be contemplated as practical by investigators. In the calculations, the value of  $\alpha$  was set at 0.40. This is a large risk of error but perhaps not as large as at first apparent. It is a 2-sided risk of error, indicating that if the treatment were inactive, there would be a probability of 0.20 of proceeding to phase III and a probability of 0.20 of concluding with equal force that the treatment is doing harm. The latter conclusion is of limited interest, as the 2 actions available are to take the treatment forward for further study or not. Even so, 0.20 is a large risk of taking forward an inactive treatment. An error is likely to be put right at phase III, so this is not the public's risk of receiving an inactive treatment. Of course, it would be optimal to keep type I error small and power large. For conventional error rates, the sample size for an analysis based on lesion volume is given above as  $n=468$ . It remains to be seen whether investigators would or should commit such resources to phase II studies.

The numeric findings of this article are only as good as the data on which they are based. Trial planners may wish to rework these calculations with larger databases or databases more relevant to the patient population that they wish to study. When devising our own design, we found the VISTA database to be the most extensive available. It is of interest to note that the marginal distribution of the mRSs shown at the foot of Table 1 is similar to that found for placebo patients in the first SAINT study,<sup>2</sup> which reported the following respective percentages: 11, 20, 12, 13, 21, and 24 (categories 5 and 6 being merged).

Phase II trials often include 2 or 3 dose levels of the investigational drug in addition to placebo. In that case, sufficient power is usually required to make each pairwise comparison with control. In the example of this article, this would lead to 63 patients per arm and 252 patients altogether for a 4-arm trial. It will often be better to reduce the number of dose levels, maybe down to 1, rather than reducing power below the already low level of 0.80.

The use of the concept of number needed to treat is not essential to the approach presented. The concept of number needed to treat has been criticized,<sup>14</sup> and taking expectations over a nominal (rather than interval) scale such as mRS is also problematic. Nevertheless, as a means of establishing what magnitude of treatment effect might be of interest, expressing that effect as an expected reduction in mRS of one third can be helpful (whether or not one then inverts this value to give number needed to treat=3).

It is of interest to compare the approach presented here with that of earlier related work.<sup>15,16</sup> Those studies established that the correlation between MRI measures and neurologic outcomes is statistically significant and determined sample sizes for a phase II study based on the former end points. There are 3 principal differences between this earlier approach and the method presented here: (1) They used

percentage reperfusion, whereas we used lesion volume; (2) In considering continuous measures, they calculated sample sizes using a bootstrapping approach, whereas we used an explicit formula; and (c) They powered the phase II study for a treatment difference in terms of imaging outcome that was selected arbitrarily, whereas we set this difference in terms of the corresponding effect on the neurologic outcome sought. The last is the only fundamental difference and constitutes our main message: here we present a rationale for choosing a treatment effect in terms of the imaging outcome that relates to a neurologic effect of a size that is both credible and clinically important.

### Acknowledgments

We would like to thank Xigen SA for funding this research and for allowing the publication of the results. We should also like to thank the VISTA Steering Committee for providing the data necessary for its completion.

### Disclosures

At the time when the main part of this research was conducted, J.W., K.B., and E.V.-M. were working for the Medical and Pharmaceutical Statistics Research Unit at the University of Reading, a self-financing research group within the university funded by grants and collaborative research contracts with the pharmaceutical industry. A.L. is an employee of Xigen SA. Part of this work was commissioned and funded by Xigen SA.

### References

1. National Institute of Neurological Diseases and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995;333:1581–1587.
2. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, Grotta J, Lyden P, Shuaib A, Hårdemark HG, Wasiewski WW, for the stroke-acute Ischemic NXY Treatment (SAINT 1) Trial Investigators. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006;354:588–600.
3. Shuaib A, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM, Diener HC, Ashwood T, Wasiewski WW, Emeribe U, for the SAINT II Trial Investigators. NXY-059 for the treatment of acute ischemic stroke. *N Engl J Med*. 2007;357:562–571.
4. Lees KR, Asplund K, Carolei A, Davis SM, Diener H-C, Kaste M, Orgogozo J-M, Whitehead J. Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. *Lancet*. 2000;355:1949–1954.
5. Schwamm LH, Koroshetz WJ, Sorensen AG, Wang B, Copen WA, Budzik R, et al. Time course of lesion development in patients with acute stroke: serial diffusion- and hemodynamic-weighted magnetic resonance imaging. *Stroke*. 1998;29:2268–2276.
6. Kalowska E, Rostrup E, Rosenbaum S, Petersen P, Paulson OB. Acute MRI changes in progressive ischemic stroke. *Eur Neurol*. 2008;59:229–236.
7. Grotta JC, Jacobs TP, Koroshetz WJ, Moskowitz MA. Stroke program review group: an interim report. *Stroke*. 2008;39:1364–1370.
8. Ali M, Bath PMW, Curram J, Davis SM, Diener H-C, Donnan GA, Fisher M, Gregson B, Grotta J, Hacke W, Hennerici MG, Hommel M, Kaste M, Marler JR, Sacco RL, Teal P, Wahlgren N-G, Warach S, Weir CJ, Lees KR. The Virtual International Stroke Trials Archive. *Stroke*. 2007;38:1905–1910.
9. McCullagh P. Regression models for ordinal data. *J R Stat Soc B*. 1980;43:109–142.
10. Whitehead J. Sample size calculations for ordered categorical data. *Stat Med*. 1993;12:2257–2271.
11. Saver JL, Johnston KC, Homer D, Wityk R, Koroshetz W, Truskowski LL, Haley EC, for the RANTTAS Investigators. Infarct volume as a surrogate or auxiliary outcome measure in ischemic stroke clinical trials. *Stroke*. 1999;30:293–298.
12. The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study Group. Effect of intravenous recombinant tissue plasminogen activator on ischemic stroke lesion size measured by computed tomography. *Stroke*. 2000;31:2912–2919.
13. Stroke Therapy Academic Industry Roundtable II (STAIR II). Recommendations for clinical trial evaluation of acute stroke therapies. *Stroke*. 2001;32:1598–1606.
14. Hutton JL. Number needed to treat: properties and problems. *J R Stat Soc A*. 2000;163:403–419.
15. Barber PA, Parsons MW, Desmond PM, Bennett DA, Donnan GA, Tress BM, Davis SM. The use of PWI and DWI in the design of ‘proof-of-concept’ stroke trials. *J Neuroimaging*. 2004;14:123–132.
16. MR Stroke Collaborative Group. Proof-of-principle phase II MRI studies in stroke. *Stroke*. 2006;37:2521–2525.