Published in final edited form as:
Stroke. 2012 March ; 43(3): . doi:10.1161/STROKEAHA.111.632935.

A simple, assumption-free and clinically interpretable approach for analysis of modified Rankin outcomes

George Howard, DrPH[1], **Jennifer L. Waller, PhD**[2], **Jenifer H. Voeks, PhD**[3], **Virginia J. Howard, PhD**[3], **Edward C. Jauch, MD**[4], **Kennedy R. Lees, MD**[5], **Fenwick T. Nichols, MD**[6], **Volker W. Rahlfs, PhD**[7], and **David C. Hess, MD**[6]

[1]Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

[2]Department of Biostatistics, Georgia Health Sciences University, Augusta, GA

[3]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

[4]Department of Neurology, Medical University of South Carolina, Charleston, SC

[5]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, Scotland, UK

[6]Department of Neurology, Georgia Health Sciences University, Augusta, GA

[7]idv Data Analysis and Study Planning, Krailling/Munich, Germany

Abstract

Introduction—There is debate regarding the approach for analysis of modified Rankin scores (mRS), the most common functional outcome scale used in acute stroke trials.

Methods—We propose to use tests to assess treatment differences addressing the metric "If a patient is chosen at random from each treatment group and if they have different outcomes, what is the chance the patient who got the investigational treatment will have a better outcome than the patient receiving the standard treatment?" This approach has an associated statement of treatment efficacy easily understood by patients and clinicians, and leads to statistical testing of treatment differences by tests closely related to the Mann-Whitney U test (a.k.a. Wilcoxon Rank-Sum test) which can be tested precisely by permutation tests (a.k.a. randomization tests).

Results—We show that a permutation test is as powerful as other approaches assessing ordinal outcomes of the mRS scale, and provide data from several examples contrasting alternative approaches.

Discussion—While many approaches to analysis of mRS outcomes have generally similar statistical performances, this proposed approach: 1) captures information from the ordinal scale, 2) provides a powerful clinical interpretation understood by both patients and clinicians, 3) has power at least equivalent to the other ordinal approaches, 4) avoids assumptions in the

Correspondence: George Howard, DrPH; Professor and Chair; Department of Biostatistics, UAB School of Public Health; Birmingham, AL 35294-0022; Phone 205-934-4905; ghoward@uab.edu.

Conflict of Interest None of the authors have a conflict of pertinent to this report.

parameterization, and 5) provides an interpretable parameter based on the same foundation as the calculation of the p-value.

Keywords

Stroke; ordinal analysis; modified Rankin score

Introduction

The stroke community is in the midst of a spirited discussion regarding the optimal approaches to the analysis and interpretation of the modified Rankin score (mRS) commonly used as an outcome in acute stroke clinical trials.[1-6] The mRS is a 7-point scale ranging from 0 (no symptoms) to 6 (dead) (see Table 1). The most severe scores, representing severe disability (mRS = 5) and death (mRS = 6), are frequently pooled as these outcomes are considered equivalently bad. Two general analysis approaches for examining mRS outcomes have been employed:

1. Categorical statistical approach where the mRS is dichotomized. Most studies have taken the approach of either dichotomizing the scale at a fixed point (as was done by the ALIAS 1 investigators[7]) or using a "sliding dichotomy" where a good outcome is defined as a function of baseline severity (as was done by the PAIS investigors[8]). This approach has the advantage of simplicity of clinical interpretation (and has been strongly "encouraged" by the FDA) but has the disadvantage of failing to harness information from the entire spectrum of mRS outcomes. Analysis of the mRS dichotomous groups has been generally implemented by chi-square testing or logistic regression.
2. Ordinal statistical approaches of the distribution across the entire mRS scale. Two analytic ordinal approaches have been employed, both having the advantage of increased statistical power provided by using information from the entire mRS scale and capturing information nested within the strata of outcomes defined by the dichotomization approach above. For example, the recently reported ALIAS 1 results showed 36.2% good outcomes (16.1% with mRS of 0, and 20.1 with mRS of 1) in the albumin-treated group, which on a relative basis is 29% (relative risk 1.29) higher than the 28.2% good outcomes in the saline-treated group (13.7% with mRS of 0, and 14,5% with mRS of 1).[7] However, within the poor outcome strata of 2 through 6, there was also an increase in the proportion of patients with very poor outcomes of 5 and 6; specifically a 20.1% death or severely disabled outcome (4.0% with mRS of 5, and 16.5 with an mRS of 6 in the albumin-treated group) which is 20% (relative risk 1.20) higher than the 16.8% death or severely disabled outcomes in the saline-treated group (3.8% with mRS of 5 and 13.0% with mRS of 6).[7] Because the redistribution of scores is within the "poor outcome" stratum, the difference is not revealed in the dichotomous analysis.

There have been two commonly used statistical approaches for the analysis of ordinal scales, both with advantages and disadvantages, specifically:

a. *The Proportional Odds Model (POM)*. This is a straightforward generalization of logistic regression where the odds ratio is calculated for each cut-point across the mRS (for example 0 versus 1 to 6, then 0 to 1 versus 2 to 6, and so on), and then a summary odds ratio is calculated from the individual odds ratios under the assumption that the individual odds ratios are the same.[9] This approach has the advantage of providing a clinically interpretable parameter (the estimated summary odds ratio), but the disadvantage of requiring the assumption that the individual odds ratios are the same (the "proportional odds assumption"). A second

disadvantage is the potential for misinterpretation of clinical meaning of the summary odds ratio, where odds ratios are frequently inappropriately interchanged with relative risk[10], and additional misinterpretation of the summary nature of the summary odds ratio.

b. *The Cochran-Mantel-Haenszel (CMH) test.* The CMH is a test of the similarity of the mean rank (across the mRS scale) for two groups.[11] The CMH test has the advantage of requiring virtually no assumptions for calculation of the p-value, but the disadvantage of not providing a clinically interpretable parameter.

In the analysis of the ordinal outcome, some studies (such as SAINT 1), used the CMH test treatment differences, and once significance was established, used the POM to provide a clinically interpretable parameter.[2, 12] While this approach is generally acceptable, it does have the disadvantage of basing the assessment of the significance on a different approach (different metric) than the estimation of the magnitude of the effect, as the two approaches can potentially give disparate results.

Thus, commonly used current approaches require the choice between the easily interpretable dichotomous analyses that fails to capture the information across the entire mRS spectrum versus ordinal analyses that either provide no measures of clinical efficacy or provide measures that are commonly misinterpreted.

An alternative approach

We set out to find an alternative approach that: 1) captures information from the ordinal scale, 2) provides a powerful and easily understood clinical interpretation, 3) has statistical power at least equivalent to the CMH or POM, 4) avoids assumptions in the parameterization, and 5) has the clinically interpretable parameter based on the same foundation as the calculation of the p-value. This process began by attempting to simply state treatment effects in terms that patients and clinicians would understand, and we suggest that the simple statement “If a patient is chosen at random from each treatment group and if they have different mRS outcomes, what is the chance that the patient who got the investigational treatment will have a better outcome than the patient receiving the standard treatment?” captures the essential information necessary for clinical decision making. This statistical statement of treatment efficacy can be rephrased to be easily understood by both patients and treating clinicians. For example, the findings of the NINDS tPA trial[13] would be explained as “Out of 100 patients treated with tPA instead of placebo, 48 will be better with tPA whereas 31 will be better with placebo. Twenty one (21) appear the same with either treatment.” We suggest this simple statement provides the information patients value by focusing on the chance that they will “do better” with a particular treatment. We are also confident that this statement can be further refined with future experience working with patients with the interpretation of results.

This approach was initially proposed over 60 years ago as an approach now referred to as the *Mann-Whitney U Test*.[14] This idea of whether a randomly chosen person from one group has “better” outcome (i.e., lower mRS) than a randomly chosen person from the other group has been shown to be mathematically equivalent to a ranking approach proposed two years earlier (the *Wilcoxon Rank-Sum Test)*.[15] Both tests were developed for a continuous outcome where ties are impossible (or at least very rare), but this is not the case in their use in the analysis of the mRS outcome. Approaches to handle the ties, however, have been subsequently proposed and are now very well accepted.[16, 17] While these two tests are among the most commonly used non-parametric tests, normally the focus is normally on the p-value with little emphasis on the real life, patient-centric description of the magnitude of

the clinical effect. However, it is this description of the effect magnitude that is was the impetus for the proposed alternative approach.

Because our alternative approach arose from discussions with clinicians to develop a statement of clinical effect, it differs slightly from the Mann-Whitney U test on the accounting for tied scores. Specifically, the Mann-Whitney U test assumes that half the pairs with tied scores are superior for one treatment and half are superior in the other (i.e., half the tied scores had a lower mRS score for the patient assigned to investigational treatment, and half had lower mRS score for the standard-treatment patient). While this approach is mathematically attractive, it leads to an awkward clinical statement of treatment differences, perhaps "If a patient is chosen at random from each treatment group, and we assume that half the pairs of patients who have the same mRS score did in fact have a better outcome in the investigational-treatment group and half had a better outcome in the standard-treatment group, then what is the chance that the patients receiving the investigational treatment will have a better outcome than the patient with standard treatment?" We consider this statement awkward to the point that understanding and explaining the clinical effect is a barrier.

Hence, our approach differs from the Mann Whitney U test in how it accounts for the tied values, where we consider only untied pairs of patients and the Mann Whitney U test assumes half to be superior in each group. Hence, the Mann Whitney U test does not precisely test our proposed clinical statement; however, this can be done by a permutation test (a.k.a. randomization test). The permutation test first calculates a "test statistic" for the observed data, in this case we used how far the observed proportion of non-tied pairs was from the null-hypothesis of 50%. The permutation test randomly assigns treatments to individuals, ensuring no association between treatment and the test statistic (guaranteeing the null hypothesis is true). This process is repeated many times and the distribution of the test statistic under the null hypothesis is estimated. Whether the observed test statistic is "unusual" under the null hypothesis (that is, the p-value) is simply the location of the observed test statistic in distribution of test statistics under the null hypothesis. For example, if 234 of 1000 test statistics calculated under the null hypothesis are greater than the observed test statistic, then the p-value is 0.234. While the understanding the details of the construction of the test is slightly complex and performing this test does require a bit of computer programming, we provide SAS code for the calculations in the supplemental material. This concept of permutation test is also not new, but was proposed by Fisher in 1935 as part of the pivotal "Lady Tasting Tea" experiment,[18] and the approach was generalized to randomization test (a.k.a. re-randomization test) when the sample size is large and complete enumeration of all outcomes is impractical.[19] Finally, the correlation between p-values from the permutation test and the Mann Whitney U test in the simulations considered for this report was 0.989, and while this estimated correlation will change with the distribution of mRS scores, this high correlation suggests that the Mann Whitney U test provides an "approximate" p-value for the more correct permutation test.

The calculations underlying our approach can be demonstrated using data from the NINDS tPA trial.[3, 13] Here the distribution of the mRS in the tPA treated patients is shown in rows, while the mRS for placebo treated patients are shown in columns (Table 2). The first patient from the first group is matched up with each patient in the second group, the second patient in the first group with each patient in the second group, and so on to the last patient in the first group matched up with each patient in the second group; hence, the number of pairs of patients is simply the product of the number of patients in each combination. For example, there were 29 tPA treated patients with a mRS of 3, and 30 placebo treated patients with a mRS of 2; hence, there are $29 \times 30 = 870$ pairs of patients with this combination of scores. Altogether, there are 30,178 pairs of patients with higher (worse) mRS than for placebo, 46,016 pairs of patients with lower (better) scores for tPA, and 19,596 pairs of patients with

a tie between placebo and tPA. As such, 48% (46,016 / 95,790) of pairs of patients had lower (better) scores for tPA, 20% (19,596/95,790) had tied scores, and 31% (30,178/95,790) lower (better) scores for placebo. Hence, these data support the statement "Out of 100 patients given tPA instead of placebo, 48 will be better with tPA whereas 31 will be better with placebo. Twenty one (21) appear the same with either treatment."

The approach is also flexible. For example, if the primary hypothesis requires stratification by a covariate (for example, stroke severity), adjustment can be incorporated into the analysis by totaling the of pairs of patients that are better, worse or tied within each stratum of the covariate; and then summing across the covariate strata. Importantly, this calculating the statistic until after the summations have been made, hence not requiring large sample sizes in each strata and allowing for simultaneous control for several confounders. In addition, the potential for effect modification (interaction with treatment) within a trial can be assessed by calculating the proportion with better outcomes within each stratum, and using a permutation test testing if there is a "large" difference between strata.

Not surprisingly, the statistical power to detect treatment effects using this approach is equivalent to the CMH or POM. As demonstrated in Table 3, we assumed a "standard treatment" mRS distribution of the placebo-treated patients in the NINDS tPA study, then created a distribution for patients treated with the investigational drug by "shifting" an increasing larger proportion of individuals to lower mRS strata (details the approach for shifting is provided as supplemental material). The power of our approach is indistinguishable from the power of the POM, both of which are generally marginally more powerful than the CMH test. We stress that it is possible also to specify distributions where the CMH will be marginally more powerful than either this proposed approach or the POM, and it is also possible to specify more extreme distributions where the dichotomous analysis is most powerful.

Finally, Table 4 provides a comparison of the alternative ordinal approaches for four recently reported studies. As expected, the statistical test evaluating treatment effects is concordant among the three approaches.

Discussion

Our analytical approach meets all the goals we established to achieve in defining a statistical approach, specifically: 1) it captures information from the ordinal scale, 2) it provides a clinical interpretation that is easily understood, 3) it has power at least equivalent to the CMH or POM, 4) it avoids assumptions in the calculation of the significance of treatment differences, and 5) the clinically interpretable parameter is based on the same foundation as the calculation of the p-value. We suggest that that this approach is:

- Superior to the dichotomous approach as it captures the entire spectrum of the mRS scale, and therefore will be generally more powerful.
- Superior to the CMH test as it provides an easily interpretable index of treatment efficacy that is based on the same metric as the significance test (i.e., does not require the use of the POM, a test using a different metric, to provide a measure of efficacy).
- Superior to the proportional odds model as it avoids the proportional odds assumption.

We would suggest that those using this approach report results visually as has been done by many studies using stacked horizontal bar graphs (sometimes referred to in the acute stroke

literature as “Grotta” bars), also provide the estimate of clinical effect and the p-value for the difference in the treatment.

On the surface, since ties are not considered in calculations, it would seem that the proposed permutation test would have lower power than the Mann-Whitney U, CMH, or POM tests, all of which incorporate the observations with ties into p-value calculations and as such have a larger sample size. However, the Mann-Whitney U test assigns half the ties to be superior for treatment “A” and half to be superior for treatment “B,” and as such within the MWU these observations provide no information to the test (i.e., implicitly excluding the observations). Preliminary investigations have shown that even in situations where the large majority of observations are tied, the relationship of the p-value from the MWU and the proposed test persists. Since the Mann-Whitney U is largely equivalent to the CMH, which is largely equivalent to the POM models, there will also be little difference of power between these tests. While additional investigations are warranted, we believe that the permutation test is generally equivalent to the MWU, CMH or POM.

As the proposed statistical approach is closely related to the Mann-Whitney U test (only differing by the accounting for tied pairs), it should not be surprising that the resulting p-value from the two techniques are remarkably similar. We suggest that the proposed new approach has the advantage that the calculation of the p-value is based on the same “metric” as the statement of clinical efficacy, which we see as a substantial advantage. However, the Mann-Whitney U test is a reasonable alternative that also has substantial advantages. Perhaps the greatest among these is the advantage of computational ease and the implementation the test in all major statistical packages, including the computation of confidence bounds for effects and adjustment for covariates. While we offer the program for the calculation of the index and p-value (see supplemental appendix), the extension of the approach for these extensions of the technique will both require modest additional programming and will also be more computationally intensive (noting, however, that computational burden is becoming less of an issue). Others, including coauthors of this work (VWR), are developing methods for re-expression of measures of association from the Mann-Whitney U to include NNT and measures of effect; however, we hope that the measure of association of the proposed approach is sufficiently straightforward to be easily understood. Hence, we suggest that the choice of the proposed method versus the alternative Mann Whitney U test is primarily a matter of personal preference that should focus on the benefit of having the basis for the measure of the association and p-value arising from the same metric versus the additional computational burden of the proposed approach.

One could easily (and perhaps correctly) argue that since the proposed approach, the POM, and the CMH are generally concordant, and the choice among them is not a major study design issue. However, the discussion in the neurological community regarding the strengths and weaknesses of the analytic approach for the mRS outcome seems to be a never-ending spirited debate. We do not claim that our approach is universally more powerful than the others, but rather that it is generally as powerful but superior because it avoids the pitfalls present in other approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

None.

Sources of Funding This work was supported by R01 NS055728 (David C Hess PI).

References

1. Saver JL, Gornbein J. Treatment effects for which shift or binary analyses are advantageous in acute stroke trials. Neurology. 2009; 72:1310–1315. [PubMed: 19092107]
2. Koziol JA, Feng AC. On the analysis and interpretation of outcome measures in stroke clinical trials: lessons from the SAINT I study of NXY-059 for acute ischemic stroke. Stroke. 2006; 37:2644–2647. [PubMed: 16946150]
3. Saver JL, Gornbein J, Starkman S. Graphic reanalysis of the two NINDS-tPA trials confirms substantial treatment benefit. Stroke. 2010; 41:2381–2390. [PubMed: 20829518]
4. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. Stroke. 2007; 38:3055–3062. [PubMed: 17916765]
5. Savitz SI, Benatar M, Saver JL, Fisher M. Outcome analysis in clinical trial design for acute stroke: physicians' attitudes and choices. Cerebrovasc Dis. 2008; 26:156–162. [PubMed: 18560219]
6. Savitz SI, Lew R, Bluhmki E, Hacke W, Fisher M. Shift analysis versus dichotomization of the modified Rankin scale outcome scores in the NINDS and ECASS-II trials. Stroke. 2007; 38:3205–3212. [PubMed: 17975102]
7. Hill MD, Martin RH, Palesch YY, Tamariz D, Waldman BD, Ryckborst KJ, et al. The Albumin in Acute Stroke Part 1 Trial: An Exploratory Efficacy Analysis. Stroke. 2011; 42:1621–25. [PubMed: 21546491]
8. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, et al. The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial. Lancet Neurol. 2009; 8:434–440. [PubMed: 19297248]
9. Hosmer, DW.; Lemeshow, S. Applied Logistic Regression. 2nd ed. John Wiley & Sons; New York, NY: 2000. p. 288-304.
10. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol. 1994; 47:881–889. [PubMed: 7730891]
11. Agresti, A. Analysis of ordinal categorical data. 2nd ed. John Wiley & Sons; New York, NY: 2010. p. 165-176.
12. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, Grotta J, Lyden P, Shuaib A, Hårdemark HG, Wasiewski WW. NXY-059 for acute ischemic stroke. N Engl J Med. 2006; 354:588–600. [PubMed: 16467546]
13. The National Institute of Neurological Disorders and Stroke rt-PA stroke study group. Tissue plasminogen activator for acute stroke. N Engl J Med. 1995; 333:1581–1587. [PubMed: 7477192]
14. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist. 1947; 22:125–128.
15. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull. 1945; 1:80–83.
16. Siegel, S. Nonparametric statistics for the behavioral sciences. McGraw-Hill Book Company; New York: 1956. p. 116-127.
17. Conover, WJ. Practical nonparametric statistics. John Wiley and Sons; New York: 1971. p. 216-223.
18. Fisher, RA. The design of experiments. 8th edition. Hafner; New York: 1966. p. 11-26.
19. Ludbrook J, Dudley H. Why permutation tests are superior to t and F tests in biomedical research. American Statistician. 1998; 52:127–132.
20. Lampl Y, Boaz M, Gilad R, Lorberboym M, Dabby R, Rapoport A, et al. Minocyline treatment in acute stroke: an open label, evaluator-blinded study. Neurology. 2007; 69:1404–1410. [PubMed: 17909152]

Table 1

The modified Rankin scale

Rankin Score	Clinical Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs 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

Table 2

Distribution of mRS for the NINDS-tPA trial,[3] showing the number of pairs of patients with each combination of mRS. For example, there were 29 tPA treated patients with a mRS of 3, and 30 placebo treated patients with a mRS of 2; hence, there are 29 × 30 = 870 pairs of patients with this combination of scores. There are 30,178 pairs of patients with higher (worse) mRS for than for placebo (light gray shade), 46,016 pairs of patients with higher scores for placebo than for tPA (unshaded), and 19,596 pairs of patients with a tie between placebo and tPA (black with white font)

		mRS for placebo-treated patients						
		0 (n = 23)	**1 (n = 31)**	**2 (n = 30)**	**3 (n =29)**	**4 (n = 90)**	**5 or 6 (n =106)**	
mRS for tPA-treated patients	0 (n = 49)	1127	1519	1470	1421	4410	5194	Total number of pairs with higher score for placebo = 46,016
	1 (n = 52)	1196	1612	1560	1508	4680	5512	
	2 (n = 26)	598	806	780	754	2340	2756	
	3 (n = 29)	667	899	870	841	2610	3074	
	4 (n = 68)	1564	2108	2040	1972	6120	7208	
	5 or 6 (n = 86)	1978	2666	2580	2494	7740	9116	
		Total number of pairs with higher score for tPA than placebo = 30,178						Total number of ties = 19,596

Table 3

Statistical power to detect differences between a standard and investigational treatment as a function of the difference between treatment groups (shift) and sample size for the Mann-Whitney U (MWU) test, proportional odds model (POM) and Cochran-Mantel-Haenszel (CMH) test. Because this table represents multiple replications, where within each replication multiple permutations are required, we note that for this table we used the MWU approximation to the proposed test

	Shift	N=500			N=1000			N=1500		
		MWU	POM	CMH	MWU	POM	CMH	MWU	POM	CMH
Smaller treatment effects	**10**	8.9	8.9	8.3	10.8	10.8	9.2	15.3	15.3	12.1
	15	10.2	10.2	9.9	18.2	18.2	14.5	27.0	27.0	21.4
	20	20.2	20.2	17.6	31.1	31.2	25.9	43.5	43.5	35.9
	25	24.5	24.5	20.1	44.6	44.7	36.8	62.4	62.4	52.0
	30	34.4	34.6	28.5	57.6	57.6	49.2	76.0	76.1	66.7
	35	41.7	41.7	35.1	72.1	72.1	62.8	89.0	89.0	82.4
	40	53.5	53.6	45.1	84.3	84.3	73.1	95.7	95.7	90.2
	45	63.4	63.4	54.2	90.6	90.6	82.4	98.5	98.5	94.9
	50	73.7	73.7	64.2	96.1	96.1	92.9	99.5	99.5	98.5

Table 4

Performance of the dichotomous analysis (dichotomized at 0-1 versus 2-6), the Cochran-Mantel-Haenszel (CMH) test, proportional odds model (POM), and proposed model. For dichotomous analyses: A =active treatment (investigational treatment), P=placebo. For the proportional odds model: the test of proportional odds is the probability that the assumption of proportional odds is met in the study

	Dichotomous Outcome		Ordinal outcome Approach								
			CMH Test	Proportional Odds Test			Proposed Approach (with Permutation/Randomization Test)				
	% mRS Score of 0/1	Treatment p-value	CMH P-Value	Test of Proportional Odds Assumption	Treatment Difference P-value	Odds Ratio (95% CI)	Proportion Better	Proportion Worse	Proportion Same	Proportion of changes better for active treatment	Treatment Difference p-value
tPA[3, 17] (n = 619)	A: 33% P: 17%	<0.0001	<0.0001	0.14	0.0002	1.69 (1.28 - 2.25)	48.0%	31.5%	20.5%	60.4%	0.0001
SAINT 1[12] (n = 1,699)	A: 33% P: 31%	0.28	0.138	0.059	0.153	1.13 (0.96 - 1.34)	43.1%	39.2%	17.7%	52.4%	0.151
Alias 1[7] (n = 257)	A: 28% P: 37%	0.16	0.654	0.43	0.625	1.11 (0.72 - 1.71)	43.4%	40.0%	16.6%	52.1%	0.626
Minocycline Trial[20] (n = 151)	A: 80% P: 25%	<0.0001	<0.0001	0.78	<0.0001	13.16 (6.54 - 27.03)	76.7%	9.4%	13.9%	89.1%	<0.0001