



Published in final edited form as:

Int J Stroke. 2014 February ; 9(2): 226–231. doi:10.1111/ijss.12052.

Regression Analysis of Ordinal Stroke Clinical Trial Outcomes: An Application to the NINDS t-PA Trial

Stacia M DeSantis, PhD¹, Christos Lazaridis, MD², Yuko Palesch, PhD¹, and Ramesh Ramakrishnan, PhD¹

¹Department of Medicine, 135 Cannon St, Suite 303, Charleston, SC 29425

²Department of Neurosciences, Division of Vascular Neurology and Neurocritical Care, 96 Jonathan Lucas St, STE 428, Charleston, SC 29425

Abstract

Background—The modified Rankin scale (mRS) is the most common functional outcome assessed in stroke trials. The proportional odds model is commonly used to analyze this ordinal outcome but it requires a restrictive assumption that a single odds ratio applies across the entire outcome scale.

Aims—To model the effect of tissue-type plasminogen activator on ordinal mRS, test model assumptions, and compare fits and predictive ability of the statistical models.

Methods—Several ordinal regression methods are presented and applied to a re-analysis of the 1995 NINDS tissue-type plasminogen activator study. Violations of the proportional odds assumption are demonstrated using graphs and statistical tests, and the partial proportional odds model is introduced and recommended as an alternative for the analysis of mRS.

Results—The partial proportional odds model relaxes the assumptions about treatment effect on the ordinal outcome scale and provides a better fit to the data than the commonly used proportional odds model (likelihood ratio test chi square = 8.05, $p=0.005$). It provides easily interpretable odds ratios and it is able to detect efficacy at the lower end and a lack of efficacy at the upper end of the mRS scale. Further, it provides lower prediction error than the proportional odds model (0.002 versus 0.005).

Conclusions—Assuming proportional odds when it does not hold can mask differential treatment effects at the upper end of the ordinal mRS scale and has implications for reduced power when studies are designed under this assumption.

Corresponding Author: Stacia M. DeSantis, Phone: 843-876-1593, Fax: 843-876-1126, desantis@musc.edu.

Conflicts of Interest: None declared

Author Contributions

Stacia M. DeSantis – Study concept and design, acquisition of data, analysis and interpretation, critical revision of the manuscript

Christos Lazaridis – Study concept and design, interpretation, critical revision of the manuscript

Yuko Palesch – Interpretation, critical revision of the manuscript

Ramesh Ramakrishnan – Study concept and design, analysis and interpretation, critical revision of the manuscript

Keywords

tissue plasminogen activator; ordinal regression; modified Rankin scale; proportional odds; partial proportional odds; clinical trial

Introduction

The modified Rankin Scale (mRS), one of the commonly used outcome measure in acute stroke trials, is a 7-point ordinal measure ranging from 0 (no symptoms) to 6 (dead) (1-3). Although acute neurological clinical trials collect ordinal outcome data, trials are usually designed and analyzed based on a dichotomized outcome obtained by collapsing them into “good” and “bad” categories (4). This does not allow for the examination of treatment effects at finer gradations of the scale and in some situations decreases the statistical power of the study. A meta-analysis by the Optimising Analysis of Stroke Trials (OAST) Collaboration showed that statistical approaches that analyze the data using the ordinal functional outcomes in their original form are more efficient than those applied to preprocessed data that do not exploit the ordinality (4). Specifically, when assessed by how many trials were statistically significant, those tests which do not collapse the data into groups out-performed the other approaches (i.e., 26% of non-collapsed versus 9% of collapsed trials were significant).

Recently, alternatives have been proposed using the full ordinal scale in the analysis under the assumption of proportional odds (1, 4-6). Under the proportional odds assumption the odds ratio comparing t-PA to placebo in patients with mRS of 0 versus 1 – 6, then 0 – 1 versus 2 – 6, and so on, are assumed to be the same. The analysis under this assumption is performed by fitting a model to the cumulative logits, called the proportional odds model (POM). If the assumption of proportional odds holds, fitting the POM is parsimonious and does not require a strict dichotomy based on an arbitrary cut off, and can increase statistical power over a dichotomous analysis. However, if the proportional odds assumption fails to hold, this analysis has the capacity to mask a lack of or harmful effects at one end of the ordinal outcome spectrum.

The statistical test for verifying the assumption of proportional odds (score test) is not well-powered (7). Consequently, the justifications for using the POM are not satisfactory. Consider the data from the NINDS t-PA trial (8). The score test for proportional odds results in a p-value slightly above the 5% significance level (p-value = 0.06). In Figure 1 the cumulative log odds of each mRS score for t-PA versus placebo are shown. The difference at each point on the ordinal scale (for each value on the x-axis) is equivalent to the log odds ratio. If the proportional odds assumption held, the line for t-PA would be parallel to the line for the placebo indicating a constant difference in the cumulative log odds. However, since the lines intersect, the assumption of proportional odds may be inappropriate. In such cases, alternative approaches that use the entire spectrum of the ordinal mRS scale should be considered.

Several acute stroke trials such as the SAINT I and II pooled analyses have utilized assumption free ordinal tests such as Cochran Mantel Haenzel and van Elteren test for

stratified data that use the whole distribution of mRS and avoid the potential issue of non proportional odds (e.g., 9,10), and are well-established alternatives to proportional odds models. These authors have also drawn attention to the different results obtained by using assumption free tests versus those based on the proportional odds model. In fact, the ESO Working Group on Outcomes distinguishes between statistical testing (not necessarily invoking the proportional odds assumption) and the expression of effect size (perhaps with common odds ratio). Howard et al. discuss the violation of the assumption of proportional odds in the context of the NINDS t-PA and SAINT 1 trials for acute stroke, and propose an alternative nonparametric permutation test for the analysis of the ordinal mRS that circumvents the need for modeling and hence any assumptions regarding proportional odds (11). However, the assumption free and non parametric approaches do not result in odds ratios, which are commonly understood by a clinical audience and in addition, these approaches do not easily allow one to control for covariates, which can be important if there are baseline group imbalances.

This article proposes a model-based alternative for analyzing ordinal outcomes that does not require the proportional odds assumption. The approach is based on a well-established theoretical framework in the statistical analysis of categorical data (12-14). Applications of alternative ordinal regression models have been mostly absent from the acute stroke trials literature. The purpose of this article is therefore to introduce commonly used and alternative approaches to analyzing the mRS using the entire ordinal scale, illustrate their use on stroke data, and provide various presentation methods to guide the assumptions and interpretations of these models. Specifically, two alternative logistic regression models that have been widely studied in the statistical literature, namely the partial proportional odds model (PPOM) and the adjacent categories logit (ACAT) model, which does not require strong assumptions, will be introduced (12, 13). The NINDS t-PA trial data will be used to illustrate them.

METHODS

Trial Data

The National Institute of Neurological Disorders and Stroke Tissue Plasminogen Activator trial was performed in 1995 to test effectiveness of t-PA versus placebo on four outcomes - the Barthel index, modified Rankin scale, Glasgow outcome scale, and National Institutes of Health Stroke Scale (NIHSS) in 624 patients at 3 months. The primary analysis using a generalized estimating equations approach resulted in a significant global test score for the four outcomes at 3 months (15). Results indicated a sustained improvement in t-PA versus placebo groups and no difference in mortality across groups. Specifically, much attention has been paid to analysis of the mRS outcome. The ordinal regression analyses of mRS presented in this paper are based on 581 completers of the study (those who had mRS data at 3 months).

Proportional Odds Model

The proportional odds model is a commonly used generalization of logistic regression, where a single odds ratio (OR) is calculated by summarizing ORs that are cumulatively

calculated over the mRS scale. The summary OR is a clinically interpretable parameter that represents “the odds of a higher versus lower mRS score in the t-PA versus placebo group” that holds for any cut point of the mRS scale. The model is favorable for presentation because it parsimoniously describes the data with only a single odds ratio but is meaningful only if the proportional odds assumption is valid.

The proportional odds assumption is often tested via a global score test that is routinely produced by statistical software (16, 17). However, this test lacks power and therefore often fails to reject when the alternative hypothesis of non-proportional odds is true (7), for example, when the non-proportionality exists for some portion of the scale, as illustrated for the t-PA data in Figure 1. Often, score test results are not reported in clinical journals, irrespective of their significance (18).

Partial Proportional Odds Model

The partial proportional odds model (PPOM) can be used in two situations. The first situation is where one covariate but not another meets the assumption of proportional odds (7, 13, 19). The second situation, which is more relevant for the NINDS t-PA data, is where a “linear deviation” from the proportional odds assumption occurs as one moves up the mRS scale (13, 14). Linear deviation simply means a violation of the proportional odds assumption in one direction. To adjust the POM for this scenario, an additional term using a second parameter that allows for the odds ratios to increase proportional to the outcome scale is introduced. The PPOM is a special and more parsimonious case of a fully (non parsimonious) nonproportional odds model. Specifically, in the current case, it addresses the situation where the cumulative logits cross, as shown in Figure 1. The appropriateness of the PPOM model over the POM is formally tested either by testing the significance of the added (second) parameter (using a Wald test), or equivalently by using a one degree of freedom likelihood ratio test (LRT). A significant p-value on these tests provides evidence against the proportional odds, favoring evidence of a linear deviation from proportional odds.

Adjacent Categories Logit Model

Like the POM and PPOM, the adjacent categories logit model (ACAT) assumes there is an inherent ordering of responses but does not require the assumption of proportional odds. Instead, separate odds ratios are simultaneously calculated for each adjacent category of response in relation to covariates. This approach is natural when one wants to describe treatment effects in terms of odds relating to particular response categories (13). For example, one may be interested in comparing how well the treatment works for those in mRS category 1 in reference to 0, or category 2 in reference to 1, and so on. Essentially, this could be used as a surrogate for measuring the effect of the treatment in reducing patients’ mRS score by one category. Further, the entire matrix of pairwise ORs can be computed under this model, which allows for comparisons of reduction in mRS of two or more categories. Unlike the POM, the ACAT model makes no additional assumptions about the mRS response scale beyond ordinality, consequently the model has the disadvantage of being less parsimonious (requiring the interpretation of 6 different odds ratios for each adjacent category of the 7 point ordinal scale). However, it is a well-studied alternative approach for situations where neither the POM nor PPOM fit the data well. A reduced

ACAT model that assumes the same linear logit effect applies simultaneously for all pairs of adjacent categories is often fit as a more parsimonious form of the ACAT. The ACAT model and the POM or PPOM cannot be compared by likelihood ratio tests, and therefore the adequacy of ACAT is determined by comparing information criteria (such as Akaike information criteria, AIC, and Bayesian information criteria, BIC) that measure the deviation of unknown “true” model, from the fitted model. Smaller values of AIC and BIC indicate better fitting models.

Model Comparison

Model comparison is based on both the fit as well as the predictive ability of the models. This is presented through graphs, through formal likelihood ratio hypothesis tests (for nested models) and AIC and BIC (for non-nested models), and through the prediction error estimated by the square of the difference between the observed probability and the estimated probability. Odds ratios and 95% confidence intervals (CI) are presented as OR [95% CI]. Odds ratios are also transformed to the probability scale for ease of interpretation.

Results

The anti-conservative score test for the proportional odds assumption resulted in a borderline p-value ($p=0.06$). Further, it is apparent from Figure 1 that the assumption is not appropriate. To further substantiate this, odds ratios for all collapsed mRS categories are shown in Figure 2. If the assumption of proportional odds held, the odds ratios would be constant across all possible mRS categorizations. However it is clear from Figure 2 that the odds ratios show an increasing trend across the categories (with ORs increasing from 0.51 to 0.97), indicating that they are not constant and therefore a summary odds ratio produced under the POM is questionable.

In Table 1 the odds ratios with 95% CI resulting from the POM and PPOM models are presented. The summary odds ratio from the POM indicates the odds of a higher mRS in the t-PA versus placebo group is 0.81 [0.71, 0.94], demonstrating a protective effect of t-PA. Via the introduction of a second parameter for linear deviation, the PPOM model results in the six odds ratios in Table 1, corresponding to the six possible categorizations. The PPOM odds ratios are closer to the observed odds ratios shown in Figure 2 than the POM odds ratio. The PPOM results indicate that the t-PA has a significant benefit irrespective of whether 0, or 0-1 or 0-2 is defined as the “favorable” outcomes to dichotomize the data. These dichotomizations are widely discussed in the literature, in the context of responder analyses or ‘sliding dichotomy’ (20).

The POM and PPOM are formally compared using likelihood ratio tests (Table 2). Since there is only one additional parameter in PPOM, the test statistic is a chi-square with one degree of freedom (df). The chi-square value is 8.05(df=1) ($p = 0.005$), which is highly significant at 5% level indicating PPOM provides a significantly better fit than the POM. Thus the PPOM is better able to replicate the true underlying treatment effects.

The fitted probabilities from a well-fitting model should replicate the observed probabilities. Figure 3 displays the observed and fitted probabilities of being in each mRS category at 3

months for the placebo and t-PA groups. Observed probabilities are calculated from the actual dataset while fitted probabilities are calculated from fitting the respective models to the data set. Figure 3 indicates that the fitted probabilities obtained from the PPOM better replicate the observed (true) data. This is further substantiated by the smaller prediction errors resulting from the PPOM (0.002 versus 0.005). The reason for the better fit of the PPOM is that in order to exploit the proportional odds assumption of the POM for the t-PA data, the fitted probabilities for t-PA and placebo groups are forced by the POM to be closer together across the spectrum of mRS as compared to the observed probabilities, resulting in a substantial overestimation of a treatment effect at the higher end of the spectrum.

The ACAT odds ratios for adjacent categories are presented at the bottom of Table 1. In terms of the AIC and BIC (Table 2), the ACAT is comparable to the POM, which is not surprising since AIC penalizes for the large number of parameters in the ACAT model. The ACAT model results are presented below. The easily interpretable category-wise odds ratios offer some additional clinical insights. Although the study was not powered under an ACAT model, given the current sample size, the only significant odds ratio under this model is the one comparing mRS category 2 to category 1 (OR=0.42 [0.22, 0.79]). One could interpret this to suggest that the most relevant impact of t-PA is to result in the reduction in odds of observing a category 2 versus category 1 on the mRS scale at 90 days, which is arguably one of the most important mRS contrasts. The effects of t-PA versus placebo for other adjacent categories are insignificant, demonstrating no effect of t-PA for some pairwise comparisons. The pairwise odds ratios resulting from this adjacent categories regression confirm a reversal in direction when comparing categories at the higher end of the mRS spectrum. In Table 3, the upper-triangle of the entire symmetric matrix of pairwise comparisons is presented for completeness. The relative merits of this model is that it has fewer assumptions and pairwise comparisons can be represented in matrix form; however, it is underpowered in most scenarios.

In the interest of parsimony, a reduced ACAT model was fit, which assumes the same effect applies to all pairs of adjacent categories was also fit to the data. This model also has only one odds ratio describing effects of treatment on adjacent categories (e.g., one OR for 1 vs. 0, 2 vs. 1, etc). The log likelihood for this model is approximately the same as that for the POM but the AIC and BIC are larger indicating that this model does not provide a better fit to the data than any of the other models considered in Table 2. This is not surprising since one summary odds ratio for all adjacent categories is impractical since these ORs are not in the same direction.

Conclusion

In most stroke trials, ordinal outcomes are analyzed by dichotomizing the outcome (21). Dichotomizing discards information and therefore could be less powerful than the methods that model the probabilities using the whole spectrum, under appropriate assumptions. For outcomes measured on an ordinal scale, when the assumption of proportionality is not violated, the proportional odds model analysis is the most powerful and recommended option in addition to or in place of the use of assumption free tests (9-11). We note that the assumption was not violated in the majority of acute stroke trials (1). Under appropriate

assumptions the POM, in straightforward trial design approaches, results in an easily interpretable common odds ratio, and a tighter confidence interval than other ordinal methods or approaches that dichotomize the ordinal scale. However, any assumptions made in the analysis should be scrutinized carefully using appropriate hypothesis tests and diagnostic tools. The acceptance of the proportional odds assumption in the NINDS t-PA trial was based on a marginal significance of the score test that rightly drew criticism in Howard et al (11). Authors in other clinical research fields have also cautioned against the use of the proportional odds model when the assumption of proportionality is violated (18, 22, 23). The violation of this assumption was supported in this paper by the graphical illustrations of a deviation from proportional odds. Several alternatives to proportional odds model, such as the sliding dichotomy based on a baseline covariate (20), assumption free tests (10), or the randomization test (11), have been proposed in the literature. While these approaches address the non-proportionality of the odds, they fail to exploit all the information provided by the ordinal scale of the outcome measure, which is not optimal in all applications (4).

Using the analysis of the NINDS r-tPA study the article demonstrates that fitting a proportional odds model and reporting a summary OR can potentially mask contradictory effects observed at the upper end of the mRS scale. Alternatively, partial proportional odds models and adjacent category logit models should be considered. For the current data, these alternatives are demonstrated to fit better and indicate a significant effect of t-PA at the lower end of the mRS scale but a diminishing effect of t-PA at the upper end of the mRS scale that was undetectable under other approaches. Moreover, the adjacent categories logit model allows one to fill in a table of odds ratios for every possible category-wise mRS comparison, which is helpful for clinicians wishing to risk-stratify patients.

The results presented in this paper are potentially consistent with prior evidence demonstrating intravenous t-PA to be less effective in patients with severe baseline deficits and high NIHSS; the one year outcome data of the 1995 NINDS study (8) found a benefit for t-PA versus placebo, with one exception, namely the subgroup of patients with NIHSS scores of more than 20 at baseline (24). Patients with high NIHSS often have proximal vessel occlusions (carotid artery or proximal middle cerebral artery) known to be less susceptible to the thrombolytic effect of intravenous t-PA (25). The alternative approaches presented in this paper provide numerical quantification of the efficacy of t-PA among the different mRS groups via the application of an appropriate statistical model. This is not meant to discourage consideration of the use of t-PA to eligible patients even in the higher spectrum of the NIHSS scale, but to better inform clinical decision making and future clinical trial design and interpretation. We emphasize that results of any model must be interpreted considering the clinical importance of the mRS grades. The difference between mRS of 0 versus 1 could be quite different in societal terms from the difference between mRS of 4 versus 5 thus it should be underscored that under the PPOM, t-PA is effective in the most clinically important region of the scale.

Due to the possibility of misclassification of the ordinal mRS scale after treatment, it is arguable that the average treatment effects for the sample or even treatment effects derived from dichotomized outcomes could be more relevant. We note that ordinal regression has

also been extended to the setting of such misclassification error, and could therefore be considered if misclassification is a concern (26-27).

In light of the fact that researchers have recently proposed that design of acute stroke trials are based on ordinal outcomes the current message is timely (2,3). Specifically, Bath et al. (1) showed that sample size needed for ordinal outcome data are in general lower and therefore they argue that the whole spectrum should be considered, especially for acute stroke trial designs. Since such a design would subsequently necessitate the use of statistical approaches for ordinal outcomes, it will be important to assess in the design stage whether the proportional odds assumption is expected to hold after the data have been collected. The reason is that if a study is powered on an ordinal outcome assuming proportional odds, the assumption would have to hold for the treatment and for all covariates in the regression model. If the assumption consequently proves to be wrong for some or all covariates, power will be sacrificed and therefore the study may not lead to definitive conclusions. Thus to specify the mRS scale (or another ordinal scale such as the Barthel index or Glasgow Outcome Scale) as a primary outcome in acute stroke trials, there must be strong a priori insight on the legitimacy of the proportional odds assumption. The alternative proposed in this article, the PPOM could also be considered at the design stage, and it would also be straightforward to develop simulation-based power analyses under different scenarios to determine the sample size that adequately powers the study for the ordinal analysis of interest. Currently, sample size under a proportional odds model is obtained by specifying the single odds ratio (often from preliminary data) and estimating the treatment group-specific proportions from that odds ratio (i.e., marginal cell probabilities), from which the sample sizes are estimated (28-29). In the case of PPOM, an odds ratio and a linear deviation from that odds ratio need to be specified from previous trials or clinical experience in acute stroke. From there, it would be similarly straightforward to calculate the expected proportions in each treatment group from which the sample sizes could be estimated.

Although only one example is presented here, it is important to note that the proportional odds assumption has been questionable in other acute neurological studies as well. For example, in the intent-to-treat analysis of the SAINT 1 trial, the p-value for the test of proportional odds was also marginal at 0.059. In light of this borderline finding, other diagnostic tools could have been used to scrutinize this assumption. In the current study, assumption-checking using a variety of plots are recommended and demonstrated to be useful in identifying the nature of the violation of the proportionality assumption.

REFERENCES

1. Bath PMW, Lees KR, Schellinger PD, Altman H, Bland M, Hogg C, et al. Statistical analysis of the primary outcome in acute stroke trials. *Stroke*. 2012; 43:1171–1178. [PubMed: 22426314]
2. Tilley BC. Contemporary outcome measures in acute stroke research. *Stroke*. 2012; 43:935–937. [PubMed: 22426318]
3. Lees KR, Bath PMW, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke*. 2012; 43:1163–1170. [PubMed: 22426313]
4. The Optimising Analysis of Stroke Trials (OAST) Collaboration. Can we improve the statistical analysis of stroke trials? Statistical re-analysis of functional outcomes in stroke trials. *Stroke*. 2007; 38:1911–1915. [PubMed: 17463316]

5. Brown DL, Coffey CS. Stroke trials: A shift to shift analysis? *Neurology*. 2009; 72:1292–1293. [PubMed: 19365050]
6. 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]
7. Peterson B, Harrell FE. Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society Series C*. 1990; 39(2):205–217.
8. The National Institute of Neurological Disorders and Stroke rtPA Stroke Study Group. Tissue plasminogen activator for acute stroke. *N Engl J Med*. 1995; 333:1581–1587. [PubMed: 7477192]
9. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Dienne H-C, et al. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006; 354:588–600. [PubMed: 16467546]
10. Diener HC, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM, Shuaib A, Ashwood T, Wasiewski W, Alderfer V, Hardemark HG, Rodichok L. NYX-059 for the treatment of acute stroke: Pooled analysis of the SAINT I and II trials. *Stroke*. 2008; 39(6):1751–8. [PubMed: 18369171]
11. Howard G, Waller JL, Voeks JH, Howard VJ, Jauch EC, Lees KR, et al. A simple, assumption-free, and clinically interpretable approach for analysis of modified Rankin outcomes. *Stroke*. 2012; 43(3):664–9. [PubMed: 22343650]
12. Agresti, A. *Categorical Data Analysis*. 2nd edition. Wiley; 1990.
13. Agresti, A. *Analysis of Ordinal Categorical Data*. 2nd edition. Wiley; 2010.
14. Agresti, A. *Examples of Using R for Modeling Ordinal Data, Supplement for Analysis of Ordinal Categorical Data*. 2nd edition. Wiley; 2010.
15. Tilley B, Marler J, Geller N, Lu M, Legler J, Brott T, et al. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke tPA trial. *Stroke*. 1996; 27:2136–2142. [PubMed: 8898828]
16. Rao CR. Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proc. Camb. Phil. Soc.* 1973; 44:50–57. *Linear Statistical Inference and its Applications*. 2nd edn. Wiley; New York:
17. McCullagh P. Regression models for ordinal data (with Discussion). *JRRS B*. 1980; 42:109–142.
18. Bender R, Grouven U. Using Binary Logistic Regression Models for Ordinal Data with Non-proportional Odds. *J Clin Epi*. 1998; 51(10):809–816.
19. Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rtPA stroke trial. *Stat Med*. 2010; 29:546–557. [PubMed: 19943331]
20. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees RK, et al. Design and Analysis of Phase III Trials with Ordered Outcome Scales: The Concept of the Sliding Dichotomy. *J Neurotrauma*. 2005; 22(5):511–517. [PubMed: 15892597]
21. Wardlaw J, Sandercock P, Warlow C, Lindley RI. Trials of thrombolysis in acute ischemic stroke: does the choice of primary outcome measure really matter? *Stroke*. 2000; 31:1133–1135. [PubMed: 10797177]
22. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epi*. 1997; 50:45–55.
23. Liu I, Mukherjee B, Suesse T, Sparrow D, Park SK. Graphical diagnostics to check model misspecification for the proportional odds regression model. *Stat Med*. 2009; 28:412–429. [PubMed: 18693299]
24. Kwiatowski TG, Libman RB, Frankel M, Tilley BC, Morgenstern LB, Lu M, et al. Effects of Tissue Plasminogen Activator for Acute Ischemic Stroke at One Year. *N Engl J Med*. 1999; 340:1781–1787. [PubMed: 10362821]
25. del Zoppo GJ, Poeck K, Pessin MS, Wolpert SM, Furlan AJ, Ferbert A, et al. Recombinant tissue plasminogen activator in acute thrombotic and embolic stroke. *Ann Neurol*. 1992; 32(1):78–86. [PubMed: 1642475]
26. Mwalili SM, Lesaffre E, Declerck D. A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographic oral health study. *App Stat*. 2005; 54(1):77–93.

27. Shirkey, BA. An Ordinal Logistic Regression Model with Misclassification of the Outcome Variable and Categorical Covariate. ProQuest; 2009. University of Texas School of Public Health Dissertation
28. Whitehead J. Sample size calculations for ordered categorical data. *Stat Med.* 1993; 12:2257–2271. 1993. published erratum appears in *Stat Med* 1994; 13(8):871. [PubMed: 8134732]
29. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ.* 1995; 311(7013):1145–1148. [PubMed: 7580713]

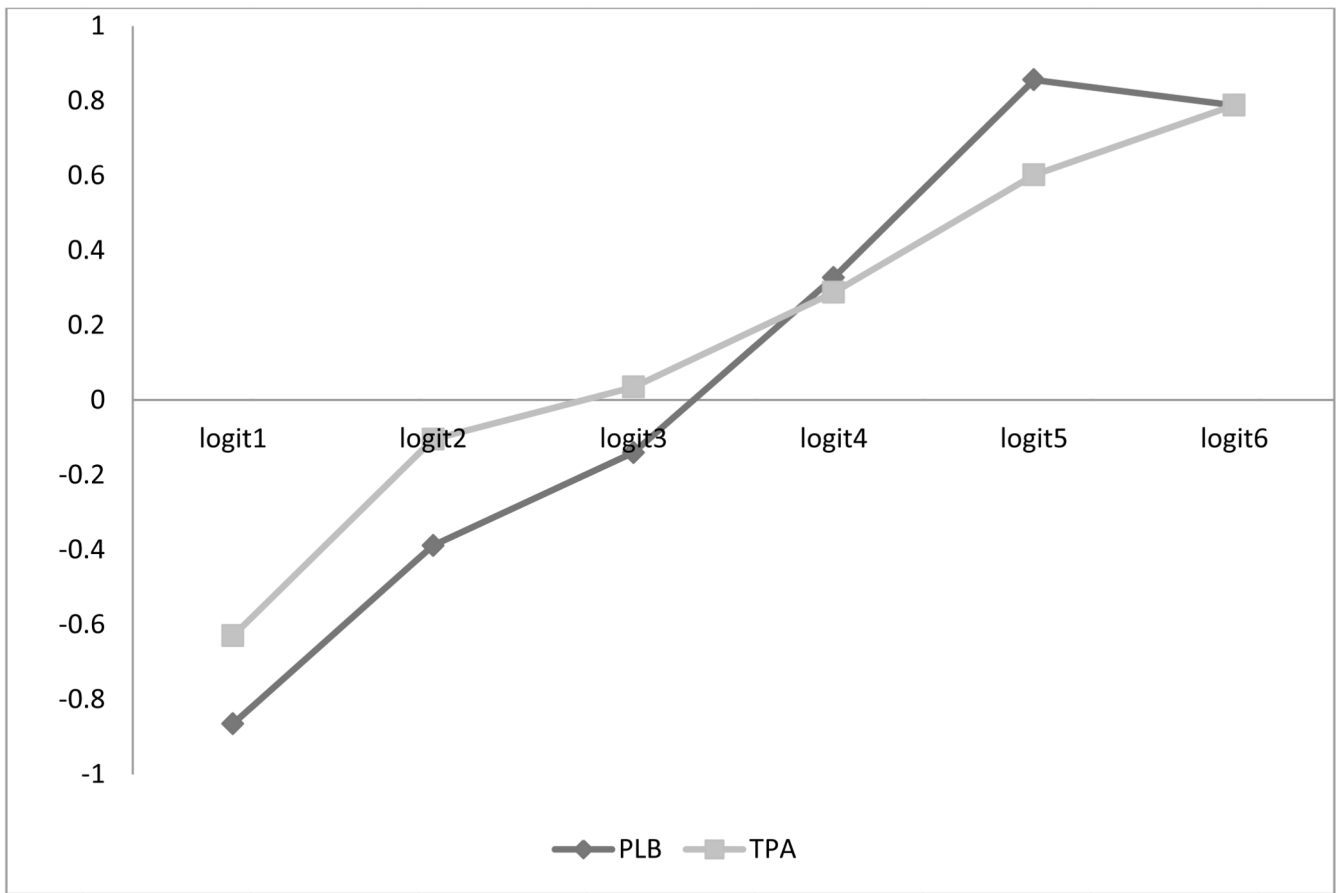


Figure 1. Cumulative log odds for the t-PA and Placebo (PLB) groups indicating a violation of the proportional odds assumption

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Odds Ratios

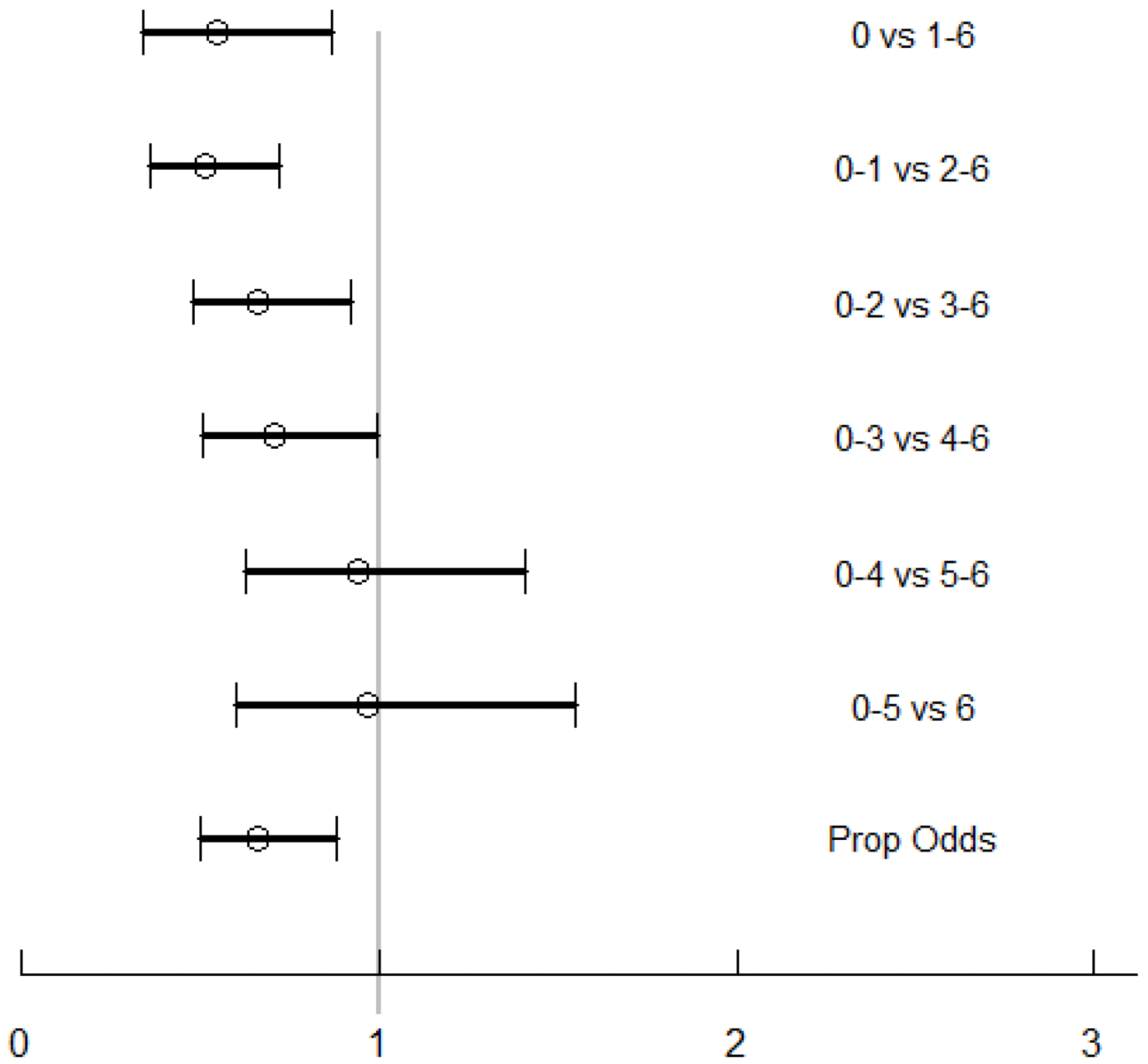


Figure 2. Odds ratios for and 95% confidence intervals (t-PA versus placebo) for individual logistic regressions and the proportional odds model

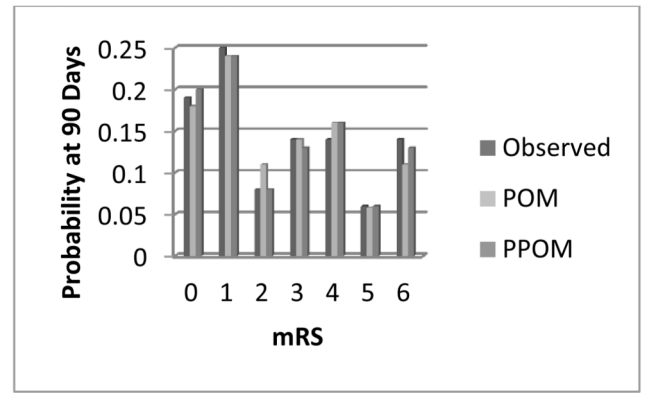
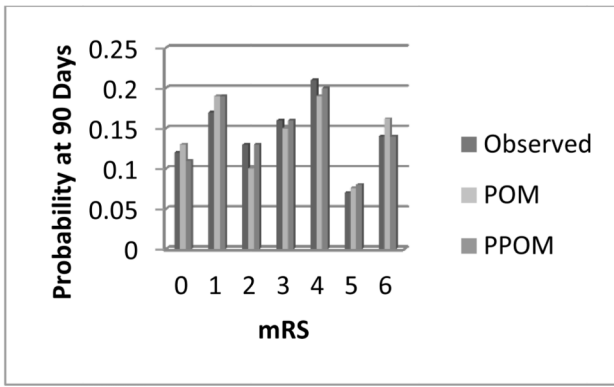


Figure 3. Observed, POM, and PPOM probabilities for each mRS category at 90 days for the placebo group (left) and t-PA group (right). The PPOM better replicates the observed probabilities (prediction error = 0.002 for the PPOM versus 0.005 for the POM)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Odds ratio [95% CIs] for the three ordinal regression models

	OR	95% CI
POM	0.81	[0.81,0.94]
Partial POM		
1-6 vs 0	0.48	[0.33,0.71]
2-6 vs 0-1	0.56	[0.41,0.78]
3-6 vs 0-2	0.64	[0.48,0.86]
4-6 vs 0-3	0.74	[0.55,1.01]
5-6 vs 0-4	0.86	[0.61,1.22]
6 vs 0-5	0.99	[0.65,1.51]
ACAT		
1 vs 0	0.86	[0.49,1.50]
2 vs 1	0.42	[0.22,0.79]
3 vs 2	1.43	[0.73,2.80]
4 vs 3	0.78	[0.43,1.38]
5 vs 4	1.38	[0.66,2.89]
6 vs 5	1.05	[0.49,2.64]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Model comparison for 4 ordinal models. A significant value indicates the model with more parameters (listed first) is preferred

	<u>-2Log L</u>	<u>Number of parameters</u>	<u>Prediction Error</u>	<u>AIC</u>	<u>BIC</u>
POM	2194.4	1	0.005	2208.4	2238.6
PPOM	2186.3	2	0.001	2202.3	2237.2
ACAT	2184.1	6	0	2208.1	2260.5
Reduced ACAT	2195.5	1	0.005	2209.5	2240.1
Likelihood Ratio Test	<u>Chi Square</u>	<u>DFs</u>	<u>P-value</u>		
Partial vs POM	8.05	1	0.005		
ACAT vs mod Acat	11.42	5	0.044		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3
Pairwise ORs for mRS categories resulting from the adjacent categories logit model

0	1	2	3	4	5	6
0	0.86	0.36	0.51	0.40	0.55	0.58
1	1.00	0.42	0.61	0.47	0.64	0.68
2		1.00	1.43	1.11	1.53	1.61
3			1.00	0.78	1.07	1.13
4				1.00	1.38	1.45
5					1.00	1.05
6						1.00

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript