Garofolo et al. Trials 2013, 14:98
http://www.trialsjournal.com/content/14/1/98

TRIALS

METHODOLOGY Open Access

The effect of covariate adjustment for baseline severity in acute stroke clinical trials with responder analysis outcomes

Kyra M Garofolo[1], Sharon D Yeatts[1], Viswanathan Ramakrishnan[1], Edward C Jauch[2], Karen C Johnston[3] and Valerie L Durkalski[1*]

Abstract

Background: Traditionally in acute stroke clinical trials, the primary clinical outcome employed is a dichotomized modified Rankin Scale (mRS). New statistical methods, such as responder analysis, are being used in stroke studies to address the concern that baseline prognostic variables, such as stroke severity, impact the likelihood of a successful outcome. Responder analysis allows the definition of success to vary according to baseline prognostic variables, producing a more clinically relevant insight into the actual effect of investigational treatments. It is unclear whether or not statistical analyses should adjust for prognostic variables when responder analysis is used, as the outcome already takes these prognostic variables into account. This research aims to investigate the effect of covariate adjustment in the responder analysis framework in order to determine the appropriate analytic method.

Methods: Using a current stroke clinical trial and its pilot studies to guide simulation parameters, 1,000 clinical trials were simulated at varying sample sizes under several treatment effects to assess power and type I error. Covariate-adjusted and unadjusted logistic regressions were used to estimate the treatment effect under each scenario. In the case of covariate-adjusted logistic regression, the trichotomized National Institute of Health Stroke Scale (NIHSS) was used in adjustment.

Results: Under various treatment effect settings, the operating characteristics of the unadjusted and adjusted analyses do not substantially differ. Power and type I error are preserved for both the unadjusted and adjusted analyses.

Conclusions: Our results suggest that, under the given treatment effect scenarios, the decision whether or not to adjust for baseline severity when using a responder analysis outcome should be guided by the needs of the study, as type I error rates and power do not appear to vary largely between the methods. These findings are applicable to stroke trials which use the mRS for the primary outcome, but also provide a broader insight into the analysis of binary outcomes that are defined based on baseline prognostic variables.

Trial registration: This research is part of the Stroke Hyperglycemia Insulin Network Effort (SHINE) trial, Identification Number NCT01369069.

Keywords: Responder analysis, Sliding dichotomy, Clinical trials, Acute stroke, Modified rankin scale, Baseline severity

* Correspondence: durkalsv@musc.edu
[1]Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425, USA
Full list of author information is available at the end of the article

BioMed Central

© 2013 Garofolo et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

Stroke is a potentially debilitating medical event that affects approximately 800,000 people in the United States each year, leaving as many as 30% of survivors permanently disabled [1]. Given this impact, there is great demand for treatments that significantly improve functional outcome following a stroke. To date, few clinical trials for the treatment of acute stroke have succeeded; of over 125 acute stroke clinical trials, only three successful treatment methods have been identified [2,3].

One of the possible reasons for the excessive number of neutral or unsuccessful stroke trials is the definition of successful outcome utilized in the studies [4]. In clinical trials, stroke outcome is most commonly measured by the modified Rankin Scale (mRS) of global disability at 90 days. The mRS is a valid and reliable measure of functional outcome following a stroke [5]. Past trials have dichotomized mRS scores into "success" and "failure", scores of 0 to 1 (or 0 to 2) were considered to be "successes" while scores greater than 1 (or 2) were considered to be "failures," regardless of baseline stroke severity [6-9]. This method fails to take into account the understanding that baseline severity is highly correlated with outcome. New methods, such as the global statistic, shift analysis, permutation testing and responder analysis, are evolving to make better use of the outcome data with the hopes of providing higher sensitivity to detect true treatment effects [2,4,6,9-17].

Responder analysis, also known as the sliding dichotomy, dichotomizes ordinal outcomes into "success" and "failure," but addresses the drawbacks of traditional dichotomization by allowing the definition of success to vary by baseline prognostic variables. Various trials have implemented the responder analysis where baseline severity is defined by one or many baseline prognostic factors [18-20]. Those study subjects in a less severe prognosis group at baseline must achieve a better outcome to be considered a trial "success," whereas a less stringent criterion for success is applied to subjects in a more severe baseline prognosis category. The currently enrolling Stroke Hyperglycemia Insulin Network Effort (SHINE) trial employs responder analysis for its primary efficacy outcome [18].

The SHINE trial is a large, multicenter, randomized clinical trial designed to determine the efficacy and safety of targeted glucose control in hyperglycemic acute ischemic stroke patients. While the methodological details of the SHINE trial are discussed elsewhere [18], it should be noted that the primary outcome for efficacy is the baseline severity adjusted 90-day mRS score dichotomized as "success" or "failure" according to a sliding dichotomy. Eligibility criteria for SHINE require that a subject's baseline NIHSS score must be between 3 and 22, inclusively. Those with a "mild" prognosis, defined by a baseline NIHSS score of 3 to 7, must achieve a 90-day mRS of 0 to be classified as a "success." Those with a "moderate" prognosis, defined by a baseline NIHSS score of 8 to 14, must achieve a 90-day mRS of 0 to 1 to be classified as a "success." Finally, those subjects with a "severe" prognosis, defined by a baseline NIHSS score of 15 to 22, must achieve a 90-day mRS of 0 to 2 to be classified as a "success." By using responder analysis with a trichotomized NIHSS, the threshold for success is stringent for the milder strokes, while the moderate to severe strokes are allowed to have more residual deficits in the threshold for success.

One of the questions that arose from the trial's Data and Safety Monitoring Board was that of covariate adjustment. Statistical analyses often adjust for prognostic factors, or covariates, that may be predictive of the primary outcome, such as baseline severity [21,22]; however, in the case of SHINE, this prognostic variable is also used to define the outcome. While the literature provides many resources on the design and implementation of responder analysis, as well as examples of trials which used responder analysis, there are no clear resources discussing whether or not statistical analyses should be adjusted for the prognostic variables used to define successful outcome.

This research aims to investigate the effect of covariate adjustment in the responder analysis framework, particularly when the covariate is involved in the definition of successful outcome. The cut-points for the SHINE trial are clinically, rather than statistically, defined and so it is conceivable that adjustment for baseline severity in the statistical analysis may account for additional variation and increase the power to detect a true treatment effect. A simulation study is conducted to assess the operating characteristics (power and type I error) of categorically-adjusted and unadjusted analyses under several possible treatment effect scenarios. In addition, treatment effect estimates and their standard errors are examined across the various scenarios. Since the primary outcome for the SHINE trial is binary, we expect to see an increase of standard error on the treatment effect estimates, consistent with the findings of Robinson and Jewell [23]. However, also consistent with Robinson and Jewell, we expect to see this increase in standard error to be balanced by a movement of the treatment effect estimate away from the null hypothesis.

By examining the effect of covariate adjustment in responder analysis, we aim to define the most appropriate statistical approach to identify true treatment effects. Our findings are not only applicable to the SHINE and other stroke trials which use the mRS for the primary outcome, but also provide insight into the appropriate use of categorical baseline prognostic variables in other trials which use an ordinal scale as a primary outcome measure.

Methods

Simulation studies were performed to examine the performance of logistic regression models that were unadjusted

and adjusted by a trichotomized baseline severity category. Baseline severity category and criteria for successful outcome were defined as in the SHINE trial described above, and are summarized in Table 1. The type I error rate and power were calculated and compared for each method, as were the treatment effect estimates and their standard errors.

The simulation parameters were guided by the SHINE trial design. A total of 1,000 clinical trials were simulated at sample sizes ranging from 498 to 1,958. This sample size range allowed us to cover the planned SHINE sample size of 1,400 while also examining model behavior at smaller and larger sample sizes. A 1:1 randomization scheme was assumed for the purposes of this investigation. All analyses were performed using SAS version 9.2 (SAS Institute, Cary, NC, USA).

The prevalence of each baseline severity category was guided by data from two prior pilot trials of hyperglycemia management in acute stroke, the Glucose Regulation in Acute Stroke Patients (GRASP) [24] and Treatment of Hyperglycemia in Ischemic Stroke (THIS) [25] pilot trials. In the simulations, 42% of subjects were classified as "mild" at baseline, 32% classified as "moderate", and the remaining 26% classified as "severe". This distribution of prognosis categories was imposed using a uniform (0, 1) random variable. In order to simulate 90-day mRS scores for the control group, we examined the distribution of 90-day mRS scores for the control groups in the GRASP and THIS pilot trials. Though the simulation of 90-day mRS scores was primarily driven by the results of the GRASP and THIS pilot trials, the National Institute of Neurological Disorders and Stroke tissue Plasminogen Activator (NINDS tPA) trial control data [26] were used to aid in the approximation of mRS outcome distributions within each of the baseline severity strata. The NINDS tPA control data helped smooth the distribution of mRS scores, as the GRASP and THIS pilot trials each had small sample sizes that resulted in several empty cells after baseline severity stratification. The exact control group distribution of 90-day mRS scores used in the simulation study is shown in Additional file 1: Table S1.

Type I error rates for each method of analysis were obtained by using the same proportion of success for both the control and intervention groups, simulating the null hypothesis of "no treatment effect". In order to assess the power of each method, a treatment effect was simulated in the data by altering the success prevalence for the intervention group. A 7% treatment effect was used, as this was the minimal clinically relevant absolute difference in favorable outcome between the two treatment groups in the SHINE study plan. For these analyses, power was examined under several scenarios as illustrated in Table 2: (1) a "flat" scenario, in which the 7% treatment effect was held constant over the three baseline severity strata; (2) a "varying" scenario, in which the overall treatment effect is still 7%, but the magnitude within strata is varied, where the mild and moderate groups see the most benefit; (3) another "varying" scenario, in which the severe group sees the most benefit; (4) a "mild harm" scenario, where the mild group sees a harmful treatment effect; and (5) a "severe harm" scenario, in which the severe group sees a harmful treatment effect.

In the first varying scenario, we applied an 8.6% treatment effect in the mild category, a 9% treatment effect in the moderate category and a 2% treatment effect in the severe category; that is, there was an 8.6% increase in prevalence of the 0 mRS for the mild stratum, a 9% increase in the prevalence of the 0 to 1 range of mRS scores for the moderate stratum, and a 2% increase in the prevalence of the 0 to 2 range of mRS scores for the severe stratum. This scenario is relevant to the SHINE trial; it is similar to what we may observe if the investigational treatment is largely beneficial to mild and moderate stroke victims, but only marginally beneficial to victims of severe stroke. The second varying treatment effect scenario applies an opposite effect in which the intensive glucose control intervention is largely beneficial to more severe strokes, but only slightly beneficial to those subjects having mild strokes. Additional file 1: Table S1 shows the exact distribution of 90-day mRS scores for the treatment groups under each of these treatment effect scenarios. These distributions were used to randomly assign 90-day mRS scores to each simulated subject in each simulated trial, with the proportions of success following the scenarios in Table 2. Given a subject's simulated baseline severity stratum (mild, moderate or severe), an assignment of "success" or "failure" was made according to the sliding dichotomy definitions.

Logistic regression was used to investigate each of these scenarios. The unadjusted case models "success" as a function only of treatment group, while the categorically-adjusted case models "success" as a function of treatment group and severity category. Severity was defined as "mild," "moderate" or "severe" based on the NIHSS prognosis group discussed in the introduction. Power and type I error rate were based on the proportion of simulated trials at a given sample size which rejected the null

Table 1 Sliding dichotomy criteria for successful outcome in SHINE trial

Baseline NIHSS	Prognosis group	90-day mRS for successful outcome
3 to 7	Mild	0
8 to 14	Moderate	0, 1
15 to 22	Severe	0, 1, 2

Table 2 Success prevalence for simulated treatment effect scenarios

Baseline severity	Treatment effect scenarios					
	No treatment effect	Flat	Varying 1	Varying 2	Mild harm	Severe harm
Mild	25%	32%	33.6%	27%	23%	33%
Moderate	35%	42%	44%	44%	50%	48%
Severe	15%	22%	17%	27.6%	26.7%	13%

hypothesis at a nominal level of 0.05. The treatment effect and its standard error were estimated for each trial.

Results

The type I error rate at each sample size for each analysis method is plotted in Figure 1. The nominal 5% reference line is shown, along with the upper and lower 95% confidence limits on this nominal level of significance. The confidence limits were calculated using the formula for binomial proportion 95% confidence intervals. The confidence limits remain the same at each sample size, as they are based on the number of trials at each sample size (1,000) rather than the sample size itself. The type I error rates for both the unadjusted and categorically-adjusted methods are within the 95% confidence limits for all the sample sizes, hovering close to the nominal 5% level.

The first investigation of power was under a "flat" treatment effect of 7% where the success rates in the control group were 25%, 35% and 15% in the mild, moderate and severe prognosis groups, respectively. The power estimates for this "flat" treatment effect scenario are plotted in Figure 2. The unadjusted and categorically-adjusted methods do not significantly differ, with the categorically-adjusted method having slightly greater power for most of the sample sizes. As planned by the SHINE study investigators, the 80% power threshold is crossed between 650 and 700 subjects per arm (1,300 to 1,400 subjects total).

The next two scenarios varied the treatment effects across the mild, moderate and severe baseline categories as 8.6%, 9% and 2%, respectively and 2%, 9% and 12.6%, respectively. The power results for these two scenarios are shown in Figure 3. As in the flat treatment effect scenario, there is no drastic difference in the unadjusted and categorically-adjusted methods with respect to power in these varying treatment effect scenarios.

As previously mentioned, it is conceivable that one of the prognosis groups may experience a slightly harmful treatment effect. When 2% harm is experienced in either the mild or the severe baseline prognosis category, the unadjusted and adjusted analyses still appear to have a similar performance, as shown in Figure 4. In the mild harm scenario, the unadjusted and adjusted power curves are still nearly stacked upon one another, with the power curve for the adjusted analysis pulling slightly above that of the unadjusted analysis at a few points. A more noticeable difference can be seen in the severe harm scenario,

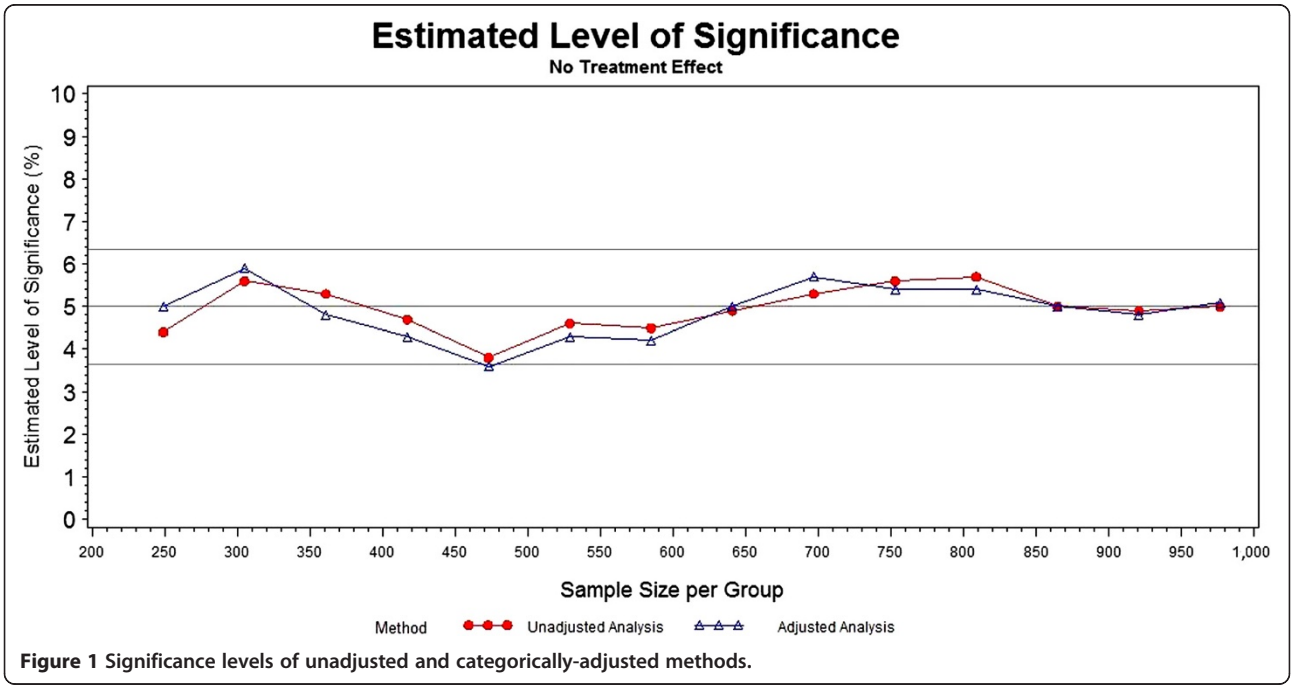

Figure 1 Significance levels of unadjusted and categorically-adjusted methods.

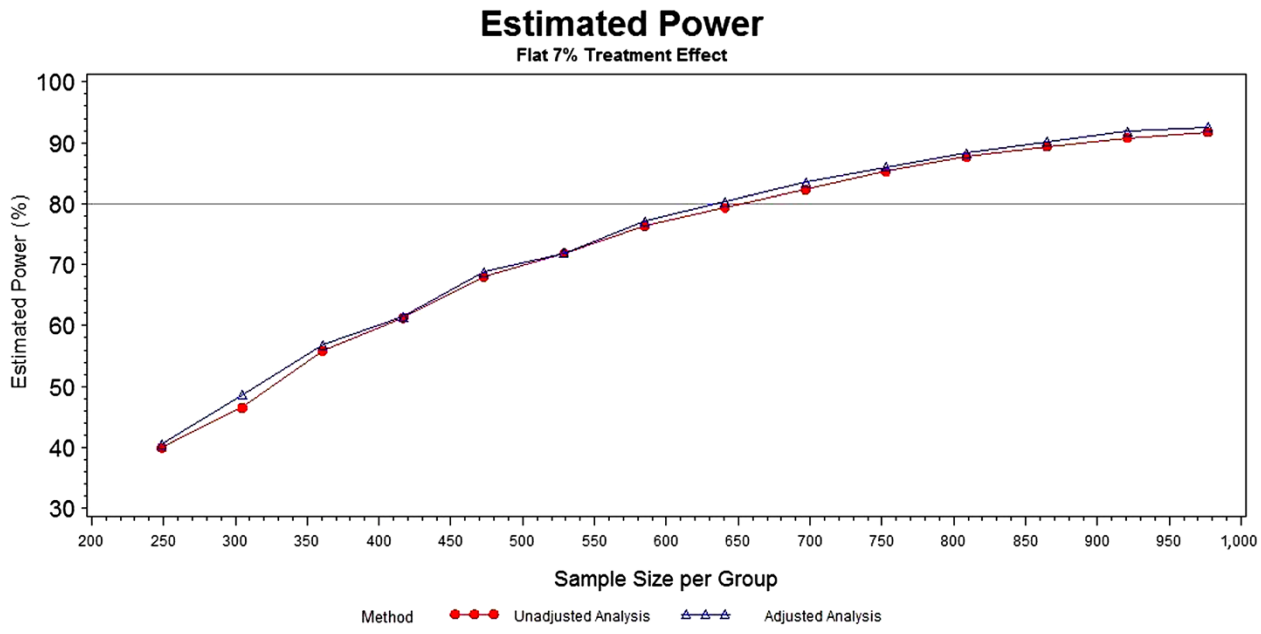

Figure 2 Power of unadjusted and categorically-adjusted methods under a flat 7% treatment effect.

where the adjusted analysis consistently has a slightly, though not dramatically, higher power than that of the unadjusted analysis.

In addition to the plots in Figures 2, 3 and 4, we also observed the treatment coefficient estimates and their standard errors for the adjusted and unadjusted models under the various treatment effect scenarios at selected sample sizes. The sample sizes of 498, 722, 946, 1,170 and 1,394 were chosen because they are the closest sample sizes to those at which the planned interim and final analyses will take place for SHINE. In addition to model estimates, the true treatment effect coefficient was calculated by pooling the nominal log-odds ratios for each prognosis group. To visualize the bias of the estimate of each treatment effect parameter and their standard errors, the simulation mean squared error (MSE) was plotted against the squared bias

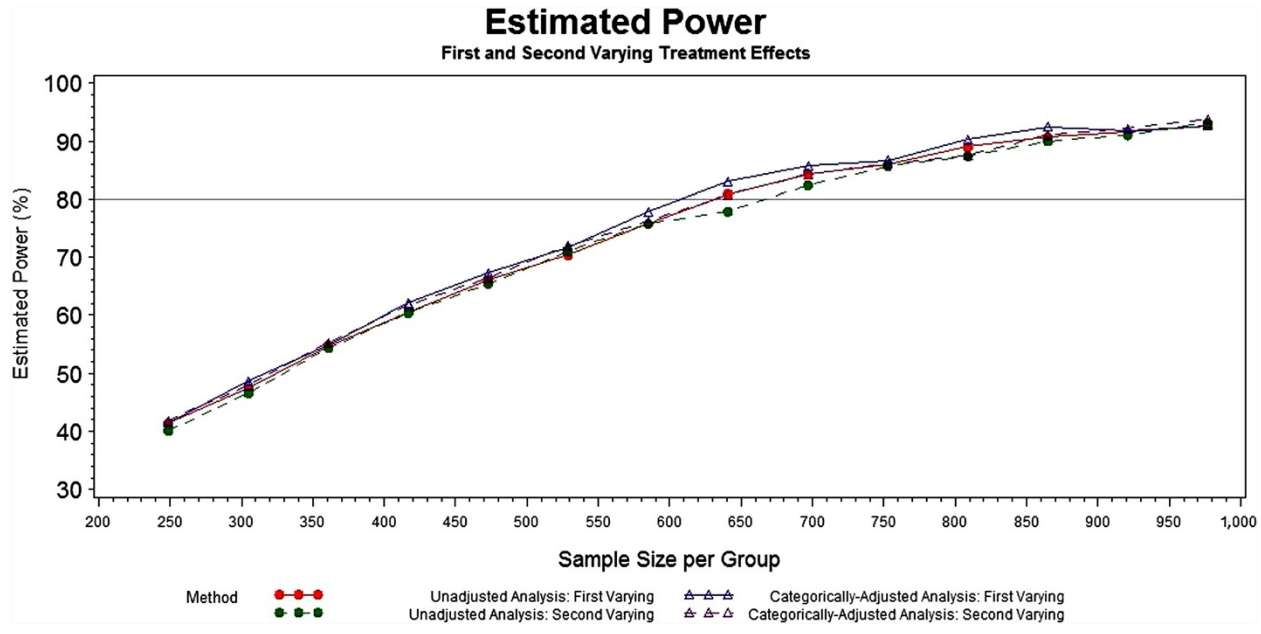

Figure 3 Power of unadjusted and categorically-adjusted methods under the first and second varying 7% treatment effects.

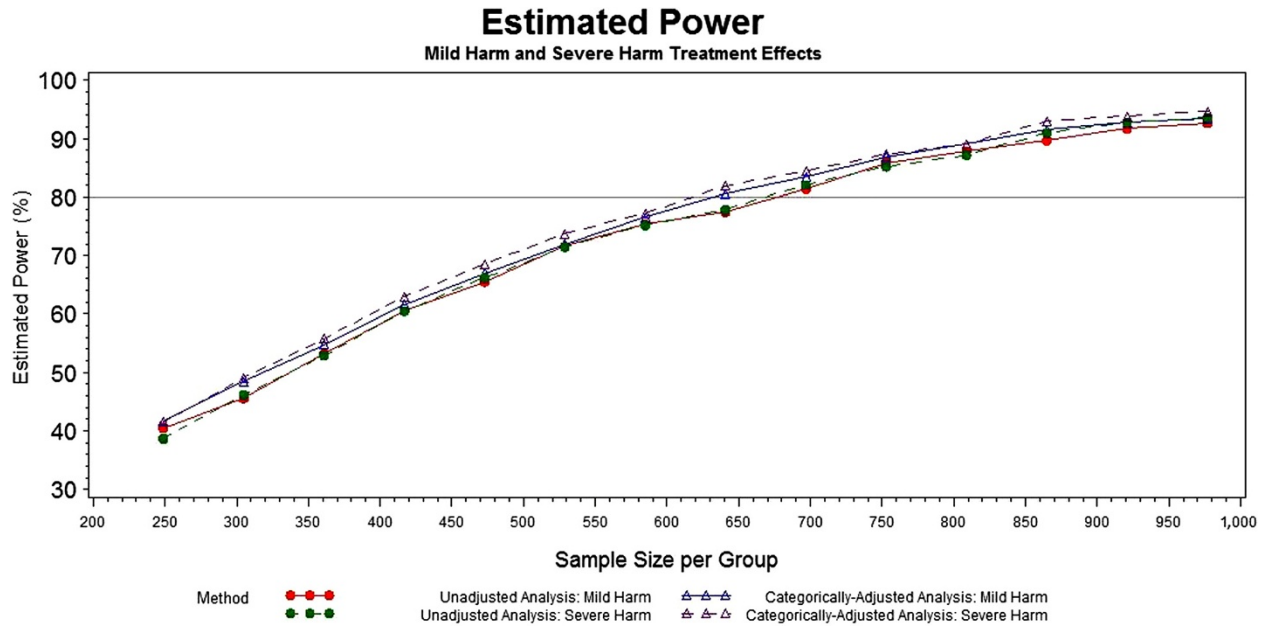

Figure 4 Power of unadjusted and categorically-adjusted methods under mild and severe harm effects.

in Figure 5. The MSE quantifies the accuracy and precision of an estimate in terms of both the bias (the difference between the true and estimated treatment effect) and the variance of the estimate. By plotting the MSE against the squared bias, we can illustrate the adequacy of the estimator. In Figure 5, the squared bias is depicted on the x-axis and the MSE on the y-axis. While the bias decreases with increasing sample size, the adjusted estimates of the treatment effect parameter are consistently less biased than the unadjusted estimates. For smaller sample sizes, the MSEs for the adjusted analyses are negligibly larger than those for the unadjusted analyses. The treatment coefficients and standard errors are provided in Additional file 2: Table S2.

Discussion

Successful stroke treatments are desperately needed given stroke's large and detrimental effect on the worldwide population. Consequently, statistical methods that offer high power to detect a true treatment effect are also needed. With this simulation study, we sought to determine whether adjustment for baseline severity within the responder analysis setting would be beneficial or harmful in terms of power and type I error rates when compared to an unadjusted analysis.

The type I error rates did not differ substantially between the two methods. The experimental type I error rates for both of the methods stayed within the 95% confidence bounds. This is a welcomed result, as a test that is either too liberal or too conservative, (rejects the null hypothesis either more or less than the nominal level, respectively), has implications on the power of the test. The oscillation around the nominal 5% level of significance is likely due to chance, and is to be expected in simulated data. Since neither method shows consistently larger type I error rates than the other, we can conclude that there is no meaningful difference between the two methods with respect to type I error.

The power appears to be approximately the same or slightly higher for the adjusted analyses in the selected scenarios. In the cases where the power is slightly higher, the magnitude is not remarkable and offers little evidence to suggest that adjusting by the single covariate leads to significantly more power. Although the simulation study presented is not exhaustive and, therefore, does not provide additional insight regarding this, the literature by Choi and Hernández suggest that an increase in power could occur as other important prognostic variables are added to the model [27,28]. It is reassuring, however, that neither method appears to be detrimental to power under the given scenarios.

In terms of bias, the unadjusted analyses consistently underestimate the nominal treatment effect, while the adjusted analyses tend to be less biased, but often slightly overestimate the nominal treatment effect. Given the magnitude of the coefficient estimates and their standard errors, neither of these bias tendencies is substantial. In terms of MSE, the two methods do not differ greatly as the sample size increases. At the smaller sample sizes, the adjusted analyses have larger MSE values due to increased standard error; however, as the sample size increases, the MSE values for the two methods converge.

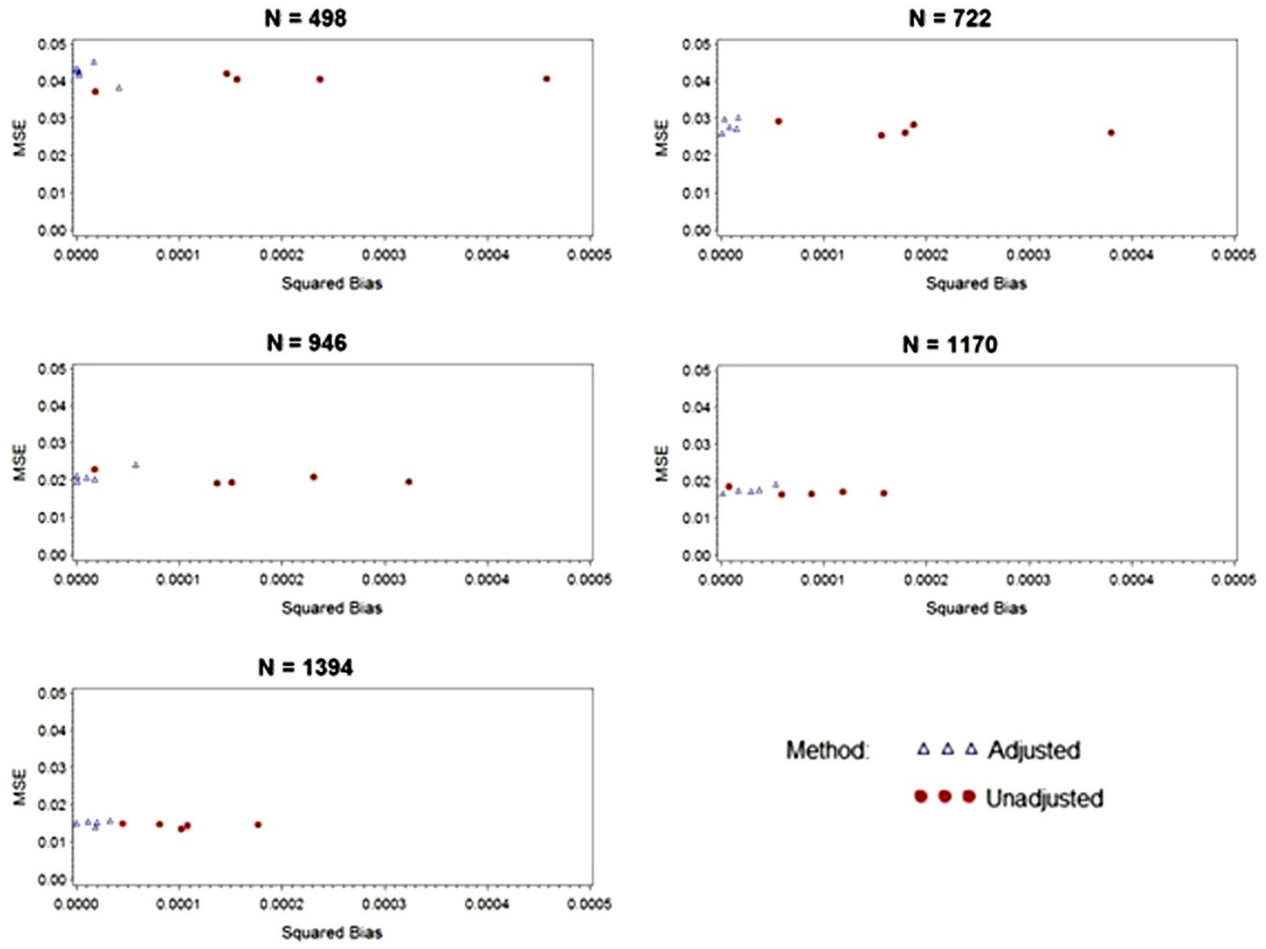

Figure 5 Mean squared error versus squared bias at selected sample sizes.

Though negligible differences were identified between the adjusted and unadjusted models, researchers should keep the randomization scheme of the study in mind when deciding whether or not to adjust for baseline severity. In general, it is advisable to "analyze as you randomize," meaning that any variable used as a stratification variable during randomization should be included as a covariate in the analysis in order to preserve nominal type I error rates and power [22,29]. Baseline severity is often used as a stratification variable in the randomization of acute stroke clinical trials, and should be included as a covariate in these cases.

It is important to note that these analyses adjust categorically for baseline severity. The categories - mild, moderate and severe - are defined by the NIHSS score, which is a larger scale ranging from 0 to 42 (limited to 3 through 22 in SHINE's inclusion criteria). A one-unit change in the NIHSS cannot easily be interpreted, as this change may have different meanings depending on the combination of neurological impairments and location along the scale. Despite this issue, the NIHSS is sometimes used as a continuous measure in the literature [30,31]. This is not necessarily straightforward and should be done with caution. It is possible that adjusting by the actual NIHSS score will provide additional information to the model and increase or maintain power in some treatment effect scenario(s). However, due to uncertainties in the clinical interpretation of a continuous NIHSS variable, adjustment by actual NIHSS score has been left as a topic for future research.

Adjustment for other baseline prognostic variables may also impact study power under the given scenarios. The inclusion of additional covariates that were not used in defining the primary outcome has not been examined in these scenarios, as it is outside the primary focus of this research. It is conceivable that the addition of multiple covariates could reduce overall power due to the increasing standard error on the treatment effect estimate, as studied by Robinson and Jewell [23] and discussed in the Background section of this paper.

Conclusions

Our results show negligible differences between analysis methods in the responder analysis setting, suggesting that in most treatment effect scenarios, adjustment for baseline severity in the primary analyses may best be guided by individual study needs rather than a blanket guideline for all studies. Though we have not shown the results here, we did examine other treatment effect scenarios which yield similar results. These scenarios included a flat and varying 15% treatment effect (instead of the 7% specified in the SHINE study plan), as well as a scenario in which the mild group experienced 5% harm.

Overall, these results shed light on the important concept of adjustment in the context of responder analysis. Though this study only examined a single severity scale, its findings are not restricted to use in stroke studies; they can provide insight into the treatment of categorical baseline prognostic covariates in other studies which use responder analysis to define their primary outcome of interest.

Additional files

Additional file 1: Table S1. Distribution of 90-day mRS scores.

Additional file 2: Table S2. Treatment coefficient estimates and their standard errors for unadjusted and adjusted methods under different treatment effect scenarios.

Abbreviations

GRASP: Glucose Regulation in Acute Stroke Patients; mRS: modified Rankin Scale; MSE: mean square error; NIHSS: National Institute of Health Stroke Scale; NINDS tPA: National Institute of Neurological Disorders and Stroke tissue Plasminogen Activator; SHINE trial: Stroke Hyperglycemia Insulin Network Effort trial; THIS: Treatment of Hyperglycemia in Ischemic Stroke.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KG carried out the simulation programming and interpretation, manuscript drafting and finalization. SY helped conceive the study concept, assisted in simulation programming and interpretation, and aided in manuscript drafting and finalization. VR helped conceive the study concept, aided in statistical interpretation, as well as manuscript drafting and finalization. KJ and EJ assisted with study concept, design, clinical interpretation, manuscript drafting and finalization. VD helped conceive the study concept, aided in design, analysis and interpretation, as well as manuscript drafting and finalization. VD provided overall supervision as the primary mentor of the first author. All authors read and approved the final manuscript.

Authors' information

KG recently completed her master's degree in biostatistics from the Medical University of South Carolina. This manuscript is a result of her master's thesis and her work as a graduate student on the SHINE grant. The other authors were her committee members and VD was her primary mentor.

Acknowledgements

This work was funded by both the SHINE trial NIH/NINDS grant 1U01-NS069498 and the Neurological Emergencies Treatment Trial (NETT) Statistical and Data Management Center NIH/NINDS grant U01-NS059041.

Author details

[1]Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425, USA. [2]Division of Emergency Medicine, Department of Medicine, Medical University of South Carolina, 169 Ashley Avenue, Charleston, SC 29425, USA. [3]Department of Neurology, University of Virginia School of Medicine, 81 Hospital Drive, McKim Hall Room 2026, Charlottesville, VA 22908, USA.

Received: 18 November 2012 Accepted: 13 March 2013
Published: 11 April 2013

References

1. Sacco RL, Frieden TR, Blakeman DE, Jauch EC, Mohl S: **What the Million Hearts Initiative means for stroke: a presidential advisory from the American Heart Association/American Stroke Association.** *Stroke* 2012, **43:**924–928.
2. Saver JL: **Optimal endpoints for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices.** *Stroke* 2011, **42:**2356–2362.
3. Hong KS, Lee SJ, Hao Q, Liebeskind DS, Saver JL: **Acute stroke trials in the 1st decade of the 21st century.** *Stroke* 2011, **42:**e314.
4. The Optimising Analysis of Stroke Trials (OAST) Collaboration, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J: **Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials.** *Stroke* 2007, **38:**1911–1915.
5. Banks JL, Marotta CA: **Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials.** *Stroke* 2007, **38:**1091–1096.
6. 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.
7. Young FB, Lees KR, Weir CJ: **Strengthening acute stroke trials through optimal use of disability end points.** *Stroke* 2003, **34:**2676–2680.
8. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, Marmarou A, Mendelow AD, Steyerberg EW, Taylor GS, Teasdale GM, Weir CJ: **Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy.** *J Neurotrauma* 2005, **22:**511–517.
9. 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.
10. Kasner SE: **Clinical interpretation and use of stroke scales.** *Lancet Neurol* 2006, **5:**603–612.
11. Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, Lyden P, Grotta J: **Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial.** *Stroke* 1996, **27:**2136–2142.
12. Cobo E, Secades JJ, Miras F, Gonzalez JA, Saver JL, Corchero C, Rius R, Dàvalos A: **Boosting the chances to improve stroke treatment.** *Stroke* 2010, **41:**e143–e150.
13. Saver JL, Gornbein J: **Treatment effects for which shift or binary analyses are advantageous in acute stroke trials.** *Neurology* 2009, **72:**1310–1315.
14. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD: **A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project.** *Clin Trials* 2010, **7:**44–57.
15. Howard G, Waller JL, Voeks JH, Howard VJ, Jauch EC, Lees KR, Nichols FT, Rahlfs VW, Hess DC: **A simple, assumption-free, and clinically interpretable approach for analysis of modified Rankin outcomes.** *Stroke* 2012, **43:**664–669.
16. Saver JL, Yafeh B: **Confirmation of tPA treatment effect by baseline severity-adjusted end point reanalysis of the NINDS-tPA stroke trials.** *Stroke* 2007, **38:**414–416.
17. Young FB, Lees KR, Weir CJ: **Improving trial power through use of prognosis-adjusted end points.** *Stroke* 2005, **36:**597–601.
18. Bruno A, Durkalski VL, Hall CE, Juneja R, Barsan WG, Janis S, Meurer WJ, Fansler A, Johnston KC: **The stroke hyperglycemia insulin network effort (SHINE) trial; design and methodology.** *Int J Stroke*. In press.
19. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, Koudstaal PJ, Dippel DW, PAIS Investigators: **The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial.** *Lancet Neurol* 2009, **8:**434–440.

20. Adams HP, Effron MB, Torner J, Davalos A, Frayne J, Teal P, Leclerc J, Oemar B, Padgett L, Barnathan ES, Hacke W: **Emergency administration of abciximab for treatment of patients with acute ischemic stroke: results of an international phase III trial.** *Stroke* 2008, **39:**87–99.
21. Piantadosi S: *Clinical Trials: A Methodological Perspective.* 2nd edition. Hoboken, New Jersey: John Wiley and Sons; 2005:470–473.
22. Harrell FE: *The Role of Covariable Adjustment in the Analysis of Clinical Trials.* [http://biostat.mc.vanderbilt.edu/twiki/pub/Main/FHHandouts/covadj.pdf].
23. Robinson LD, Jewell NP: **Some surprising results about covariate adjustment in logistic regression models.** *Int Stat Rev* 1991, **58:**227–240.
24. Johnston KC, Hall CE, Kissela BM, Bleck TP, Conaway MR, GRASP Investigators: **Glucose Regulation in Acute Stroke Patients (GRASP) trial: a randomized pilot trial.** *Stroke* 2009, **40:**3804–3809.
25. Bruno A, Kent TA, Coull BM, Shankar RR, Saha C, Becker KJ, Kissela BM, Williams LS: **Treatment of Hyperglycemia in Ischemic Stroke (THIS): a randomized pilot trial.** *Stroke* 2008, **39:**384–389.
26. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group: **Tissue plasminogen activator for acute ischemic stroke.** *N Eng J Med* 1995, **333:**1581–1588.
27. Choi SC: **Sample size in clinical trials with dichotomous endpoints: use of covariables.** *J Biopharm Stat* 1998, **8:**367–375.
28. Hernández AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema DF, Maas AI: **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma* 2006, **23:**1295–1303.
29. Committee for Proprietary Medicinal Products: Committee for Proprietary Medicinal Products (CPMP): **points to consider on adjustment for baseline covariates.** *Stat Med* 2004, **23:**701–709.
30. Lin HJ, Chang WL, Tseng MC: **Readmission after stroke in a hospital based registry: risk, etiologies, and risk factors.** *Neurology* 2011, **76:**438–443.
31. Dai DF, Thajeb P, Tu CF, Chiang FT, Chen CH, Yang RB, Chen JJ: **Plasma concentration of SCUBE1, a novel platelet protein, is elevated in patients with acute coronary syndrome and ischemic stroke.** *J Am Coll Cardiol* 2008, **51:**2173–2180.

doi:10.1186/1745-6215-14-98
Cite this article as: Garofolo *et al.*: The effect of covariate adjustment for baseline severity in acute stroke clinical trials with responder analysis outcomes. *Trials* 2013 **14**:98.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central