Trials

METHODOLOGY

Open Access

# Improving analysis of cognitive outcomes in cardiovascular trials using different statistical approaches

Shun Fu Lee[1,2*], William Whiteley[3,4], Jackie Bosch[1,5], Laura Sherlock[3], Tali Cukierman-Yaffe[6,7], Martin O'Donnell[8,9], John W. Eikelboom[1,10], Hertzel C. Gerstein[1,10], Shrikant I. Bangdiwala[1,2] and Graciela Muniz-Terrera[3,11]

## Abstract

**Background** The Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) questionnaires are commonly used to measure global cognition in clinical trials. Because these scales are discrete and bounded with ceiling and floor effects and highly skewed, their analysis as continuous outcomes presents challenges. Normality assumptions of linear regression models are usually violated, which may result in failure to detect associations with variables of interest.

**Methods** Alternative approaches to analyzing the results of these cognitive batteries include transformations (standardization, square root, or log transformation) of the scores in the multivariate linear regression (MLR) model, the use of nonlinear beta-binomial regression (which is not dependent on the assumption of normality), or Tobit regression, which adds a latent variable to account for bounded data. We aim to empirically compare the model performance of all proposed approaches using four large randomized controlled trials (ORIGIN, TRANSCEND, COMPASS, and NAVIGATE-ESUS), and using as metrics the Akaike information criterion (AIC). We also compared the treatment effects for the methods that have the same unit of measure (i.e., untransformed MLR, beta-binomial, and Tobit).

**Results** The beta-binomial consistently demonstrated superior model performance, with the lowest AIC values among nearly all the approaches considered, followed by the MLR with square root and log transformations across all four studies. Notably, in ORIGIN, a substantial AIC reduction was observed when comparing the untransformed MLR to the beta-binomial, whereas other studies had relatively small AIC reductions. The beta-binomial model also resulted in a significant treatment effect in ORIGIN, while the untransformed MLR and Tobit regression showed no significance. The other three studies had similar and insignificant treatment effects among the three approaches.

**Conclusion** When analyzing discrete and bounded outcomes, such as cognitive scores, as continuous variables, a beta-binomial regression model improves model performance, avoids spurious significance, and allows for a direct interpretation of the actual cognitive measure.

**Trials registration** ORIGIN (NCT00069784). Registered on October 1, 2003; TRANSCEND (NCT00153101). Registered on September 9, 2005; COMPASS (NCT01776424). Registered on January 24, 2013; NAVIGATE-ESUS (NCT02313909). Registered on December 8, 2014.

**Keywords** Bounded, Ceiling effect, Cognitive, Generalized linear regression, Beta-binomial, Tobit, Transformations

*Correspondence:
Shun Fu Lee
shunfu.lee@phri.ca
Full list of author information is available at the end of the article

## Background

Over the last two decades, there have been extensive efforts to test interventions to slow cognitive decline and prevent dementia. Results have been at best inconsistent, and researchers have questioned whether limitations of commonly used methodological approaches might be obscuring real effects. For instance, commonly used standard cognitive tests such as the Mini-Mental State Examination (MMSE) [1, 2] and the Montreal Cognitive Assessment (MoCA) [3] have strong ceiling and floor effects (i.e., highly skewed). The floor effect is often a result of the cognitive measures' poor sensitivity in detecting severe cognitive impairment [4], leading to a clustering of scores at the low end. This precludes their analysis as continuous variables with the consequence that true differences between randomly allocated groups in a trial may not be detected. Indeed, the MMSE and MoCA are often analyzed with generalized linear regression models such as analysis of covariance (ANCOVA) that assumes normality of outcomes for pre- and post-scores or change scores that do not fully account for their psychometric properties. As it is rare for the results of cognitive tests to be normally distributed, the use of linear regression models (which assume normality) may be inappropriate and hamper the ability to detect a true effect.

When distributional assumptions are violated, alternative approaches for the analyses of data include the transformation of the cognitive scores to create a normal distribution (after which linear regression can be used), the use of different regression models that are not dependent on normality, or the use of other assessments where the scores are normally distributed. For example, in international studies, researchers normalize a skewed distribution using country-standardized z-scores [5], or transformed scores such as $\sqrt{\text{Max}+1}-\sqrt{(\text{Max}+1)-\text{score}}$ and $\log(\text{Max}+1)-\log((\text{Max}+1)-\text{score})$ [6]. Regressions that are not dependent on normality (due to bounded data) are beta-binomial regression [7–9] and Tobit regression [10–12]. In the beta-binomial regression, the outcome measure derived from the sum of all discrete questions is modeled as a discrete binomial and a beta distribution to account for overdispersed data. Tobit regression treats the outcome responses as if they were normally distributed but censors them if they are outside a given range. Although these three approaches have been used in observational studies [8, 9], they are rarely used in clinical trials.

This study aims to empirically evaluate model performance of different approaches analyzing continuous cognitive scores versus linear models of the untransformed data using four large randomized controlled trials. We also compared treatment effects across methods that use the same unit of measurement (i.e., untransformed MLR, beta-binomial, and Tobit). The lessons learnt from these post hoc analyses may inform analyses of data from other cognitive measurements characterized by discrete and bounded outcomes in clinical trials with non-normally distributed scores.

This paper is organized as follows. Data section presents a description of the four trials. Methods section starts by introducing generalized linear regressions to transformed outcomes and non-linear regressions and follows by describing the analysis used to compare the different approaches against the standard or untransformed generalized linear regression. The analysis includes assessing the distribution of the scores, estimating the effect size (except for the ones with the transformation), computing the confidence intervals, and evaluating model fit measured by the Akaike information criterion (AIC) [13]. Finally, in the Results section, we provide a brief discussion of the results obtained and conclude with some general conclusions and recommendations in the "Discussion" section.

## Methods

### Overview of the four clinical trials

We used data from 4 international clinical trials. The Outcome Reduction with an Initial Glargine INtervention (ORIGIN) trial investigated the effect of insulin glargine (versus standard care) on cardiovascular events in patients with diabetes and pre-diabetes [14]. The MMSE was administered at baseline, year 2, year 5, and the last follow-up. The effect of the interventions on the MMSE was assessed with repeated-measures ANOVA before and after adjusting for selected baseline covariates [15]. The Telmisartan Randomised AssessmeNt Study in ACE iNtolerant subjects with cardiovascular Disease (TRANSCEND) trial assessed the effect of the angiotensin-receptor blocker telmisartan (versus placebo) in patients intolerant to ACE inhibitors with cardiovascular disease or diabetes with end-organ damage [16]. The MMSE was administered at baseline and last follow-up and then dichotomized and analyzed using logistic regression to assess the treatment effect [17]. The Cardiovascular Outcomes for People Using Anticoagulation Strategies (COMPASS) trial evaluated whether rivaroxaban alone or in combination with aspirin would be more effective than aspirin alone for secondary cardiovascular prevention [18]. The New Approach Rivaroxaban Inhibition of Factor Xa in a Global Trial versus ASA to Prevent Embolism in Embolic Stroke of Undetermined Source (NAVIGATE ESUS) trial tested whether rivaroxaban (versus aspirin) would reduce in result in a lower risk of recurrent stroke than aspirin [19, 20]. Both COMPASS and

NAVIGATE-ESUS only collected the MoCA scores at baseline and study end. COMPASS has not yet published cognitive measures by the treatment group. However, NAVIGATE-ESUS assessed changes in MoCA scores between treatment groups using the Mann–Whitney U or Kruskal–Wallis test [20].

## Overview of the analysis methods

### *Transformation to enable the use of multivariable linear regression (MLR) Models*

The generalized linear regression model [21] assumes normality of outcome and is summarized below.

Let $y_i$ be the continuous outcome for individual $i$, $i = 1, \ldots, n$. The continuous outcome $y_i$ is expressed as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 \text{Baseline}_i + \beta_3 \text{Age}_i + \beta_4 \text{Sex}_i + \beta_5 \text{Edu}_i + e_i \quad (1)$$

- $\beta_0$ is the intercept;
- $\beta_1$ is the fixed treatment effect for the experimental intervention relative to the control;
- $x_i$ is the binary treatment indicator with 1 for intervention and 0 for control for the $i$th individual.
- $\beta_k$ for $k = 2, 3, 4,$ and 5 are the fixed effects for the covariates: baseline score, age, sex, and education level, respectively.
- $e_i \sim N(0, \sigma_i^2)$ is the error term for individual $i$.

One transformation, such as standardizing cognitive measures, was initially introduced to address methodological issues where the cognitive performance of individuals without impairment, particularly those with higher education, was challenging. This approach was further extended to standardize scores by country, using country-specific baseline parameters, with the baseline mean and standard deviation calculated separately for every country in a multinational trial. This method has the added benefit of accounting for systematic country differences. Other approaches are based on the fact that given that the distributions of the MMSE or MoCA scores are left skewed and that the highest score of each is 30. In these approaches, transformations [expressed as $h(y_i)$] were done to satisfy the normality assumption of linear regression models [22] as follows: $h(\text{score}) = \sqrt{\text{Max} + 1} - \sqrt{(\text{Max} + 1) - \text{score}}$ and $h(\text{score}) = \log(\text{Max} + 1) - \log((\text{Max} + 1) - \text{score})$.

### *Regression models not dependent on normality: beta-binomial regression*

Given the cognitive scores often are calculated by summing equal 0 or 1 scores from a series of questions (e.g., the MMSE was calculated by summing up 30 yes–no questions), the beta-binomial distribution has a similar data structure that consists of a finite sum of Bernoulli variables with a probability parameter analyzed as a random variable that follows a beta distribution.

Let $y_i$ be a sum of all binary responses of $n$ questions for individual $i$ conditioned on the random variable with a probability parameter $p_i$ of a Bernoulli distribution. Then $p_i$ follows a beta distribution with parameters $(a, b)$ and,

$$E(p_i) = \mu = \frac{a}{a+b}, V(p_i) = \mu(1-\mu)\rho$$

where $\rho = (1 + a + b)^{-1}$ is the intraclass correlation coefficient [23]. Then $y_i$ follows a beta-binomial distribution with

$$E(y_i) = n\mu, V(y_i) = n\mu(1-\mu)[1 + (n-1)\rho].$$

The linear predictor of the beta-binomial hierarchical generalized linear model [7] is

$$\text{logit}\,(p_i) = \beta_0 + \beta_1 x_i + \beta_2 \text{Baseline}_i + \beta_3 \text{Age}_i + \beta_4 \text{Sex}_i + \beta_5 \text{Edu}_i + v_i$$

where $v_i$ is the random intercept effect attributed to individual $i$ and follows a beta distribution. For the covariate $x_i$, $E(p_i) = \exp(\beta_0 + \beta_1 x_i)/(1 + \exp(\beta_0 + \beta_1 x_i))$ [24]. A beta-binomial regression has the random component following a beta distribution with more flexibility on the shape of distribution instead of the common normal distribution with a symmetrical bell shape to model binomial overdispersed. For example, the beta-binomial takes on a U-shaped if both a and b are less than 1 and approximates the binomial distribution if both a and b are greater than 1 [25].

### *Regression models not dependent on normality: Tobit regression*

Tobit regression is an alternative under normality assumptions and in the presence of moderate ceiling or floor effect, by setting a continuous variable as the response constrained or censored to a closed interval [10, 11].

Let $y_i^*$ be the random latent variable that is not censored for individual $i$. Furthermore, it is assumed that $y_i^*$ can be observed for a given range $[l, u]$ and is censored when $y_i^*$ falls outside of that range. The Tobit model with $y_i^*$ conditional on the base-specific parameters $b_i$ with the linear regression model as the underlying model for $y_i^*$ is given by

$$y_i^* | b_i = \beta_0 + \beta_1 x_i + \beta_2 \text{Baseline}_i + \beta_3 \text{Age}_i + \beta_4 \text{Sex}_i + \beta_5 \text{Edu}_i + z_i b_i + e_i$$

with $b_i \sim N(0,\ \text{D})$. Then $y_i$ is obtained from $y_i^*$ as

$$y_i = l \text{ for } y_i^* \le l$$

$$y_i = y_i^* \text{ for } l < y_i^* \leq u$$

$$y_i = u \text{ for } y_i^* \geq u$$

Although the approach is justified, the estimated coefficients obtained with this approach could go beyond the closed interval due to the artificial situation based on an assumption of censoring and may result in unrealistic interpretations.

### Statistical methods

Since the study duration and the frequency of data collection varied across studies, the model used for different analysis approaches had the last follow-up score as the dependent variable and the treatment effect as the independent variable adjusting for the baseline score, age, and sex.

The distributions of the last follow-up MMSE, as well as the country-standardized MMSE score, the square root MMSE, and the log-transformed MMSE, were assessed by plotting histograms and computed skewness and kurtosis. Skewness is a measure of the asymmetry and kurtosis is a measure of "peakedness" of a distribution. A distribution with either an absolute skew value larger than two or an absolute kurtosis larger than seven could be considered as substantial non-normal.

The treatment effect (treatment versus control) in clinical trials on the cognitive function was then analyzed using the untransformed MLR, MLR with country-standardized score, MLR with square root transformation, MLR with log transformation, beta-binomial regression, and Tobit regression. The covariates for the model included baseline score, age, sex, and education (< 12 years of education, ≥ 12 years of education) for both MMSE and MoCA, even though the MoCA score already accounts for the education level by adding one point for those with < 12 years of education as part of its score.

The model performance for the different analysis approaches of the same dataset was compared using the Akaike information criterion (AIC) [13], a measure of the log-likelihood penalized for the number of variables. To ensure comparability of AIC values across methods, the AIC values for the MLRs were obtained using the maximum likelihood estimation. Smaller AIC values indicate a better model fit. The AICs for the models with the transformed outcomes need to have their likelihood multiplied by the corresponding Jacobian matrix to be comparable with other AICs. For the MLR, the normality of residuals was further assessed by plotting the histograms and Q-Q plots. The treatment effects were only compared among untransformed MLR, beta-binomial regression, and Tobit regression given that the units of cognitive measure of transformed MLR became different from the original measure. The treatment effects were reported as mean differences with corresponding 95% confidence intervals (CI) and CI width. The assumption of linearity was assessed using the residual vs predicted plot. In addition, we conducted mixed models with repeated measures (MMRM) and beta-binomial regression for ORIGIN, which is the only study that collected MMSE data multiple times. Analyses were performed using SAS software, Version 9.4 and R Version 4.2.3.

## Results

Participants with baseline and last follow-up MMSE scores were included from ORIGIN ($n = 11{,}691$ mean age, 63.5 years, 65% male, 38% higher education, mean follow-up, 5.9 years) and TRANSCEND ($n = 5815$ mean age, 66.9 years, 57% male, 34% higher education, mean follow-up, 4.6 years) (Table 1). Participants with baseline and last follow-up MoCA scores were included from COMPASS ($n = 17{,}864$, mean age 68.2 years, 78% male, 47% higher education, mean follow-up, 1.9 years) and NAGIVATE-ESUS ($n = 7016$, mean age, 66.9 years, 62%

**Table 1** Baseline characteristics for ORIGIN, TRANSCEND, COMPASS, and NAVIGATE-ESUS

| Description | ORIGIN | TRANSCEND | COMPASS | NAVIGATE-ESUS |
|---|---|---|---|---|
| *N* | 11,691 | 5815 | 17,864 | 7016 |
| Study follow-up (years), mean (SD) | 5.9 (1.4) | 4.6 (1.0) | 1.9 (0.7) | 1.0 (0.6) |
| Age(years), mean (SD) | 63.5 (7.8) | 66.9 (7.3) | 68.2 (7.9) | 66.9 (9.8) |
| Male, n(%) | 7625 (65) | 3319 (57) | 13,996 (78) | 4327 (62) |
| Education > 12 years, *n* (%) | 4422 (38) | 1974 (34) | 8360 (47) | 3244 (46) |
| Baseline MMSE, mean (SD) | 27.9 (2.9) | 27.5 (3.1) | –- | –- |
| Baseline MMSE, median (Q1–Q3) | 29 (27, 30) | 29 (26, 30) | –- | –- |
| Baseline MoCA, mean (SD) | –- | –- | 24.5 (3.9) | 23.1 (5.5) |
| Baseline MoCA, median (Q1–Q3) | –- | –- | 25 (23, 27) | 24 (21, 27) |

*SD* standard deviation, *Q1* first quartile, *Q3* third quartile, *MMSE* Mini-Mental Status Examination, *MoCA* Montreal Cognitive Examination

male, 46% higher education, mean follow-up, 1 year). The mean baseline and follow-up MMSE scores for ORIGIN and TRANSCEND were 27.9 (SD = 2.9) vs. 27.3 (SD = 3.5) and 27.5 (SD = 3.1) vs. 27.4 (SD = 3.5), respectively, while for COMPASS and NAVIGATE-ESUS, the mean baseline and follow-up MoCA scores were 24.5 (SD = 3.9) vs. 24.4 (SD = 4.0) and 23.1 (SD = 5.5) vs. 23.5 (SD = 5.6), respectively (Tables 1 and 2). There was no significant difference in scores between baseline and follow-up in any of the four studies.

The follow-up MMSE distributions in the ORIGIN and TRANSCEND studies were similarly bounded, with most participants having a maximum score of 30 (Fig. 1). The distribution of change scores from baseline, since the models were adjusted for baseline, exhibited a lack of normality as well (Supplemental Fig. 1). However, the absolute skewness greater than 2 and the absolute kurtosis greater than 8 (Table 2) indicated that the distributions were substantially non-normal. Even after country-standardization of MMSE scores, the distributions still showed large absolute skewness and kurtosis values. Applying a square root or log transformation substantially reduced the skewness and kurtosis values. In contrast, the follow-up MoCA distributions in the COMPASS and NAVIGATE-ESUS studies were left-skewed but still considered normal, with most participants having scores close to the maximum (Fig. 1). The absolute skewness and kurtosis were high but still within the range expected for a normal distribution. Applying a square root or log transformation further improved the skewness of the distribution (Table 2).

In terms of fitting the MLR with transformations, we found that the model performance significantly improved for the MMSE scores, and slightly improved for the MoCA scores when we applied either the square root or log transformation as compared to the untransformed MLR (Table 2). For instance, in ORIGIN with MMSE scores, the AIC decreased from 58,013 for the untransformed MLR to 50,317 with the square root transformation, and to 47,162 with the log transformation. In COMPASS with MoCA scores, the AIC decreased from 84,925 for the untransformed MLR to 82,322 with the square root transformation but increased to 85,861 with the log transformation. The histograms of residuals (Supplemental Figs. 2 and 3) became more normally distributed by reducing the initially high skewness and kurtosis values to a more acceptable range after applying square

**Table 2** Last follow-up cognitive measures applying transformations and model performance after fitting a mixed linear model

| | Follow-up cognitive measures | | | | MLR |
|---|---|---|---|---|---|
| | Mean (SD) | Median (Q1-Q3) | Skewness | Kurtosis | AIC |
| ORIGIN (*N* = 11,691) | | | | | |
| MMSE | 27.3 (3.5) | 29 (26, 30) | −2.2 | 9.8 | 58,013 |
| STD MMSE | −0.2 (1.2) | 0.2 (−0.7, 0.7) | −1.7 | 6.2 | 62,601 |
| SQRT MMSE | 3.8 (0.8) | 4.2 (3.3, 4.6) | −1.1 | 4 | 50,317 |
| LOG MMSE | 0.9 (0.8) | 0.7 (0, 1.6) | 0.4 | 2.1 | **47,162** |
| TRANSCEND (*N* = 5815) | | | | | |
| MMSE | 27.4 (3.5) | 29 (26, 30) | −2.1 | 8.9 | 28,048 |
| STD MMSE | −0.1 (1.1) | 0.3 (−0.5, 0.7) | −1.9 | 7.0 | 30,628 |
| SQRT MMSE | 3.8 (0.8) | 4.2 (3.3, 4.6) | −1.1 | 3.9 | 24,214 |
| LOG MMSE | 0.93 (0.8) | 0.7 (0, 1.6) | 0.4 | 2.1 | **22,697** |
| COMPASS (*N* = 17,864) | | | | | |
| MoCA | 24.4 (4.0) | 25 (22, 27) | −1.3 | 5.2 | 84,925 |
| STD MoCA | 0 (1.0) | 0.1 (−0.6, 0.7) | −1.1 | 4.9 | 90,561 |
| SQRT MoCA | 3.1 (0.8) | 3.1 (2.6, 3.6) | −0.4 | 3.1 | **82,322** |
| LOG MMSE | 1.7 (0.7) | 1.8 (1.4, 2.2) | −0.5 | 3.3 | 85,861 |
| NAVIGATE-ESUS (*N* = 7016) | | | | | |
| MoCA | 23.5 (5.6) | 25 (21, 27) | −1.6 | 5.8 | 36,515 |
| STD MoCA | 0.1 (1.0) | 0.3 (−0.4, 0.8) | −1.4 | 5.8 | 38,780 |
| SQRT MoCA | 3 (1.0) | 3.1 (2.4, 3.6) | −0.6 | 3.2 | **34,179** |
| LOG MMSE | 1.8 (0.8) | 1.8 (1.4, 2.3) | −0.4 | 2.8 | 35,472 |

*SD* standard deviation, *Q1* first quartile, *Q3* third quartile, *Skewness* a measure of the asymmetry of a distribution (substantial non-normal: absolute skewness ≥ 2), *Kurtosis* a measure of "peakedness" of a distribution (substantial non-normal: absolute kurtosis ≥ 7), *MLR* Multivariate linear regression model for treatment effect adjusting for baseline score, age, sex, and education, *AIC* Akaike information criterion (smaller is better), *MMSE* Mini-Mental Status Examination, *MoCA* Montreal Cognitive Examination, *STD* country-standardized score

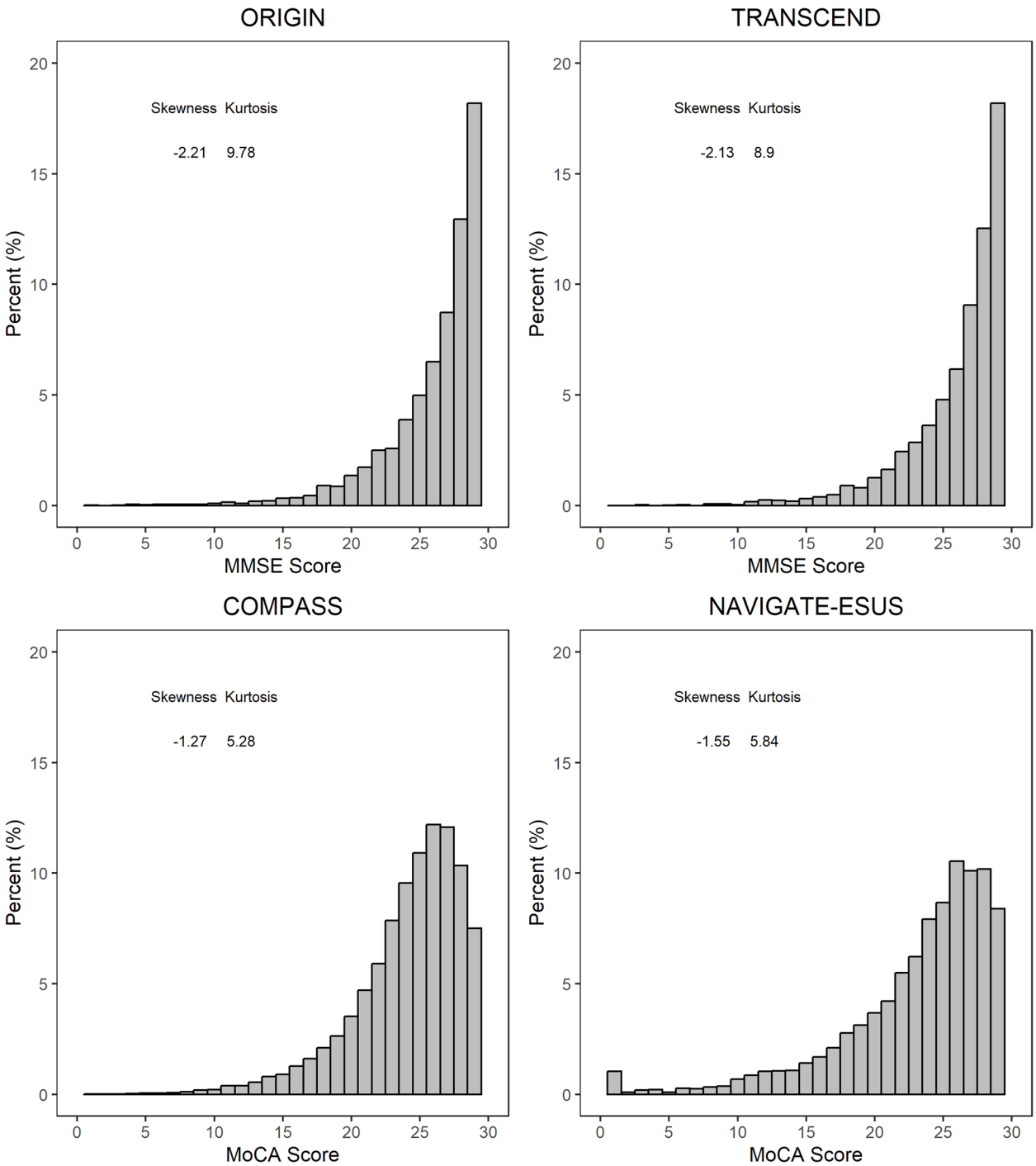



**Fig. 1** Histograms for the MMSE of ORIGIN and TRANSCEND studies and the MoCA of COMPASS and NAVIGATE-ESUS studies

root or log transformations. This was more evident in the Q-Q plots, where most points aligned closely to the straight line for the cognitive measures with a log transformation (Supplemental Figs. 4 and 5).

SQRT: $\sqrt{\text{Max}+1} - \sqrt{(\text{Max}+1) - \text{score}}$

LOG: $\log(\text{Max}+1) - \log((\text{Max}+1) - \text{score})$

AIC for Beta-binomial were 46,342, 46,342, 22,387, 81,848, and 34,288 for ORIGIN, TRANSCEND, COMPASS, and NAVIGATE-ESUS, respectively.

AIC for Tobit were 49,095, 23,588, 84,112, and 35,641 for ORIGIN, TRANSCEND, COMPASS, and NAVIGATE-ESUS, respectively.

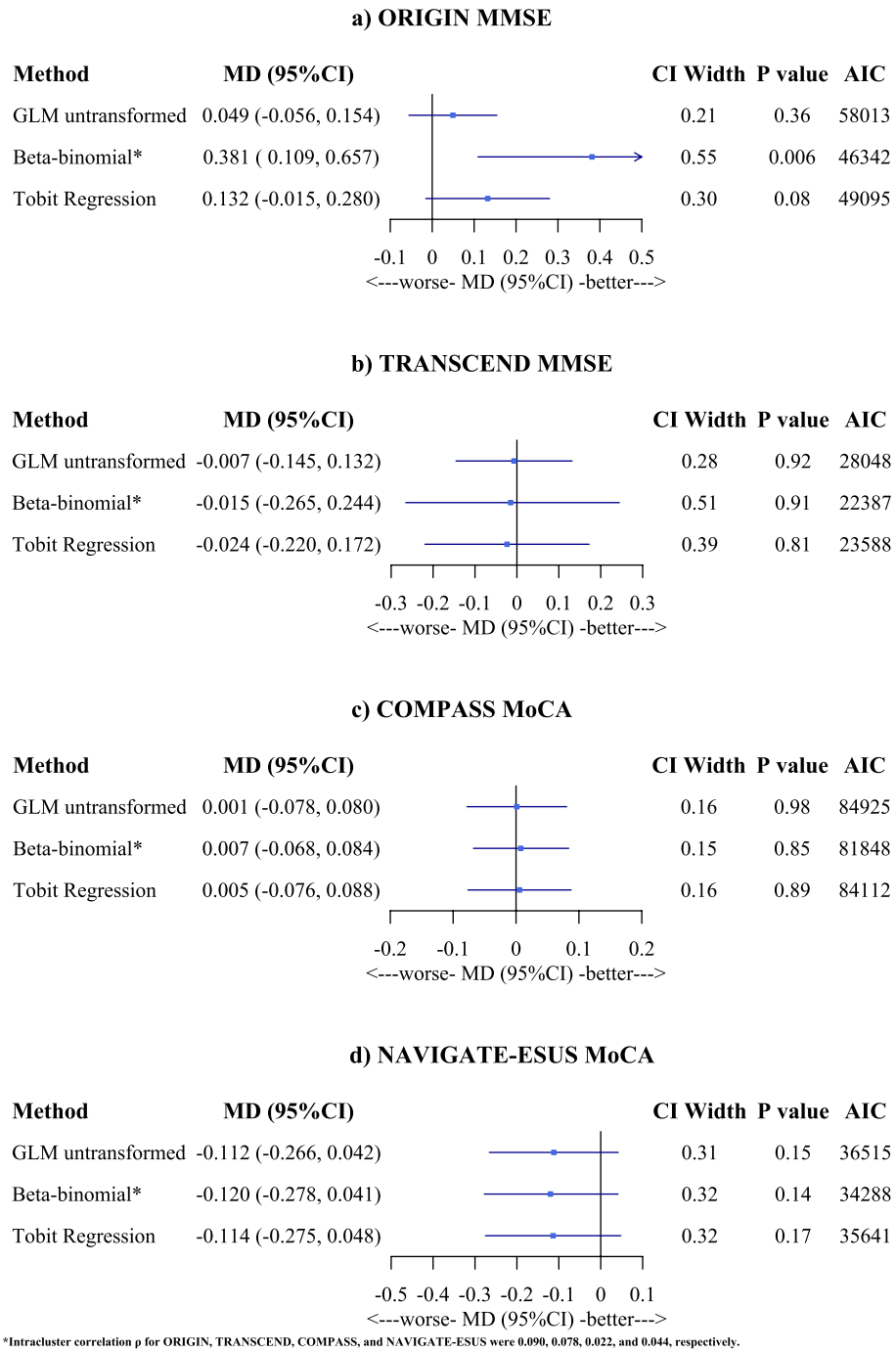
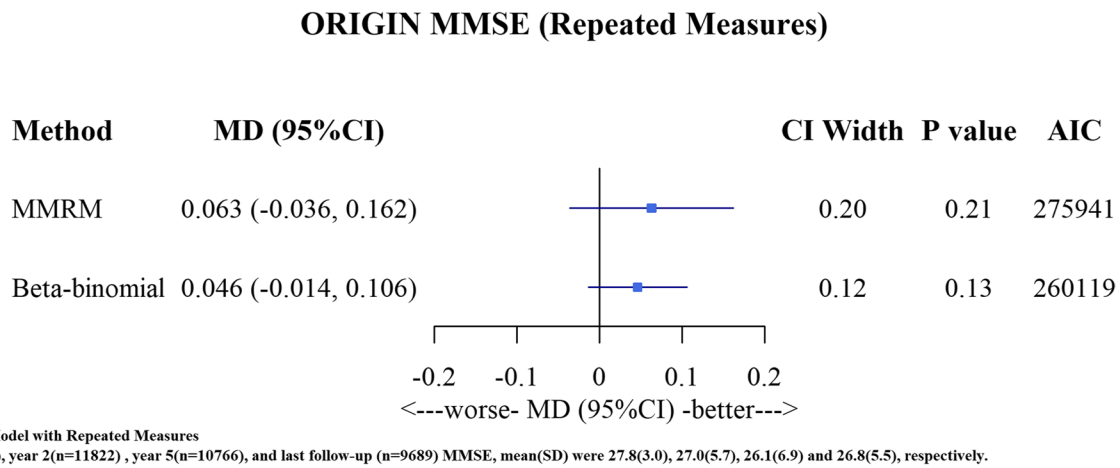



**Fig. 2** Between-group mean difference and model performance given by AIC values for the selected methods

Next, we compared the MLR with two other regression models not only the model performance but also the treatment effects with the three models had the same unit of measure for the cognitive scores.

In the ORIGIN study, the beta-binomial regression model showed a lysignificant higher MMSE score of 0.38 units (95% CI, 0.11 to 0.66; CI width = 0.55, $p = 0.006$; $AIC = 46{,}342$) in the treatment group as compared to the control group. In contrast, both the untransformed MLR (0.049; 95% CI, − 0.06 to 0.15; CI width = 0.21; $p = 0.36$; AIC = 58,013) and the Tobit regression model (0.13; 95% CI, − 0.02 to 0.28; CI width = 0.30; $p = 0.08$; AIC = 49,095) showed no significant treatment effect (Fig. 2). A substantial AIC reduction was found in comparing between




MMRM: Mixed Model with Repeated Measures
Baseline(n=12354), year 2(n=11822) , year 5(n=10766), and last follow-up (n=9689) MMSE, mean(SD) were 27.8(3.0), 27.0(5.7), 26.1(6.9) and 26.8(5.5), respectively.

**Fig. 3** Between-group mean difference and model performance given by AIC values for the ORIGIN study with repeated measures

the MLR untransformed and the beta-binomial. The intracluster correlation coefficients obtained for the beta-binomial models were 0.090, 0.078, 0.022, and 0.044 for ORIGIN, TRANSCEND, COMPASS, and NAVIGATE-ESUS, respectively. Although treatment effects were insignificant across all models for the other studies, a consistent trend was observed in which all three approaches had similar treatment effect estimates. Moreover, the beta-binomial models consistently exhibited slightly lower AIC values, even when compared to the MLR with transformations.

Similar results were found in the ORIGIN study with repeated measures, where the beta-binomial and MMRM showed similar treatment effect differences. However, the beta-binomial resulted in a narrower confidence interval width and a smaller AIC (Fig. 3).

The baseline cognitive score, baseline age, sex, and education level were mostly significantly associated with the dependent variable across the MLR untransformed, the beta-binomial, and the Tobit regression. Except in the NAVIGATE-ESUS study, the association between sex and cognitive scores was only found significant in the beta-binomial, and Tobit regression, but became not significant in the untransformed MLR. The degree of the associations between the covariates and the outcomes varied with the different approaches but the direction of the association remained consistent (Supplemental Table 1). The assumption of linearity was met, with points mostly hovering around the horizontal line for the generalized linear regression model, beta-binomial, and Tobit regression (Supplemental Figs. 6–7).

## Discussion

We used data from four international clinical trials to empirically evaluate the performance of different modeling approaches to analyze cognitive measures, specifically MMSE and MoCA. The methods used were the standard MLR, country-standardized MLR, squared root MLR, log-transformed MLR, beta-binomial, and Tobit regression. The beta-binomial consistently had the lowest AIC values compared to other approaches for both MMSE and MoCA in all studies, suggesting that this approach provided a better effect size estimate associated with an improved model fit, particularly for bounded data. It is also reassuring that the beta-binomial provides a reasonable and good fit, along with unbiased estimates, when the intracluster correlation coefficient is less than 0.1, according to the simulation [26]. The squared root MLR and the log-transformed MLR show a reasonable improvement on the model performance from the MLR standard, but it was challenging to make sense of the between-group mean difference relative to the actual cognitive scores. The effect sizes were significantly different between the beta-binomial and the standard MLR models only in ORIGIN, which had a substantial AIC reduction. For the Tobit regression, there was a small, improved model fit with smaller AIC as compared with the MLR standard. All the four studies had a small effect size and varying magnitude and significance of the covariate association across different models. Based on these results, we recommend considering the beta-binomial to avoid any spurious significance in comparing treatment effects or assessing risk factor effects in discrete and bounded cognitive measures during clinical trials Several studies [7–9] support these findings. One study [7] highlighted issues with applying a linear mixed model to outcomes due to ceiling and floor effects. Another study [8] also concluded that the beta-binomial distribution provided the best fit and found similar results when comparing the performance of four distributions (normal, t-family, binomial, and beta-binomial) in analyzing longitudinal data with the MMSE as the response variable. Lastly, another study [9] demonstrated

how to analyze and interpret health-related quality of life data using beta-binomial regression.

Due to data availability in four cardiovascular clinical trials, comparisons of different approaches were primarily conducted on response outcomes at two time points (baseline and follow-up). Additionally, we compared the mixed model with repeated measures and the beta-binomial model for the ORIGIN study, which included four time points. If more time points were available, it would be worthwhile to explore the performance of these analysis approaches with a greater number of time points and different methodologies to study the score variations.

## Conclusions

These findings demonstrate that beta-binomial regression analyses of continuous cognitive scores (that are derived from clinical instruments with floor and ceiling effects) will optimally assess the effect of an intervention on cognitive status within a randomized clinical trial. It allows for a direct interpretation of results and provides an unbiased estimate in trials using cognitive scores with non-standard psychometric properties. These findings are relevant and applicable to any analysis concerning outcomes influenced by floor and ceiling effects. We aimed to provide non-simulated empirical evidence for outcomes influenced by floor and ceiling effects. Our future work will explore how to estimate the power when analyzing the outcomes using a beta-binomial regression and incorporating repeated measures.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13063-024-08482-2.

Supplementary Material 1: Supplemental Table 1. Association of the covariates used to adjust the models. Supplemental Figure 1. Histograms of the MMSE or MoCA change score from baseline score for ORIGIN, TRANSCEND, COMPASS and NAVIGATE-ESUS. Supplemental Figure 2. Histograms of the MMSE residuals from the GLM with untransformed and transformed outcomes for ORIGIN and TRANSCEND. Supplemental Figure 3. Histograms of the MoCA residuals from the GLM with untransformed and transformed outcomes for COMPASS and NAVIGATE-ESUS. Supplemental Figure 4. Q-Q plot of the MMSE residuals from the GLM with untransformed and transformed outcomes for ORIGIN and TRANSCEND. Supplemental Figure 5. Q-Q plot for the MoCA residuals from the GLM with untransformed and transformed outcomes for COMPASS and NAVIGATE-ESUS. Supplemental Figure 6. Predicted vs Standardized Residual of the MMSE from the GLM, Beta-binomial and Tobit regression for ORIGIN and TRANSCEND. Supplemental Figure 7. Predicted vs Standardized Residual of the MoCA from the GLM, Beta-binomial and Tobit regression for COMPASS and NAVIGATE-ESUS. SAS and R codes for the analysis with one timepoint and multiple time points, respectively.

**Acknowledgements**
We thank the ORIGIN, ONTARGET/TRANSCEND, COMPASS, and NAVIGATE-ESUS Investigators for the use of their data.


**Authors' contributions**
SL conducted all the analyses and prepared the manuscript with guidance from GM and SB. WW and JB also contributed to the writing of the manuscript. All authors have reviewed and approved the final version of the manuscript.

**Funding**
None.

**Availability of data and materials**
The four clinical trials' data analyzed during the current study are not publicly available due to privacy or ethical restrictions but are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no competing interests.

**Author details**
[1]Population Health Research Institute, McMaster University, Hamilton, ON, Canada. [2]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada. [3]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [4]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [5]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada. [6]Department of Epidemiology and Preventive Medicine, Sackler Faculty of Medicine, School of Public Health, Tel Aviv University, Tel-Aviv, Israel. [7]Institute of Endocrinology, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel. [8]HRB-Clinical Research Facility, National University of Ireland, Galway, Ireland. [9]Department of Geriatric and Stroke Medicine, Galway University Hospital, Newcastle Road, Galway, Ireland. [10]Department of Medicine, McMaster University, Hamilton, ON, Canada. [11]Ohio University Heritage College of Osteopathic Medicine, Athens, USA.

Received: 24 January 2024 Accepted: 17 September 2024
Published online: 02 October 2024

**References**

1. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. J Psychiatr Res. 1975;12(3):189–98.
2. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc. 1992;40(9):922-35.
3. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695-9.
4. Proust-Lima C, Amieva H, Dartigues JF, Jacqmin-Gadda H. Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. Am J Epidemiol. 2007;165(3):344–50.
5. Gerstein HC, Colhoun HM, Dagenais GR, Diaz R, Lakshmanan M, Pais P, Probstfield J, Riesmeyer JS, Riddle MC, Rydén L, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. The Lancet. 2019;394(10193):121–30.
6. Tabachnick BG, Fidell LS, Ullman JB. Using multivariate statistics. In., Seventh edition edn. New York: Pearson; 2019.
7. Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. Misuse of the linear mixed model when evaluating risk factors of cognitive decline. Am J Epidemiol. 2011;174(9):1077–88.
8. Muniz-Terrera G, Hout AVD, Rigby RA, Stasinopoulos DM. Analysing cognitive test data: distributions and non-parametric random effects. Stat Methods Med Res. 2016;25(2):741–53.

9. Najera-Zuloaga J, Lee DJ, Arostegui I. Comparison of beta-binomial regression model approaches to analyze health-related quality of life data. Stat Methods Med Res. 2018;27(10):2989–3009.
10. Tobin J. Estimation of Relationships for Limited Dependent Variables. Econometrica. 1958;26(1):24–24.
11. Twisk J, Rijmen F. Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects. J Clin Epidemiol. 2009;62(9):953–8.
12. Terrera GM, Minett T, Brayne C, Matthews FE. Education associated with a delayed onset of terminal decline. Age Ageing. 2014;43(1):26–31.
13. Akaike H. On the Likelihood of a Time Series Model. J Roy Stat Soc. 1978;27(3/4):217–35.
14. Gerstein HC, Bosch J, Dagenais GR, Díaz R, Jung H, Maggioni AP, Pogue J, Probstfield J, Ramachandran A, Riddle MC, et al. Basal Insulin and Cardiovascular and Other Outcomes in Dysglycemia. N Engl J Med. 2012;367(4):319–28.
15. Cukierman-Yaffe T, Bosch J, Diaz R, Dyal L, Hancu N, Hildebrandt P, Lanas F, Lewis BS, Marre M, Yale J-F, et al. Effects of basal insulin glargine and omega-3 fatty acid on cognitive decline and probable cognitive impairment in people with dysglycaemia: a substudy of the ORIGIN trial. Lancet Diabetes Endocrinol. 2014;2(7):562–72.
16. Yusuf S, Teo K, Anderson C, Pogue J, Dyal L, Copland I, Schumacher H, Dagenais G, Sleight P. Effects of the angiotensin-receptor blocker telmisartan on cardiovascular events in high-risk patients intolerant to angiotensin-converting enzyme inhibitors: a randomised controlled trial. Lancet. 2008;372:1174–83.
17. Anderson C, Teo K, Gao P, Arima H, Dans A, Unger T, Commerford P, Dyal L, Schumacher H, Pogue J, et al. Renin-angiotensin system blockade and cognitive function in patients at high risk of cardiovascular disease: analysis of data from the ONTARGET and TRANSCEND studies. Lancet Neurol. 2011;10(1):43–53.
18. Eikelboom JW, Connolly SJ, Bosch J, Dagenais GR, Hart RG, Shestakovska O, Diaz R, Alings M, Lonn EM, Anand SS, et al. Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. N Engl J Med. 2017;377(14):1319–30.
19. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI, Berkowitz SD, Swaminathan B, Lavados P, Wang Y, Wang Y, et al. Rivaroxaban for Stroke Prevention after Embolic Stroke of Undetermined Source. N Engl J Med. 2018;378(23):2191–201.
20. Bosch J, Pearce LA, Sharma M, Canavan M, Whiteley WN, Mikulik R, Mundl H, Yusuf S, Hart RG, O'Donnell MJ. Rivaroxaban versus aspirin on functional and cognitive outcomes after embolic stroke of undetermined source: NAVIGATE ESUS trial. J Stroke Cerebrovasc Dis. 2022;31(5):106404.
21. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data In. 1982;38:963–74.
22. Huppert FA, Cabelli ST, Matthews FE. Brief cognitive assessment in a UK population sample - Distributional properties and the relationship between the MMSE and an extended mental state examination. BMC Geriatrics .2005;5.
23. Simonoff J. Analyzing Categorical Data. New York: Springer; 2003.
24. Guimarães P. A Simple Approach to Fit the Beta-binomial Model. Stand Genomic Sci. 2005;5(3):385–94.
25. Liu CF, Burgess JF Jr, Manning WG, Maciejewski ML. Beta-binomial regression and bimodal utilization. Health Serv Res. 2013;48(5):1769–78.
26. Kim J, Lee JH. The Validation of a Beta-Binomial Model for Overdispersed Binomial Data. Commun Stat Simul Comput. 2017;46(2):807–14.


## Publisher's Note