2017

Multistate Markov Models for Ordinal Functional Outcomes of Acute Onset Disease: Application in Acute Stroke Therapy Trials

Christy Cassarly
Medical University of South Carolina

Multistate Markov Models for Ordinal Functional Outcomes of Acute Onset Disease: Application in Acute Stroke Therapy Trials

Christy Cassarly

A dissertation submitted to the faculty of the Medical University of South Carolina in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the College of Graduate Studies.

Department of Public Health Sciences

2017

Approved By:

Co-Chair, Advisory Committee

Renee' Hebert Martin, Ph.D.

Co-Chair, Advisory Committee

Yuko Y. Palesch, Ph.D.

Marc Chimowitz, MBChB

Edsel Peña, Ph.D.

Viswanathan Ramakrishnan, Ph.D.

Acknowledgements

This project was supported by the NINDS/NIH Statistical and Data Coordinating Center for the ATACH-2 grant 5U01NS061861, P.I.: Yuko Y. Palesch, Ph.D.

This dissertation required the intellectual, emotional, mental and sometimes even physical support of a network of family, friends and mentors that I will never be able to adequately express my gratitude to in a few paragraphs. Nonetheless, I will do my best to thank them.

I feel honored to have had my two advisors, Dr. Renee' Hebert Martin and Yuko Y. Palesch, guide me through this process. The person who began the graduate program in 2011 was an entirely different person than I am today. I directly attribute my intellectual growth to the support of these two women. I am proud of the progress I have made as a researcher and writer due to their guidance, motivation and knowledge. Despite busy schedules, I was able to spend countless hours in their offices discussing my research with their full attention. They were also there for me at times when I needed help through unrelated, personal life challenges. I could have never imagined what valuable pieces of my life these mentors would become and I know I have made two lifelong friends. Thank you both for believing in me, even when I didn't believe in myself.

I also want to thank Dr. Viswanathan Ramakrishnan who was a major source of statistical knowledge for my research. Dr. Ramesh would sit with me, at times for hours,

reading drafted papers and carefully scrutinizing content, language and grammar. I am so grateful for the many lessons I learned from Dr. Ramesh ranging from statistical theory to proper push-up form. Thank you, Dr. Ramesh, for being so much more than just an advisory committee member.

I was fortunate to have a well-rounded dissertation committee complete with five members who truly cared about my success. Every single member of my committee carefully scrutinized the drafts of papers I would send and consistently gave thoughtful comments and suggestions. Thank you, Dr. Marc Chimowitz, for all of your clinical expertise and helping me stay focused on the big picture. Thank you, Dr. Edsel Peña, for your ability and willingness to carefully scrutinize the theoretical details of my work. The expertise of each member of my advisory committee was essential and I would not have been able to complete this work without their guidance.

I am grateful for the experience I gained through opportunities to practice the skills I learned throughout the entire course of my education. I want to thank Drs. Martin and Palesch for allowing me to learn from and work with the Data Coordination Unit. I want to specifically thank the Antihypertensive Treatment of Acute Cerebral Hemorrhage (ATACH) II team for teaching me many lessons on the reality of the application of biostatistics to clinical trials. This invaluable hands-on experience has given me a head start on honing my skills as I begin my career.

I want to thank the dedicated faculty of the Department of Public Health at MUSC. It is hard to express how meaningful it is to have a professor that can be present in a classroom even when they are insanely busy. I also want to thank my fellow

classmates in the program. I am so grateful to have met so many talented and intelligent young biostatisticians. Thank you for being lifelines when things got hard (AQE preparation, anyone?) and helping keep me sane. I am excited to know that we will maintain our friendships into the future.

I also want to acknowledge mentors of days past. Specifically, I want to thank Dr. Marlow Lemons for being the reason I pursued statistics in the first place. Dr. Lemons' passion for statistics was evident on the first day of my introductory statistics course required for my psychology major at Virginia Tech. The infectious nature of his enthusiasm quickly converted me to statistics. With his guidance and after taking six courses, I set out to complete my Ph.D., a degree I had never previously considered. Had I never crossed paths with Dr. Lemons, I can't imagine I would find myself where I am today. Thank you, Dr. Lemons, for being a one-of-a-kind educator.

I would also like to thank all of my wonderful friends, near and far. I have the most incredible support system and there is no way to properly thank each person in it. No matter what life has thrown at me, my friends have been there for me with support, patience and love. Thank you to my South Carolina family, for making Charleston feel like home.

Finally, I want to thank my family. To articulate the gratitude I have for my mom, dad, and sister, Lauren, is an impossible task. I have always been so well supported at every stage of my life by my parents and my sister, who are my best friends. Thank you for your unconditional love. Thank you for being my rocks and your willingness to drop whatever you're doing to make sure I am okay. Thank you, Lauren, for being a friend and

a confidant- I can always count on you for your wisdom and maturity and am constantly impressed by the amazing woman you have become. Finally, thank you, Mom and Dad, for being the parents I hope to be some day- you have provided me with every opportunity I could ever imagine and your love for your family and your enjoyment of life inspires me every minute of every day. I owe everything that I am to you and I dedicate this dissertation to you.

TABLE OF CONTENTS

Abstract

CHRISTY CASSARLY. Multistate Markov Models for Ordinal Functional Outcomes of Acute Onset Disease: Application in Acute Stroke Therapy Trials. (Under the direction of RENEE' HEBERT MARTIN and YUKO Y. PALESCH)

The modified Rankin Scale (mRS), a seven-point ordinal scale ranging from no symptoms to death, is the most commonly used outcome measures in acute stroke therapy trials. Often, one visit is chosen for the primary analysis, and the scale is dichotomized leading to loss of information. Recently, alternative methods for analyzing the mRS have been explored. In addition, acute onset conditions require immediate attention and treatment, posing a challenge to assess baseline outcome measures for clinical trials. Thus, the mRS is not obtainable at baseline. Much of the progression or recovery experienced by a patient suffering from an acute onset disease is expected to occur early on. Moreover, typically, the goal of a treatment or therapeutic action is improvement in patient health compared to their baseline measure. To accurately quantify improvement, a measure of the outcome at baseline is ideal. This dissertation first explores the feasibility of multistate Markov models for the analysis of the mRS which allow for the full ordinal scale as well as the repeated measures data to be incorporated. The operating characteristics (type I error and power) of the multistate Markov model are compared with those from repeated logistic regression. Next, a framework is developed to predict and incorporate the latent baseline mRS score in a piecewise-constant multistate model. The last part of this work applies the piecewise-constant latent baseline model to real

acute stroke trial data and compares the results with alternative methods for analysis of the mRS.

1 Introduction and Significance

1.1 Overview and Specific Aims

Ordinal response outcomes are often used in clinical trials. However, rather than analyzing the full ordinal scale, many trials choose to dichotomize the primary outcome. Although models used for dichotomous outcomes are easier to implement and tend to produce summary statistics with more clinically meaningful interpretations, dichotomization can result in a loss of statistical power [1]. Additionally, in trials where long-term follow-up is planned, the outcome is collected at multiple visits. Despite the availability of the repeated measures, many trials focus the primary analysis on the data from one visit, ignoring the additional outcome data.

One example of a therapeutic area that collects an ordinal outcome at multiple visits in clinical trials is acute stroke therapy. For many such trials, the modified Rankin Scale (mRS) score at 90 days post-randomization is used as the primary outcome measure [2]. The mRS is a seven-point ordinal scale that ranges from 0 (no symptoms) to 6 (dead) and measures functional independence of stroke patients. It is commonly dichotomized to test the primary hypotheses of interest.

An emphasis has been placed on exploring alternative analytic methods for the analysis of mRS outcome data from acute stroke trials in recent years. Results indicate that the original structure of the scale needs to be maintained in analysis as much as possible [1, 3]. A number of alternative methods that preserve the ordinality of the mRS have been proposed. However, these methods have not been widely accepted in practice.

The ultimate goal of a treatment or therapeutic action is to improve patient health compared to their baseline measure at presentation, immediately following an event. To accurately quantify improvement, a measure of the outcome at baseline would be ideal. Conditions with sudden onset, such as stroke, require immediate attention and treatment, posing a challenge to assess baseline outcome measures for clinical trials. Thus, the mRS is not obtainable at baseline making the quantification of "improvement" very challenging.

The multistate Markov model (MSMM) in continuous-time analyzes ordinal data and has been used to model the course of many diseases [4]. These models are advantageous in clinical applications where a disease process naturally moves through increasing stages of severity [4]. The feasibility of MSMMs for analysis of the mRS has not been previously considered. The mRS has more disease states (here, the seven levels of the scale) than most clinical applications of MSMMs. Most of the subjects that transition to a different state experience adjacent-state transitions, with only a few non-adjacent state transitions. The combination of these two issues leads to a data structure, henceforth referred to as sparsely populated ordinal data, where small cell counts are observed for transitions to non-adjacent states. Currently, there is little information available regarding the appropriateness of MSMMs for sparsely populated ordinal outcomes.

This dissertation aims to address the issues presented here, and the specific aims of this research are as follows:

1. To explore the operating characteristics (type I error and power) of MSMMs compared with repeated logistic regression used to analyze sparsely populated repeated measures ordinal data.
2. To develop a MSMM approach with piecewise-constant transition intensities that incorporates a latent baseline state.
3. Analyze acute stroke therapy trial data using the methods developed in Aim 2 and compare the results with those from alternative methods previously suggested for the analysis of the mRS.

Realization of these aims will achieve the following: (1) feasibility of MSMMs for sparsely populated ordinal data will be demonstrated through investigation of sample size needed to achieve adequate power; (2) validity will be demonstrated through simulation studies where type I error is preserved; (3) efficiency of inclusion of the baseline mRS in the MSMM will be demonstrated; and (4) efficiency of the MSMM as compared with other methods for ordinal outcome data will be shown.

1.2 Motivation and Clinical Relevance

Each year, approximately 795,000 people have a stroke, 87% of which are ischemic [5]. Most randomized trials in acute stroke neuroprotection treatment have failed to show efficacy [3]. Several explanations have been proposed to describe the lack of positive trials in stroke, including heterogeneity in stroke pathophysiology, poor methodological and statistical standards, and incomplete preclinical testing [6]. One is poor study design and statistical methods, specifically, the analysis of the primary outcome [1]. The models that exist to fit dichotomous and continuous outcomes are easier to implement and tend to

produce summary statistics with more clinically meaningful interpretations. However, analysis of ordinal response outcomes are less straightforward. Thus, traditionally, many trials have dichotomized the ordinal mRS (Table 1.1) into success, scores of 0 or 1 (or 0 to 2), or failure, scores greater than 1 (or 2), for the primary analysis, often collected at 90 days post-randomization [7].

Table 1.1. Modified Rankin Scale.

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities but able to look after own affairs without assistance
3	Moderate disability requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent, and requiring constant nursing care and attention
6	Dead

Some patients with severe stroke may never have the potential to achieve a "success" as defined by the dichotomy because they are so severely disabled at baseline. Patients with minor strokes may achieve a successful score more easily than those who are more disabled at baseline [8]. Thus, the prognostic heterogeneity of subjects does not allow for potential equal contribution to the treatment effect estimation for all subjects [9]. In general, ignoring these differences and dichotomizing the ordinal scale reduces statistical power [1]. Any reduction in power may result in failure to find a clinically meaningful treatment effect. In addition, the recovery and outcomes of subjects following a stroke realistically lie on a continuum. Categorical analysis of the ordinal scale provides

a more comprehensive quantification of the process than the analysis of the dichotomized scale [10].

An additional drawback of the traditional analysis - dichotomization of the 90 day ordinal outcome - is the lack of use of available longitudinal data. Many acute ischemic trials assess and collect the mRS score at discharge and/or at 30 days from randomization and also at periodic intervals through 12 months, if long-term follow-up is planned in the trial. However, the longitudinal data are rarely used in the primary analysis. A more comprehensive clinical understanding of the treatment effect on outcome after a stroke may be better described using repeated measures analysis if improvement or worsening is expected beyond the 90 day primary outcome [11].

Recently, an emphasis has been placed on exploring alternative outcomes as well as other analytic methods for the analysis of mRS data from acute stroke trials (continuous analysis- t-test, linear regression; ordinal analysis- shift analysis, proportional odds model, partial proportional odds model, adjacent categories logit model; sliding dichotomy; utility weighted mRS; repeated measures analysis). The literature indicates that the mRS should be analyzed in such a way that maintains the original structure of the scale as much as possible [1, 3]. Alternative analytic strategies proposed for analysis of the mRS have not been widely accepted in practice. These strategies are reviewed in depth in the following section.

MSMMs are proposed to analyze the longitudinal mRS scores. An example of the typical structure of the observed transition matrices for the mRS over time is provided in Table 1.2. In this example, mRS outcome data from a mock acute stroke trial of 1,000 subjects are observed for four follow-up visits to illustrate the structure of sparsely

populated ordinal data. The transitions that occur from one visit to the next are presented in each of the matrices.

Table 1.2: mRS Transition Example.

mRS at Time 1	mRS at Time 2							
	0	1	2	3	4	5	6	Total
0	84	20	5	1	0	0	0	110
1	45	79	17	2	1	0	1	145
2	12	50	56	15	8	6	1	148
3	2	25	46	63	17	3	0	156
4	1	2	23	77	89	23	5	220
5	0	0	2	7	41	48	22	120
6	0	0	0	0	0	0	101	101
Total	144	176	149	165	156	80	130	1000

mRS at Time 2	mRS at Time 3							
	0	1	2	3	4	5	6	Total
0	108	26	5	3	1	1	0	144
1	38	123	12	2	0	1	0	176
2	9	31	86	18	2	1	2	149
3	3	5	32	116	8	1	0	165
4	0	1	4	34	102	13	2	156
5	0	0	0	2	18	45	15	80
6	0	0	0	0	0	0	130	130
Total	158	186	139	175	131	62	149	1000

mRS at Time 3	mRS at Time 4							
	0	1	2	3	4	5	6	Total
0	126	27	4	0	1	0	0	158
1	21	146	14	4	1	0	0	186
2	5	26	93	12	2	1	0	139
3	0	4	28	129	11	1	2	175
4	0	2	0	25	94	7	3	131
5	0	0	0	1	10	34	17	62
6	0	0	0	0	0	0	149	149
Total	152	205	139	171	119	43	171	1000

In this example it is clear that the majority of subjects actually stay in the same state from one time period to the next as shown by the largest numbers along the main diagonal. The second largest numbers are to adjacent states and very few to non-adjacent states. For example, 79 of the 145 subjects that had mRS = 1 at Time 1 also had mRS = 1 at Time 2. Only 17 of the 145 subjects with mRS = 1 at Time 1 transitioned to mRS = 2 at Time 2. This is an example of an adjacent-state transition, one where a subject moves from one state (mRS = 1) to an adjacent state (mRS = 2). The other adjacent-state transition for mRS = 1 is the transition to mRS = 0. Throughout the table, a majority of observations are of subjects that remain in the same state, or have the same mRS score from one time to the next. Most of the subjects that transition to a different state experience adjacent-state transitions, with only a few non-adjacent state transitions.

In order to assess the application of the MSMM and number of states modeled, a literature review was conducted. Using the following keywords: multistate, Markov, panel, clinical, application, continuous-time, and excluding the following words: piecewise, non-homogeneous, inhomogeneous, semi-Markov, hidden Markov and random effects, a total of 40 articles were identified. An article was excluded if (a) the content was actually theoretical and there was no application, (b) it was a review with no new content, (c) multistate models were referenced, flagging it for review but the models were not actually fit, or (d) the models were actually discrete-time. Of the remaining 26 articles, 25 fit models to data with five or fewer states and two fit models to data with six states [12-36]. One publication used a six-state model to analyze a dataset with much more data than is typically collected in acute stroke trials- approximately 5,000 patients

[37]. Thus, the feasibility of MSMM for analysis of sparsely populated ordinal data with a large number of states is unclear.

Currently, there is little information available regarding the appropriateness of MSMMs for sparsely populated ordinal outcomes with a large number of possible health states, as observed in longitudinal mRS data from acute stroke trials. In this dissertation, feasibility and operating characteristics of MSMMs applied to sparsely populated ordinal data are examined. In addition, a method is proposed to incorporate the latent baseline mRS score in longitudinal MSMMs. The mRS is unavailable at baseline and the aforementioned analytic techniques used for mRS data have adjusted for baseline severity using the National Institutes of Health Stroke Scale (NIHSS), a score that ranges from 0 (no neurological deficit) to 42. Inclusion of the latent baseline mRS predicted using the baseline NIHSS and other baseline covariates in a MSMM could improve statistical efficiency to detect a significant treatment effect.

1.2.1 Motivating Examples

Data from two randomized double-blind, placebo-controlled acute stroke therapy trials are considered. The NINDS tissue plasminogen activator (t-PA) trial was designed to compare t-PA versus placebo in patients with acute ischemic stroke. The trial had two parts. Part 1 tested whether patients treated with t-PA had early improvement, as compared with those that were given placebo [38]. Part 2 was designed to determine whether there was a consistent and persuasive difference between the groups tested using four outcomes (Barthel Index, mRS, Glasgow Outcome Scale, and NIHSS) at 90 days modeled as a Global Test Score [39]. The Barthel Index is an index of independence that

scores the ability of patients to care for themselves [40]. Patients that can perform all assessed activities with complete independence are given a score of 100. The Glasgow outcome scale is a global assessment of function that ranges from 1 indicating good recovery to 5, death [41]. In order to be considered for inclusion for enrollment, there had to be deficiency measureable by the NIHSS. A total of 624 subjects were enrolled (291 in Part 1 and 333 in Part2), 312 in each group [38]. In Part 1, a benefit was observed for patients treated with t-PA in all four outcome measures. The primary analysis in Part 2, using generalized estimating equations, showed a significant global test score for the four outcomes [39]. Clinical and demographic characteristics can be found in the original paper [38].

The Albumin in Acute Stroke (ALIAS) trial was a two part trial designed to compare 25% human serum albumin (ALB) and saline in patients with acute ischemic stroke. Part 1 consisted of two separate, concurrently implemented trials designed to assess whether ALB therapy improved neuroprotection beyond standard of care in two cohorts of patients [42]. One cohort consisted of subjects that received standard thrombolytic therapy (intravenous t-PA, intra-arterial t-PA, endovascular mechanical thrombolysis or a combination of intravenous and endovascular treatment) and the other was subjects who were not thrombolysed. Part 1 was suspended for safety reasons after 434 subjects (207 albumin and 217 saline) were enrolled. More patients died in the first 30 days in the ALB group than the placebo group and deaths were increased in patients older than 83 years and patients that received excessive intravenous fluids. The study design for Part 2 was modified based on the safety findings in Part 1. The primary endpoint was a composite outcome defined as a NIHSS 0-1 and/or mRS 0-1 at 90 days

from randomization. Only patients with a baseline NIHSS of 6 or above were eligible for enrollment in the trial. Part 2 was stopped early for futility after 841 subjects were randomized (422 to albumin and 419 to saline). Clinical and demographic characteristics are described elsewhere [43].

2 Background

2.1 Previous Analytic Methods for Ordinal Outcomes

In this section, alternative analytic approaches to dichotomized analysis for the mRS are explored. A summary of all of the methods is provided in Table 2.1.

2.1.1 Continuous Analysis

In general, analysis of continuous variables with the t-test and linear regression is straightforward and produces clinically intuitive summary statistics. When these methods are applied to ordinal scales, the results are less interpretable. Non-integer values from an ordinal scale do not have a clear meaning when they are considered to be continuous. When compared to ordinal analysis, continuous analysis has been shown to have comparable power; however, the normality assumption is not met in most studies of stroke outcome [3]. In order to consider an ordinal outcome to be continuous, the sample size must be large enough for the normal approximation to be valid. Even in large datasets, the mRS is skewed and there are no recommendations on how to normalize it.

2.1.2 Shift Analysis

The Cochran-Mantel Haenszel (CMH) shift test, or the van Elteren test, can be used to analyze the distribution of ordinal data [7]. This test can show whether a treatment causes a significant favorable shift toward better outcome. Shift analysis can account for ordered categories, has no distributional assumptions and is easy to implement. However, it is not feasible for large scale clinical trials with non-simple randomization schemes as it only allows for a limited number of covariates. Logistic regression can be used in conjunction

with the shift analysis to provide an estimate of treatment effect because there is not an associated odds ratio (OR) or effect size produced from the CMH test [44].

2.1.3 Ordinal Analyses

The proportional odds model (POM) assumes an identical effect of the predictors for each cumulative probability [45]. In other words, the OR comparing treatment to placebo in patients with mRS of 0 versus 1-6 is assumed to be the same as the OR for mRS of 0-1 versus 2-6, and so on. If the proportional odds assumption holds, statistical power can be increased compared to analysis using a strict dichotomy. If the assumption fails, this analysis could mask important effects at one end of the ordinal outcome [46]. The score test for assessing the proportional odds assumption is anticonservative and DeSantis and colleagues illustrated the lack of power of the score test using the data from the NINDS t-PA trial [46]. The score test failed to reject the assumption of proportional odds ($p = 0.06$); however, a plot of the cumulative log odds of each mRS score for each treatment group indicated that the assumption may be inappropriate.

The partial proportional odds model (PPOM) can be used when the proportional odds assumption does not hold [46]. In general, there are two types of PPOMs, an unconstrained and a constrained model [47]. The unconstrained PPOM produces cut-off point-specific odds ratios. Alternatively, if a pattern is expected in the cut-off point-specific odds ratio, for example a linear trend, constraints could be placed on the parameter to obtain an appropriate fit. A linear trend occurs when the violation of the proportional odds assumption is in one direction. This model includes an additional parameter that allows for the ORs to increase proportional to the outcome scale. One

drawback of this model is that it could require a larger sample size to be adequately powered.

The adjacent categories logit model (ACAT) is another logistic regression model that does not require the proportional odds assumption. These models utilize single-category probabilities rather than cumulative probabilities [45]. Rather than effects that refer to the entire response scale in the POM, the ACAT effects refer to the effect of a predictor on the response in any two adjacent categories. As with the PPOM, the ACAT model may require a larger sample size to be adequately powered to detect a treatment effect.

2.1.4 Sliding Dichotomy Analysis

The sliding dichotomy (or the more generalized sliding trichotomy or tetrachotomy) allows for the definition of success to vary based on patient-specific baseline prognostic variables while maintaining a dichotomized outcome [48]. Re-analysis of acute stroke therapy trials uses pre-specified cut-points for prognosis group definition based on the NIHSS score [49]. The sliding dichotomy can be used to define “mild”, “moderate”, and “severe” stroke using the baseline NIHSS score and defines “success” for each of the three groups. One example is to define favorable outcome as mRS = 0 for mild strokes, mRS = 0-1 for moderate strokes and mRS = 0-2 for severe strokes. Since baseline severity is such a strong predictor of outcome in stroke patients, this baseline severity adjusted approach has been considered for use over the traditional dichotomy [48].

Some simulation studies have shown that the utilization of the sliding dichotomy provides higher sensitivity to detect true treatment effects [50]. For example, when the

probability of favorable outcome is high (greater than 0.5), the sliding dichotomy provides higher power [9]. This is not a general result; however, as other studies have determined that the traditional dichotomy is more powerful than the sliding dichotomy in most situations [51]. When the probabilities of favorable outcome are lower, the traditional dichotomy is more powerful [9].

While the sliding dichotomy has the potential to be a powerful tool in some settings, it also has limitations. Determining the number of prognostic groups to use is not an obvious decision and can be difficult to justify. In addition, determining how to choose the cut points for the different groups can be a difficult task. Although using three groups (mild, moderate and severe) is used in the literature [48], methods used to determine severity cut points vary and need to be verified. Poor selection of the number of groups and cut points could result in a loss of power. In addition, while the sliding dichotomy allows for a baseline severity adjusted outcome, it still ignores any non- "success" transition from one mRS score to another even though each mRS category (except 5 to 6) represents a clinically meaningful difference in health state [52].

2.1.5 Utility-Weighted mRS

A recently proposed approach to transform the mRS into a patient-centered outcome measure is the utility-weighted mRS (UW-mRS) [53]. The chosen patient-centered outcome measure, utility, is the desirability of a specific health outcome to a patient [54]. Utility weights for each level of the mRS were derived by averaging values derived in two prior studies. The first study mapped mRS scores to the European Quality of Life Scale (EQ-5D) in transient ischemic attack survivors from a population-based study in

Great Britain [55]. The second study derived weights using the methodology of the World Health Organization Global Burden of Disease Project [56]. Using the utility weights is straightforward, and the UW-mRS is analyzed using continuous analysis. Although this method is easy to implement and may provide greater statistical power, it is based on only two populations and may not be generalizable to other populations. In addition, the utility values are limited with respect to interpretation compared to other methods.

2.1.6 Longitudinal Analysis

Generalized estimating equations (GEE) can be used to estimate parameters for outcomes collected at multiple time points [57]. This approach allows for covariate adjustment while incorporating within-patient correlation. GEE analysis has been used to analyze the repeated measures of the mRS [11, 58]. Analysis of the longitudinal dichotomized outcomes yields clinically meaningful odds ratios. However, as the models increase in complexity, computational issues may arise.

Table 2.1. Summary of statistics obtained from each type of analysis of the mRS.				
Method	Statistic(s)	Interpretation	Strengths	Limitations
Logistic regression	Odds ratio (OR)	The odds of good outcome in the treatment group versus placebo	• Easy to apply • Easy to interpret • Clinically intuitive	• Can result in a loss of power • Requires prespecification of expected treatment distribution
Linear regression	Difference of means	Improvement of the average mRS score in patients that received treatment	• Easy to apply	• No straightforward interpretation • Normality assumption is often not met with no recommendations for normalizing
Shift analysis (CMH test)	Probability value	The treatment group shifted in a favorable direction toward a better mRS score versus placebo	• Easy to apply • Accounts for ordered categories • No distributional assumptions	• Accommodates a limited number of covariates • Not clinically intuitive- no effect size or odds ratio
POM	Summary OR	The odds of a lower mRS the treatment group versus placebo	• Easy to apply • Clinically meaningful summary odds ratio	• May yield biased estimate if proportional odds assumption is not met • Anticonservative score test
PPOM	ORs for six dichotomies of mRS	Treatment has a significant benefit for certain definitions of good outcome	• Does not require proportional odds assumption	• Less straightforward summary odds ratios • Can require a larger sample size to be adequately powered
ACAT	ORs for six adjacent categories	The treatment group is more likely to have smaller mRS for certain adjacent mRS scores	• Does not require proportional odds assumption	• Can require a larger sample size to be adequately powered
Logistic regression of sliding dichotomy	OR	The odds of good outcome (defined by baseline severity) in the treatment group versus placebo	• Easy to apply • Easy to interpret • Clinically intuitive	• Less power in some scenarios • Choosing groups and cut-points poorly leads to loss of power
Linear regression of UW-mRS	Difference of mean utility scores	Improvement of the average utility score in patients that received treatment	• Easy to apply • Can increase power	• May not be generalizable to other studies • Limited interpretability
Repeated measures GEE (dichotomized)	OR	The odds of good outcome over the 12-month period in the treatment group versus placebo	• Utilizes all longitudinal data • Clinically meaningful odds ratio	• More complicated modeling • May have computational difficulties
MSMM	Hazard ratios for each allowable transition	The hazard (instantaneous risk) of transitioning from one mRS state to another in the treatment group versus placebo	• Utilizes full ordinal scale • Utilizes all longitudinal data • Estimates transition rates for progression and recovery	• Difficulty in estimating sample size • Computationally intensive

2.2 Multistate Markov Models

2.2.1 General Multistate Markov Models

MSMMs are an alternative approach to analyze repeated measures data with an ordinal outcome. These types of models describe how a process moves between states over time, which is desirable in the description of disease processes that naturally move through increasing stages of severity [59]. MSMMs can provide a better clinical understanding of the disease process since the information from the entire course of the disease is used to estimate the parameters of the model. These models have been used in numerous clinical applications including: multiple sclerosis [60], periodontal disease [61], alcoholism [62], and psychiatry [63].

Figure 2.1 represents a general MSMM with four states. The arrows indicate that a transition can occur between any two states. The model estimates parameters describing each of the allowable transitions.

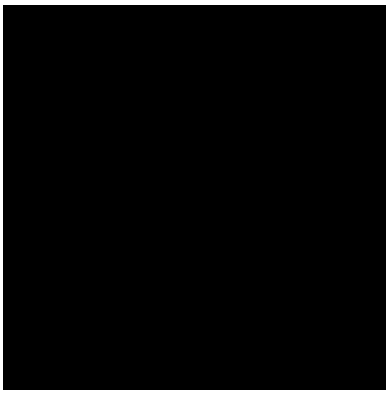


Figure 2.1: General four-state MSMM.

The use of MSMMs requires that the Markov property holds for the observed data. Consider a system with a finite state-space $\Omega = \{1,2,3,...,I\}$, where I represents the number of states. Let $X(s+t)$ be a discrete random variable that indicates the state occupied by the system at time $s+t$. Let $\mathrm{F}_{X(s)} = \{X(v) : v \le s\}$ which denotes all information pertaining to the history of X up to time s [64]. A series of observations has the Markov property if the conditional distribution of $X(s+t)$ given $\mathrm{F}_{X(s)}$, satisfies

$$P\left\{X(s+t) = j \mid \mathrm{F}_{X(s)}, X(s) = i\right\} = P\left\{X(s+t) = j \mid X(s) = i\right\} = p_{ij}(s,t),\ i, j \in \Omega \quad (2.1)$$

In other words, the present state depends only on the immediately preceding observation and not on the ones that precede it.

MSMMs may be defined for both discrete time and continuous. Although the course of disease is a continuous process, clinical trials often only collect data at intermittent follow-up visits. In the context of stroke, the exact time of progression or recovery, or change of state, of disease is not observed. Data of this type, representative of a continuous process yet observed at discrete time points, is referred to as panel data [65]. Both discrete and continuous time MSMMs can be used to describe panel data. In many acute stroke trials, the mRS is collected at follow-up visits that are not evenly spaced. In such instances, continuous time models are appropriate. A continuous model for panel data can only be used in cases where the sampling times are considered to be non-informative [66]. An example of non-informative sampling is a fixed observation scheme, where the interval of follow-up is specified in advance. However if observations are not fixed or random and are self-selected by the subject (informative), this modeling

technique is not appropriate without properly adjusting for the additional information [66]. For instance, these models cannot be used in a scenario where observations occur when a subject visits a doctor because they are in poor-condition. A model that incorporates the information from the sampling times must be used for this type of self-selected follow-up outcome data. In acute stroke trials, the follow-up visits are usually specified in advance and are non-informative so continuous modeling is appropriate.

A common assumption of continuous-time MSMMs is that of homogeneity, where transition probabilities remain constant over time. When homogeneity is assumed, the transition probabilities, $p_{ij}(t)$, are defined as

$$p_{ij}(t) = P\{X(s+t) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\} \ . \qquad (2.2)$$

Since this expression does not depend on s, the transition from state i to state j on a time interval of length t has the same probability at any time. The $p_{ij}(t)$ are elements of the transition probability matrix, $\mathbf{P}(t)$.

In order to construct continuous time Markov chains, the amount of time the process will remain in a state, $i \in \Omega$ must be determined. Suppose $X(0) = i$ and let T_i represent the amount of time a process stays in i after entering. To derive the distribution of T_i, let $s, t \geq 0$ and consider

$$
\begin{aligned}
&\Pr\{T_i > s+t \mid T_i > s\} \\
&\quad = \Pr\{X(r) = i \text{ for } r \in [0, s+t] \mid X(r) = i \text{ for } r \in [0,s]\} \\
&\quad = \Pr\{X(r) = i \text{ for } r \in [s, s+t] \mid X(r) = i \text{ for } r \in [0,s]\} \\
&\quad = \Pr\{X(r) = i \text{ for } r \in [s, s+t] \mid X(s) = i\} \qquad \text{(Markov property)} \\
&\quad = \Pr\{X(r) = i \text{ for } r \in [0, t] \mid X(0) = i\} \qquad \text{(time homogeneity)} \\
&\quad = \Pr\{T_i > t\}.
\end{aligned}
$$

Thus T_i satisfies the memoryless property and follows the exponential distribution.

The movement of a subject between states is described by λ_{ij}, the transition intensities:

$$\lambda_{ij} = \lim_{\delta t \to 0} \frac{P(X(\delta t) = j \mid X(0) = i)}{\delta t}, \text{ for } i \neq j \ . \tag{2.3}$$

The intensities represent the instantaneous rate of moving from state i to state $j \neq i$ and form the generator matrix, $\mathbf{\Lambda}$, whose rows sum to zero and the diagonal entries are $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$.

In each of the continuous time Markov chain constructions only the local behavior of the process is known. To determine the global behavior of the process, Kolmogorov differential equations to solve for the terms

$$p_{ij}(t) = \Pr\{X(t) = j \mid X(0) = i\}.$$

These are a system of ordinary differential equations describing the probabilities $p_{ij}(t)$. Two sets of Kolmogorov equations exist, forward and backward differential equations. The forward differential equations are used when the interest is to understand a process at

a future time. Backward differential equations are used to describe what happened at a previous time given that a process has a certain state in the future. Using the Chapman-Kolmogorov equation,

$$p_{ik}(\tau,t)=\sum_{j=1}^{\infty}p_{ij}(\tau,\xi)\,p_{jk}(\xi,\mathrm{t}),\ \text{where}\ \tau<\xi<t \tag{2.4}$$

the forward differential equations are derived as follows [67]:

$$
\begin{aligned}
\mathrm{p}'_{ij}(t) &= \lim_{h\to 0}\frac{p_{ij}(t+h)-p_{ij}(t)}{h} \\
&= \lim_{h\to 0}\frac{1}{h}\left(\sum_{y=1}^{k} p_{iy}(0,t)\,\mathrm{p}_{yj}(t,t+h)-p_{ij}(t)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(\sum_{y=1}^{k}\Pr\{X(t)=y\mid X(0)=i\}\Pr\{X(t+h)=j\mid X(t)=y\}-p_{ij}(t)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\Big(\Pr\{X(t)=j\mid X(0)=i\}\Pr\{X(t+h)=j\mid X(t)=j\} \\
&\qquad +\sum_{y=1,y\neq j}^{k}\Pr\{X(t)=y\mid X(0)=i\}\Pr\{X(t+h)=j\mid X(t)=y\}-p_{ij}(t)\Big) \\
&= \lim_{h\to 0}\frac{1}{h}\left(p_{ij}(t)\big(1-\lambda(j)h+o(h)\big)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\big)\big(\lambda(y,j)h+o(h)\big)-p_{ij}(t)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(\begin{array}{l} p_{ij}(t)\big(1-\lambda(j)h\big)+p_{ij}(t)o(h) \\ +\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)h\big)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)o(h)\big)-p_{ij}(t)\end{array}\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(p_{ij}(t)\big(1-\lambda(j)h\big)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)h\big)+\sum_{y=1}^{k}\big(p_{iy}(t)o(h)\big)-p_{ij}(t)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(p_{ij}(t)\big(1-\lambda(j)h\big)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)h\big)+o(h)-p_{ij}(t)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(p_{ij}(t)\big(1-\lambda(j)h-1\big)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)h\big)+o(h)\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(-p_{ij}(t)\lambda(j)h+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)h\big)+o(h)\right) \\
&= \lim_{h\to 0}\left(-p_{ij}(t)\lambda(j)+\sum_{y=1,y\neq j}^{k}\big(p_{iy}(t)\lambda(y,j)\big)+\frac{o(h)}{h}\right) \\
&= -p_{ij}(t)\lambda(j)+\sum_{y=1,y\neq j}^{k}p_{iy}(t)\lambda(y,j)
\end{aligned}
$$

$$p_{ij}(t)\lambda(j)=\sum_{y=1,y\neq j}^{k}p_{iy}(t)\lambda(y,j).$$

Using the fact that $\lambda(i)=\sum_{j=1,j\neq i}^{k}\lambda(i,j)$ the generator matrix of the Markov chain, A_{ij}, is defined as follows

$$A_{ij} = \begin{cases} -\lambda(i), & \text{if } i = j \\ \lambda(i,j), & \text{if } i \neq j \end{cases} = \begin{cases} -\sum_{j=1}^{k} \lambda(i,j), & \text{if } i = j \\ \lambda(i,j), & \text{if } i \neq j \end{cases}.$$

The Kolmogorov equations can then be rewritten in terms of the generator matrix:

$$\begin{aligned} -p_{ij}(t)\lambda(j) + \sum_{y=1, y\neq j}^{k} p_{iy}(t)\lambda(y,j) \\ &= p_{ij}(t)A_{jj} + \sum_{y=1, y\neq j}^{k} p_{iy}(t)A_{yj} \\ &= \sum_{y=1}^{k} p_{iy}(t)A_{yj} \\ &= (P(t)\mathrm{A})_{ij} \\ P'(t) &= P(t)A\,. \end{aligned}$$

Finally, the system of equations can be solved

$$P(t) = P(0)e^{tA} = e^{tA}$$

where e^{tA} is the matrix exponential and $P(0)$ is the identity matrix. The matrix exponential is defined by

$$e^{tA} = \sum_{v=0}^{\infty} \frac{t^v A^v}{v!} \text{ [59].} \tag{2.5}$$

For simple models $P(t)$ can be calculated in terms of A algebraically. In more complex cases, the Kolmogorov equations define a system of equations that cannot be solved analytically. If the eigenvalues of A are distinct, eigen-decomposition can be used to calculate $P(t)$ [68].

Let B be a diagonal matrix consisting of the eigenvalues and C be a matrix with corresponding eigenvectors as the columns. If distinct eigenvalues exist, C can be inverted and $A = CBC^{-1}$, and

$$e^{tA} = \sum_{\nu=0}^{\infty} \frac{t^{\nu}(CBC^{-1})^{\nu}}{\nu!} = C\left(\sum_{\nu=0}^{\infty} \frac{t^{\nu}B^{\nu}}{\nu!}\right)C^{-1} = Ce^{tB}C^{-1} \text{ [67].} \qquad (2.6)$$

The model parameters are estimated using maximum likelihood estimation with numerical optimization. Once the parameters are estimated, the likelihood can be calculated.

Suppose X is observed over $t_0 < t_1 < t_2 < ... < t_M$. Let $i_0, i_1, i_2, ..., i_M$ be the observed states over these time points. Then, the associated likelihood function is

$$\begin{aligned} L(\Lambda \mid x) &= P(\mathrm{X}_0 = i_0)\prod_{j=i}^{M} P(X_{t_j} = i_j \mid X_{t_{j-1}} = i_{j-1}) \\ &= P(\mathrm{X}_0 = i_0)\prod_{j=i}^{M} P(X_{t_j - t_{j-1}} = i_j \mid X_0 = i_{j-1}) \\ &= P(\mathrm{X}_0 = i_0)\prod_{j=i}^{M} P_{i_{j-1}i_j}(t_j - t_{j-1}) \,. \end{aligned} \qquad (2.7)$$

Using equations (2.5) in (2.7), the likelihood reduces to

$$L(\Lambda \mid x) = P_{i0}\prod_{j=1}^{M}\left(\sum_{k=0}^{\infty} \frac{(t_j - t_{j-1})^k (\Lambda^k)_{i_{j-1}i_j}}{k!}\right), \qquad (2.8)$$

where P_{i0} is the initial probability that the process is at i_0. For a series of observations $x_{z0},...,x_{zn_z}$ at times $t_{z0},...,t_{zn_z}$ for patients $z = 1,...N$, with covariate vectors p_i and model parameters θ, the log-likelihood is

$$l(\theta) = \sum_{z=1}^{N} \sum_{\ell}^{n_z} \log(p_{x_{i(\ell}} \quad _{\ell}(t_{\cdot\ell} \quad _{\ell}^{\cdot} \quad ^{\cdot}))) \text{ [59].} \tag{2.9}$$

Application of MSMMs requires the user to consider which transitions can realistically occur in continuous time. When the states represent levels of disease severity it is assumed that in order for a subject to travel from one state to a non-adjacent state, the subject also had to travel through the intermediate states [59]. For example, if a transition from state 3 to state 1 is observed, it is assumed that the subject traveled through state 2 at some point as well. Thus, in these applications, a reduced transition intensity matrix where non-adjacent state intensities are fixed to equal zero should be assumed, with the exception of transitions to death. If a state represents death it is called an absorbing state since transitions from death cannot occur. Figure 2.2 displays a general continuous MSMM for panel data where the states represent disease severity and state k is death. The same methods for parameter estimation from the general model apply for this reduced model.

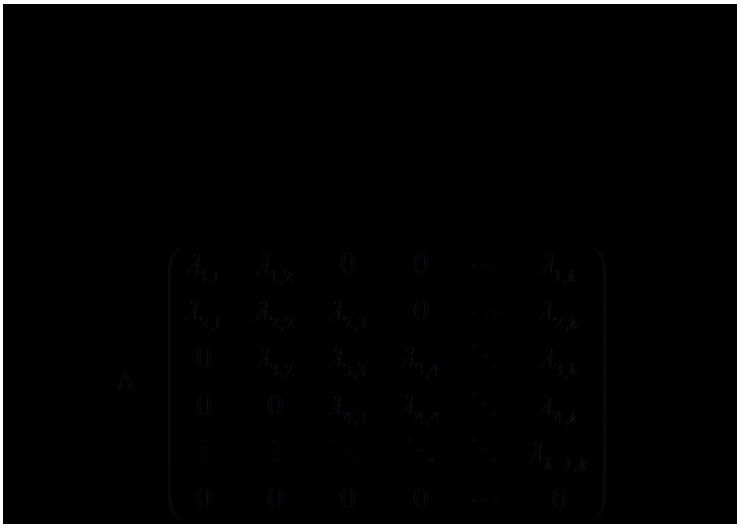


Figure 2.2: General MSMM for disease severity.

2.2.2 Piecewise-constant Multistate Markov Models

In the case of ischemic stroke occurrence and treatment, patients can get better or worse very quickly during the acute phase immediately after occurrence. For this reason, the time homogeneity assumption is expected to fail for the first transition, from the predicted baseline to the first observed outcome which typically occurs at one week or one month post stroke onset. Therefore, the assumption of homogeneity is relaxed and a non-homogeneous model is considered.

The MSMM for panel data can be extended to accommodate piecewise-constant intensity matrices for the non-homogeneous case [59]. Here, the transition probability functions are dependent on s, and the transition matrix function is $\mathbf{P}(s,t) = (p_{ij}(s,t))$. The transition intensity functions are now defined by

$$\lambda_{ij}(t) = \lim_{\delta t \to 0} \frac{P(X(\mathrm{t}+\delta t) = j \mid X(\mathrm{t}) = i)}{\delta t},\ i, j \in S,\ i \neq j\,. \tag{2.10}$$

The time-homogeneity assumption can be tested by using a likelihood ratio test to compare the time-homogenous model to piecewise-constant models with different cutoff times.

2.2.3 Predictors

The effect of a predictor, specifically treatment, is incorporated into the model as transition intensity functions [69]. Let z be a vector of observed predictors then

$$\lambda_{ij}(t;z) = \lambda_{ij}(t)\exp(z'\beta_{ij}(t)),\ \ \lambda_{ij}(t) > 0 \tag{2.11}$$

where $\beta_{ij}(t)$ is the parameter vector associated with the predictor vector z in the transition between states i and j in time t. The transitional rates are represented by $\lambda_{ij}(t;z)$ at time t for the patients with vector z.

2.2.4 Application of MSMMs to the mRS

mRS data collected from acute stroke trials is used to demonstrate the aforementioned MSMM methodology. Although MSMMs have been used to describe a number of disease processes [60-63, 70], currently literature is lacking applications of these models on data with a larger, i.e. more than four, number of states. With disease represented as seven states the mRS is a good example of where the application can be expanded. Also the mRS is collected at discrete time points, the disease process itself is not discrete, thus it is an example of panel data. Two acute stroke clinical trials are used in this application,

the NINDS t-PA trial and the ALIAS trial. In the NINDS t-PA trial, the observations occurred at 7-10 days, 90 days, 180 days, and 360 days [38]. The mRS was collected at 30 days, 90 days, 180 days, 270 days, and 360 days for ALIAS [43]. Because the observations are not evenly spaced a discrete model is not appropriate, and a continuous-time MSMM is used.

The states of this model represent the seven levels of the ordinal mRS scale. The nature of the mRS and the follow-up schedule lead to a sparsely populated matrix of observed transitions. Tables 2.2 and 2.3 show the state tables of the frequency of transitions for NINDS t-PA and ALIAS, respectively. The rows represent the state to which a subject begins, and the columns represent the state into which the subject transitions. These tables include all transitions over the course of follow up. For example, for the NINDS t-PA trial, there were 196 observations where the mRS for a subject was 0 for two consecutive time points (including from 7-10 days to 90 days, 90 days to 180 days and 180 days to 360 days).

As previously described, a large proportion of the observed pairs of the mRS are for subjects that do not transition and remain in the same state. The transitions that do occur are largely adjacent state transitions. Though non-adjacent state transitions are observed, it is assumed that a subject passed through the intermediate states; the transitions were not captured because the mRS was observed at discrete times. This data structure is unlike data in published applications of MSMMs. First, most other models have only three or four states; here, there are seven. Second, the frequencies for many of

the transitions are sparse. The feasibility of MSMMs for this type of data needs to be determined.

Table 2.2. Frequency of mRS Transitions in NINDS t-PA.

	0	1	2	3	4	5	6
0	196	46	4	0	1	0	1
1	62	224	23	4	4	1	5
2	6	53	73	20	4	0	7
3	6	21	48	137	9	3	6
4	3	18	23	78	156	21	28
5	0	2	2	14	53	79	75

Table 2.3. Frequency of mRS Transitions in ALIAS.

	0	1	2	3	4	5	6
0	524	98	13	6	3	1	2
1	158	542	85	19	4	0	2
2	31	175	432	53	7	1	3
3	7	48	150	417	43	4	5
4	1	7	28	136	391	37	9
5	0	1	0	7	60	100	23

In order to determine the feasibility of MSMMs for analysis of sparsely populated ordinal data, the operating characteristics of MSMMs are compared with repeated logistic regression in Chapter 3 of this dissertation. Once the feasibility is assessed, a MSMM approach with piecewise-constant transition intensities incorporating a latent baseline state is developed in Chapter 4. Finally, in Chapter 5, the methods developed in Chapter 4 are applied to acute stroke therapy trial data and are compared with results from the alternative methods previously described.

3 Original Manuscript 1

Title: Assessing type I error and power of multistate Markov models for panel data- A simulation study

Authors: Christy Cassarly[1], Renee' Martin[1], Marc Chimowitz[2], Edsel Peña[3], Viswanathan Ramakrishnan[1], Yuko Palesch[1]

Affiliation:

[1] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC

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

[3] Department of Statistics, University of South Carolina, Columbia, SC

Submission Status: published in *Communications in Statistics - Simulation and Computation*

Abstract

Ordinal outcomes collected at multiple follow-up visits are common in clinical trials. Sometimes, one visit is chosen for the primary analysis and the scale is dichotomized amounting to loss of information. Multistate Markov models describe how a process moves between states over time. Here, simulation studies are performed to investigate the type I error and power characteristics of multistate Markov models for panel data with limited non-adjacent state transitions. The results suggest that the multistate Markov models preserve the type I error and adequate power is achieved with modest sample sizes for panel data with limited non-adjacent state transitions.

Keywords

Multistate models; panel data; type I error; power; stroke

3.1 Introduction

Most randomized trials in acute stroke neuroprotection treatment have failed to show efficacy for new interventions [3]. Mergenthaler and Meisel (2012) provide several explanations to describe the lack of positive trials in stroke including heterogeneity in stroke pathophysiology and incomplete preclinical testing [6]. Two of the explanations cited by the Optimising Analysis of Stroke Trials Collaboration are inadequate study designs and inappropriate statistical methods, specifically the analysis of the primary outcome [1].

The modified Rankin Scale (mRS) score at 90 days post-randomization is a commonly used primary outcome measure in Phase III clinical trials of acute stroke therapy [2]. The mRS is a seven-point ordinal scale that measures degree of disability of stroke patients (Table 3.1).

Table 3.1. Modified Rankin Scale.

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities but able to look after own affairs without assistance
3	Moderate disability requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent, and requiring constant nursing care and attention
6	Dead

Analyzing the mRS as an ordinal scale has only recently gained acceptance [1, 3, 46]. Many trials have chosen to dichotomize the mRS into success, scores of 0 or 1 (or 0

to 2), or failure, scores greater than 1 (or 2), for the primary analysis [7]. Though models used for dichotomous outcomes are easier to implement and some prefer the clinically meaningful interpretations, dichotomization can result in a loss of statistical power [1]. It is intuitive that some patients with severe stroke may never have the potential to achieve a success as defined by the dichotomy. Thus, the prognostic heterogeneity of subjects does not allow for potential equal contribution to the estimation of the treatment effect for all subjects [9].

Recently, an emphasis has been placed on exploring alternate analytic methods for the mRS outcome data from acute stroke trials. Results indicate that the mRS should be analyzed in such a way that maintains the original structure of the scale as much as possible [1, 3]. Linear regression and analysis of variance have been suggested to analyze the mRS scale as a continuous variable. Although results from these models are generally intuitive, the application to the mRS leads to summary statistics that will not have a clear interpretation. Non-integer values from an ordinal scale do not have a clear meaning when they are treated as continuous.

Another popular alternative method for mRS outcome data is sliding dichotomy analysis. The sliding dichotomy method allows for the definition of success to vary based on patient-specific baseline prognostic variables while maintaining a dichotomized outcome [48]. Commonly, re-analysis of acute stroke trial data using the sliding dichotomy defines pre-specified cut-points for prognostic group inclusion based on the National Institutes of Health Stroke Scale (NIHSS) score [49]. The mRS is unavailable immediately after randomization so models of acute stroke trial data often adjust for

baseline severity using the NIHSS, a score that ranges from 0 (no neurological deficit) to 42. Often, three prognostic groups are defined using the baseline NIHSS for the sliding dichotomy as 'mild', 'moderate', and 'severe' and the definition for 'success' differs for each group. One example is to define favorable outcome as mRS = 0 for mild strokes, mRS = 0-1 for moderate strokes and mRS = 0-2 for severe strokes. Since baseline severity is a strong predictor of outcome in stroke patients, this baseline severity adjusted approach has been considered for use over the traditional dichotomy.

While the sliding dichotomy has the potential to be a powerful tool in some settings, it has limitations. Some simulation studies have shown that the utilization of the sliding dichotomy provides higher sensitivity to detect true treatment effects [50]. For example, when the probability of favorable outcome is high (greater than 0.5), the sliding dichotomy provides higher power [9]. This is not a general result; however, as other studies have shown that the traditional dichotomy is more powerful than the sliding dichotomy in many situations [51]. When the probability of favorable outcome is lower, the traditional dichotomy is more powerful [9]. In addition, determining the number of prognostic groups to use is not an obvious decision and can be difficult to justify. Moreover, determining how to choose the cut points for the different groups can be a difficult task. Although the use of three groups (mild, moderate and severe) is common in the literature, methods used to determine severity cut points vary and need to be verified [48]. Poor selection of the number of prognostic groups and cut points could result in a loss of power. Furthermore, while the sliding dichotomy allows for a baseline severity adjusted outcome, it still ignores any non-'success' transition from one mRS score to

another even though each mRS category (except 5 to 6) represents a clinically meaningful difference in health state [52]. The recovery and outcome of subjects following a stroke realistically lies on a continuum. Ordinal analysis of the mRS scores can provide a more complete understanding of this process than analysis of the dichotomized scale [10].

Recently, methods using the full ordinal scale have been demonstrated [3, 46, 71]. The proportional odds model is a cumulative logistic regression model that has been proposed for analysis of mRS outcome data, Use of this model requires the assumption of proportional odds- the odds ratio comparing treatment to control in subjects with mRS = 0 versus 1-6 is the same as the odds ratio for mRS = 0-1 versus 2-6, and so on. In data where the proportional odds assumption does not hold, shift analysis, an assumption-free ordinal test, can be used [7]. Shift analysis can be performed using the van Elteren test, an extension of the two-sample Wilcoxon rank-sum test. Though shift analysis does not require assumptions, it does not produce a summary statistic which is often desired by clinicians. Alternatively, in cases where the proportional odds assumption is unreasonable, the partial proportional odds model or adjacent categories logit model can be used. The partial proportional odds model includes an additional term to allow for the odds ratios to increase proportional to the outcome scale [46]. The adjacent categories logit model calculates odds ratios for each adjacent category of response in relation to covariates. Both of these models are more flexible than the proportional odds model but lack a single summary statistic.

An additional drawback of focusing the primary outcome on the 90-day time point is the lack of use of available longitudinal data. Many acute stroke trials collect the mRS at discharge and/or at 30 days from randomization and also at periodic intervals through 12 months, if long-term follow-up is planned. The longitudinal data are not often used in the primary analysis. Repeated measures analysis, which incorporates outcome data from all follow-up visits, may provide a more comprehensive clinical understanding of the treatment effect on outcome after a stroke [11].

Table 3.2. mRS transition example.

		mRS at time 2							
		0	1	2	3	4	5	6	Total
mRS at time 1	0	84	20	5	1	0	0	0	110
	1	45	79	17	2	1	0	1	145
	2	12	50	56	15	8	6	1	148
	3	2	25	46	63	17	3	0	156
	4	1	2	23	77	89	23	5	220
	5	0	0	2	7	41	48	22	120
	6	0	0	0	0	0	0	101	101
	Total	144	176	149	165	156	80	130	1000

		mRS at time 3							
		0	1	2	3	4	5	6	Total
mRS at time 2	0	108	26	5	3	1	1	0	144
	1	38	123	12	2	0	1	0	176
	2	9	31	86	18	2	1	2	149
	3	3	5	32	116	8	1	0	165
	4	0	1	4	34	102	13	2	156
	5	0	0	0	2	18	45	15	80
	6	0	0	0	0	0	0	130	130
	Total	158	186	139	175	131	62	149	1000

In this article, a novel approach using multistate Markov modeling is proposed for the mRS scores. Multistate Markov modeling incorporates the longitudinal ordinal data

and provides clinically relevant summary statistics to describe treatment effect. The mRS has more disease 'states' (here, the seven levels of the ordinal response) than many previously considered clinical applications of multistate Markov models. An example of the typical data structure of the observed mRS could be illustrated in Table 3.2. In this example, mRS outcome data from a mock acute stroke trial of 1,000 subjects was created for three follow-up visits. The 'transition' from one state to another that occurred from one visit to the next is described in Table 3.2. For example, 79 of the 145 subjects that had mRS = 1 at time 1 also had mRS = 1 at time 2. Only 17 of the 145 subjects with mRS = 1 at time 1 transitioned to mRS = 2 at time 2. This is an example of an 'adjacent-state' transition. Throughout the table, a majority of the observations are instances where the subjects remained in the same state, or had the same mRS score from one time to the next. Most of the subjects that transitioned to a different state display adjacent-state transitions, with a limited number of non-adjacent state transitions.

A literature review conducted of an online database yielded a total of 40 articles using the following keywords: multistate, Markov, panel, clinical, application, continuous-time, and the following excluded words: piecewise, non homogeneous, nonhomogeneous, inhomogeneous, semi Markov, hidden, random effects. An article was excluded if (a) the content was actually theoretical and there was no application, (b) it was a review with no new content, (c) multistate models were referenced, flagging it for review but the models were not actually fit, or (d) the models were actually discrete-time. Of the remaining 26 articles, 25 fit models to data with five or fewer states and two fit models to data with six states [12-36]. One publication used a six-state model to analyze

a dataset with much more data than is typically collected in acute stroke trials- approximately 5,000 patients [37].

The multistate Markov model is introduced in Section 3.2. The main focus of the paper is to approximate the power and type I error probabilities for multistate Markov models of data structures similar to the longitudinal mRS outcomes observed in acute stroke trials. In Section 3.2, continuous-time multistate Markov models are defined and the simulation scenarios for estimation of the operating characteristics of these models are described. In Section 3.3, the type I error probabilities and power are approximated for varying design elements and power of the multistate models is compared with that of repeated measures logistic regression. In Section 3.4, the findings are summarized and discussed.

3.2 Methods

Multistate Markov modeling is an alternative approach to analyze repeated measures data with an ordinal outcome. The multistate Markov model describes how a process moves between states over time, which is desirable in the description of disease processes that naturally move through increasing stages of severity [4]. Subjects can improve and worsen over the course of follow-up and these movements back and forth between disease states are all incorporated in the estimation of the model. Multistate Markov models can provide a better clinical understanding of the disease process since the information from the entire course of the disease is used to estimate the parameters of the model. These models have been used in numerous clinical applications including: multiple sclerosis [60, 70], periodontal disease [61], alcoholism [62], and psychiatry [63].

This approach has not been used before for mRS data and therefore, in this article a simulation study is performed to examine the operating characteristics of the proposed approach.

3.2.1 Multistate Markov Models

The use of multistate Markov modeling requires that the Markov property holds for the observed data. Consider a stochastic process with a finite state-space $S = \{1, 2, 3, ...I\}$, where I represents the number of states in the model. Let $X(s)$ be the state occupied at time s. The series of observations has the Markov property if the conditional distribution of $X(s+t)$, given $\mathrm{F}_{X(s)} = \{X(v) : v \le s\}$, where $\mathrm{F}_{X(s)}$ denotes all of the information pertaining to the history of X up to time s [64], satisfies

$$\Pr\left\{X(s+t) = j \mid \mathrm{F}_{X(s)}, X(s) = i\right\} = \Pr\left\{X(s+t) = j \mid X(s) = i\right\} = p_{ij}(s,t),\ i, j \in I\,. \qquad (3.1)$$

In other words, a Markov process is one such that the conditional probability distribution of the state of a process at a given time is dependent only on the immediately preceding observation and not on the earlier ones.

Markov models may be defined for discrete time as well as continuous. Although the course of disease is a continuous process, clinical trials often only collect data at intermittent follow-up visits. In the context of stroke, the exact time of progression or recovery, or change of state, of disease is not observed. Data of this type, representative of a continuous process that is only observed at discrete time points, is known as panel data [72]. Both discrete and continuous time multistate Markov models can be used to describe panel data. If the sampling times are equally spaced, a continuous model that has

been adapted for panel data is preferred over a discrete model [68]. In many acute stroke trials, the mRS is collected at follow-up visits that are not evenly spaced. In such instances, continuous time models are appropriate. A continuous model for panel data can only be used in cases where the sampling times are considered to be non-informative [4]. An example of non-informative sampling is a fixed observation scheme, where the interval of follow-up is specified in advance. However if observations are not fixed or random and are self-selected by the subject (informative), this modeling technique is not appropriate without properly adjusting for the additional information [4]. For instance, these models cannot be used in a scenario where observations occur when a subject visits a doctor because they are in poor-condition. A model that incorporates the information from the sampling times must be used for this type of self-selected follow-up outcome data. In acute stroke trials, the follow-up visits are usually specified in advance and are non-informative so continuous modeling is appropriate.

A common assumption when fitting continuous-time Markov models is the time-homogeneity assumption. This is the assumption that the transition probabilities remain constant over time. When time-homogeneity is assumed, the probability that the next move of the process is from state i to state j can be written,

$$P\{X(s+t)=j \mid X(s)=i\} = P\{X(t)=j \mid X(0)\} = p_{ij}(t). \qquad (3.2)$$

Thus, the probabilities only depend on the length of the time interval, t. The $p_{ij}(t)$ are elements of the transition probability matrix, $\mathbf{P(t)}$. The $(i,j)^{\text{th}}$ entry of $\mathbf{P(t)}$ is the probability of being in state j given the starting state is i after a time interval of t.

The movement of a subject between states is described by λ_{ij}, the transition intensities:

$$\lambda_{ij} = \lim_{\delta t \to 0} \frac{P(X(\delta t) = j \mid X(0) = i)}{\delta t}. \tag{3.3}$$

The intensities represent the instantaneous rate of moving from state i to state $j \neq i$. The intensities form the generator matrix, $\mathbf{\Lambda}$, whose rows sum to zero and the diagonal entries are $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$. $\mathbf{P(t)}$ can be solved by taking a matrix exponential of $\mathbf{\Lambda}$ scaled by the time interval,

$$\mathbf{P(t)} = e^{\mathbf{t\Lambda}} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{\Lambda}^k}{k!} \tag{3.4}$$

where $\mathbf{\Lambda}^k$ is the k$^{\text{th}}$ power of the generator matrix $\mathbf{\Lambda}$.

Suppose now that we observe X over $t_1 < t_2 < ... < t_M$. Let $i_1, i_2, ..., i_M$ be the observed states over these time points. Then, the associated likelihood function is

$$P(X_1 = i_1) \prod_{j=i}^{M} P(X_{t_j} = i_j \mid X_{t_{j-1}} = i_{j-1})$$

$$= P(X_1 = i_1) \prod_{j=i}^{M} P(X_{t_j - t_{j-1}} = i_j \mid X_0 = i_{j-1})$$

$$= P(X_1 = i_1) \prod_{j=i}^{M} P_{i_{j-1} i_j}(t_j - t_{j-1}). \tag{3.5}$$

Using (3.4) and (3.5), the likelihood is therefore

$$L(\Lambda \mid x) = P_{i_1} \prod_{j=1}^{M} \left(\sum_{k=0}^{\infty} \frac{(t_j - t_{j-1})^k (\Lambda^k)_{i_{j-1} i_j}}{k!} \right) \tag{3.6}$$

where P_{i_1} is the initial probability that the process is at i_1.

The effects of covariates can also be investigated by modeling the intensity as a function of the variables of interest, $\mathbf{z(t)}$. The transition intensity matrix elements λ_{ij} are replaced by

$$\lambda_{ij}\left(\mathbf{z}\left(\mathbf{t}\right)\right) = \lambda_{ij}^{(0)} e^{\beta_{ij}' \mathbf{z(t)}} \tag{3.7}$$

where $\lambda_{ij}^{(0)}$ represent the baseline intensities (without covariates) and β_{ij} are the effect of covariates on the transition from state i to state j [4]. To determine the significance of a covariate, a likelihood ratio test is used to compare nested models. In Section 3.3, the model including treatment is compared to a model without treatment. Thus, the resulting intensities are

$$\lambda_{ij} = \lambda_{ij}^{(0)} e^{\beta_{ij} z(t)} \tag{3.8}$$

where $\lambda_{ij}^{(0)}$ represent the intensities without the covariate and $z(t)$ is the treatment assignment (0 for control and 1 for treatment) for subject n. Thus, the null and alternative hypotheses for the test of the effect of treatment are

$$H_0 : \beta_{ij} = 0 \text{ for all } i, j$$
$$H_1 : \beta_{ij} \neq 0 \text{ for some } i, j.$$

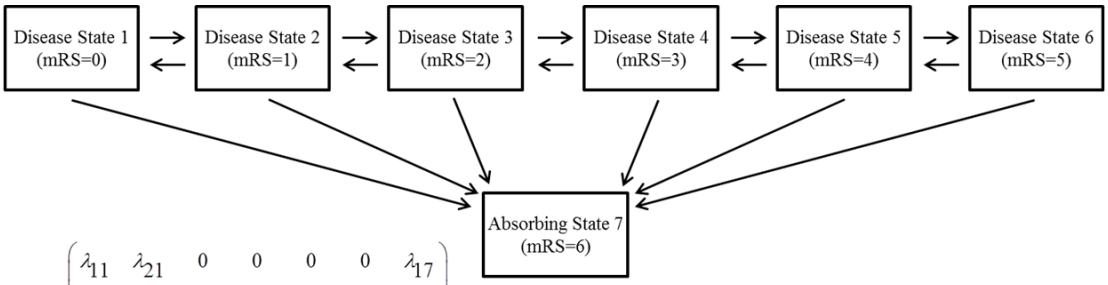
The null hypothesis will be rejected using the asymptotic distribution of $-2\ln(L_0 / L_1)$ where L_0 is the maximum value of the likelihood of the reduced model and L_1 is the

maximum value of the likelihood of the full model. For large n, this asymptotic distribution is a χ^2 with k degrees of freedom, where k is the difference in the number of parameters in the two models.

The difficult part in this process is obtaining the maximum likelihood estimates. Often methods such as Newton-Raphson can cause issues because the computation of the second derivative can be costly in terms of time. Additionally, if the Hessian matrix is non-negative definite away from the optimum, slow or non-convergence may occur. To avoid this, other approaches have been proposed. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is used to maximize the likelihood with analytic gradients and can be used with or without analytic first derivatives [4, 68]. The BFGS algorithm approximates Newton's iterative method for finding the roots of differentiable functions [73]. In this algorithm, the Hessian matrix of second derivatives is not evaluated directly. Instead, it approximates the Hessian using gradients. If too many transitions are considered with not enough data in a multistate model, the maximum likelihood estimate could lie on boundary of the parameter space (when one or more transition intensities equal 0). If this occurs, the maximum likelihood estimate may be inconsistent since asymptotic theory requires the assumption that the true parameter value lies away from the boundary.

It is important to consider which transitions can realistically occur in continuous time. When the states represent levels of disease severity it is assumed that in order for a subject to travel from one state to a non-adjacent state, the subject also had to travel through the intermediate states. Thus, in this application of these models, a reduced

transition intensity matrix where non-adjacent state transition intensities are fixed to equal zero should be assumed. The exception is with mRS = 6, we assume that death can occur from any state and transitions cannot occur out of it because it is an absorbing state. The allowable transitions are displayed in Figure 3.1.



$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{21} & 0 & 0 & 0 & 0 & \lambda_{17} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & 0 & 0 & 0 & \lambda_{27} \\ 0 & \lambda_{32} & \lambda_{33} & \lambda_{34} & 0 & 0 & \lambda_{37} \\ 0 & 0 & \lambda_{43} & \lambda_{44} & \lambda_{45} & 0 & \lambda_{47} \\ 0 & 0 & 0 & \lambda_{54} & \lambda_{55} & \lambda_{56} & \lambda_{57} \\ 0 & 0 & 0 & 0 & \lambda_{65} & \lambda_{66} & \lambda_{67} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 3.1: Multistate model for panel observed mRS data.

3.2.2 Simulation Scenarios

In this section, the procedures for examining the operating characteristics of multistate Markov models under a variety of conditions are described. First, whether or not the multistate Markov model preserves the type I error probability is examined through simulations. Next, given the type I error probability, the desired power is examined for two clinically relevant scenarios, each with two sets of follow-up trajectories for each of the models. The power of the multistate Markov model is compared with that of repeated logistic regression. The motivating example of this simulation study is the limited non-adjacent state transitions observed in mRS data. The simulation scenarios are generated such that the assigned transition probabilities mimic real acute stroke trial as closely as

possible. Data from three different phase III acute stroke trials were considered when assigning transition probabilities.

The first trial used is the National Institute of Neurologic Disorders and Stroke (NINDS) tissue-Plasminogen Activator (t-PA) study [38]. The NINDS t-PA trial compared t-PA versus placebo in subjects with acute ischemic stroke. The primary analysis showed a significant global test score for four (Barthel Index, mRS, Glasgow Outcome Scale, and NIHSS) outcomes as well as for the mRS alone [39]. To further illustrate the structure of acute stroke trial data, the mRS scores for the control and treatment groups are displayed in Sankey plots in Figures 3.2 and 3.3 [74], respectively. The Sankey plots allow for a visualization of changes within each treatment group over time. The longitudinal bar chart shows the percentage of subjects with each mRS score at each follow-up visit. In addition, the wavy lines between each bar, the links, describe the change in the number of subjects in each state, over time. A thick line indicates that a large number of subjects transition between two states. Note that as illustrated in Table 3.2 with the mock data, the percentage of transitions that occur between non-adjacent states is small.

The other two trials considered for data generation were the albumin in acute stroke (ALIAS) II trial and the Interventional Management of Stroke (IMS) III trial [43, 75]. ALIAS II was designed to compare 25% human serum albumin and saline in patients with acute ischemic stroke. IMS III was designed to compare intravenous t-PA plus an intra-arterial device therapy and/or additional intra-arterial t-PA versus t-PA alone. Both ALIAS II and IMS III were stopped early for futility.

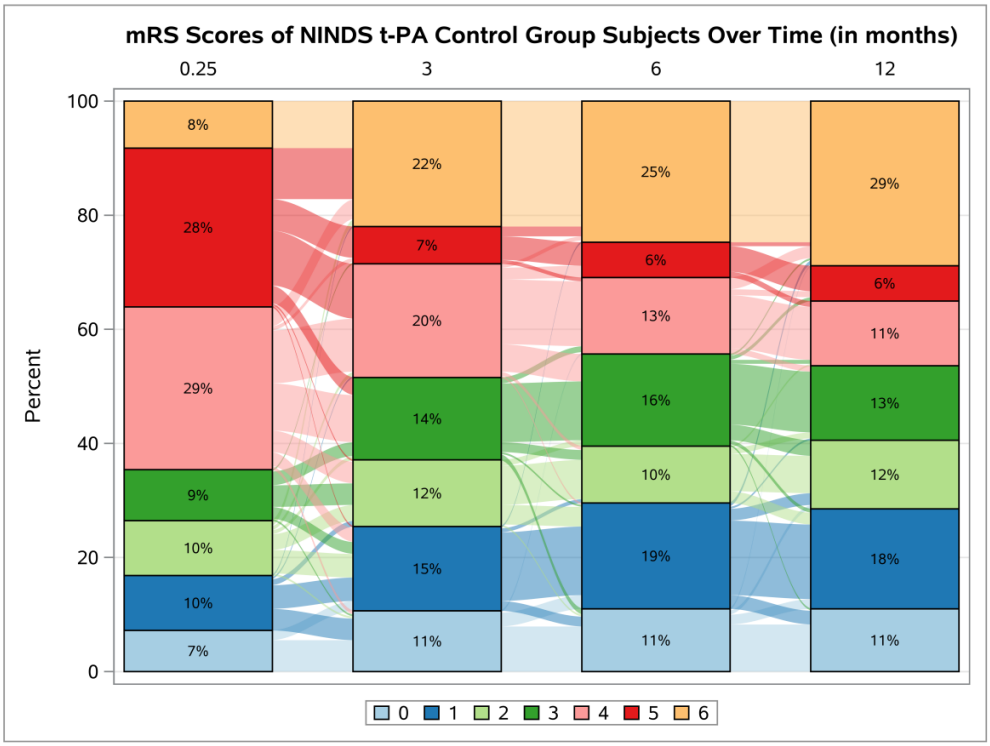


Figure 3.2: Sankey plot of NINDS t-pa control group subjects over time.

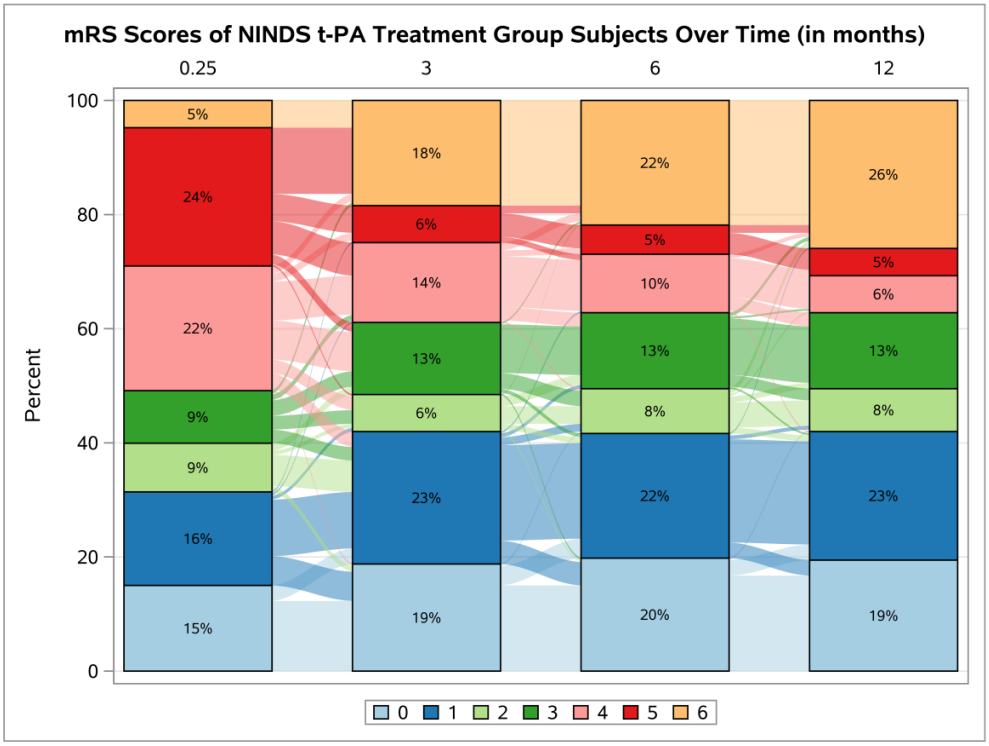


Figure 3.3: Sankey plot of NINDS t-pa treatment group subjects over time.

For each of the previously mentioned trials, the observed transition counts for each follow-up visit are combined in one table to calculate aggregate observed transition probabilities. To illustrate, data in Table 3.2 would be combined in an aggregate table with 192 (84 + 108) instances where a subject stayed in state 0 out of the total 254 (110 + 144) instances where a subject started in state 0. Thus, for example, the observed aggregate transition probability of remaining in state 0 is 0.76 (192/254), in the mock trial. These observed transition probability matrices are calculated for each study to illustrate the structure for mRS outcome data from acute stroke trials. As previously mentioned, the notable characteristic of the mRS outcome data from these trials is the limited number of non-adjacent state transitions.

To evaluate power, data are generated under the alternative hypothesis that a treatment effect exists. In each multistate Markov model, multiple parameters describe a single covariate effect. Therefore, there are many ways in which a significant treatment effect could exist. In order to simplify, we consider two different clinically relevant scenarios. The first scenario considers a case where only one of the assigned transition probabilities differs between the control and treatment groups. For this set of simulations, the transition probabilities are assigned such that they are all the same for both groups except for the transition from mRS = 3 to mRS = 2 (as well as the transition from mRS = 2 to 1, mRS = 2 to 0, and mRS = 1 to 0, as the intermediate transitions may not be observed). The second scenario for sample size estimation is one where the treatment effect exists in all transitions. The positive transitions are assigned higher probabilities in

the treatment group, making them more likely. The negative transitions are assigned a larger probability in the control group.

It is likely that other ordinal scales collected over time have a longitudinal structure similar to the mRS, where non-adjacent state transitions are sparse. In order to consider scales with differing numbers of states, we used the data generated to mimic the mRS described above and collapsed the estimated transition probability matrices to create six-, five-, and four-state models. The method in which the states are aggregated is described in Table 3.3.

Table 3.3. Simulation scenarios for power.

		Number of subjects per group	
States	Visits	One differing transition	All differing transitions
7	3	400, 500, 600, 800, 1000	300, 400, 500, 600, 800, 1000
	6	200, 300, 400, 500, 600	125, 150, 175, 200, 300, 400
6	3	400, 500, 600, 800, 1000	300, 400, 500, 600, 800, 1000
	6	150, 200, 300, 400, 500, 600	125, 150, 175, 200, 300, 400
5	3	300, 400, 500, 600, 800, 1000	200, 300, 400, 500, 600, 800, 1000
	6	100, 150, 200, 300, 400, 500, 600	75, 100, 125, 150, 175, 200, 300, 400
4	3	200, 300, 400, 500, 600, 800, 1000	200, 300, 400, 500, 600, 800, 1000
	6	100, 150, 200, 300, 400, 500, 600	50, 75, 100, 125, 150, 175, 200, 300, 400
3	3	200, 300, 400, 500, 600, 800, 1000	200, 300, 400, 500, 600, 800, 1000
	6	100, 150, 200, 300, 400, 500, 600	50, 75, 100, 125, 150, 175, 200, 300, 400

In practice, collapsing states is a decision that should be made with caution. For example, if there is clinical evidence that two health states are not distinct, it may be acceptable to combine them. If two health states are aggregated that are vastly different there could be a loss of power. In order to illustrate this point, for the 5-state (and subsequently the 4-state) model, mRS = 2 and mRS = 3 are aggregated. In the scenario

where only the transition from mRS = 3 to mRS = 2 differs, there is an expected loss in power for these aggregated models. Thus, in the case where only one transition differs, an additional scenario was considered where mRS = 2 and mRS = 3 are not combined, referred to as the 5-state* model.

The probabilities used to assign outcome trajectories are listed in Appendices 3A-3C. Using these probabilities, the data generation includes the following steps:

1. Generate a sample of treatment assignments from a random uniform(0, 1) distribution where the probability that the m^{th} subject is assigned to treatment is 0.5.
2. Generate random uniform variables for all t.
3. Assign a state for $t = 0$ using the probabilities described in the appendices.
4. For each $t > 0$ use the probabilities to assign a state conditional on the state occupied at $t - 1$.

To determine the type I error, data are generated under the null hypothesis of no treatment effect. The simulation scenarios for estimation of type I error include differing number of states and increasing sample size per group, starting at 200.

The simulation studies for power are repeated for each set of simulation parameters (Table 3.4) allowing the number of subjects in each treatment group to vary, as well as the number of follow-up visits (three or six visits). Each set of simulations is carried out using 1,000 runs. For each set of parameters the sample size is set to observe approximately 80% power. The type I error for the multistate Markov model is set to the

observed value from the previously described simulations. The resulting power is compared to that of the repeated logistic regression model.

Table 3.4. Modified Rankin Scale inclusion categories.		
Symbol	Model	mRS Scores
○	7 state	mRS 0, mRS 1, mRS 2, mRS 3, mRS4, mRS 5, mRS 6
□	6 state	mRS 0, mRS 1, mRS 2, mRS 3, mRS4, mRS 5-6
◇	5 state	mRS 0, mRS 1, mRS 2, mRS 3, mRS 4-6
△	4 state	mRS 0, mRS 1, mRS 2-3, mRS 4-6
☆	3 state	mRS 0-1, mRS 2-3, mRS 4-6

The data used were simulated using SAS 9.4 statistical software. SAS 9.4 was also used to run the Generalized Estimating Equation models for repeated measures logistic regression with PROC GENMOD. The Markov models were fitted in R statistical software version 3.3.0 using the 'msm' package for multistate Markov models [4].

3.3 Results

In this section, the behavior of the type I error and power is evaluated. The simulation results of the type I error are displayed in Figure 3.4. For the application considered, with data structured similar to the three acute stroke trials described in Section 3.2, the type I error probability is preserved for all of the multistate Markov models. In order to examine whether the chi-square approximation of the likelihood ratio test is appropriate for comparing the nested models, p-values under the appropriate chi-square distribution were obtained and are shown in Appendix 3D. The p-values appear to be approximately uniform and the test-statistic sampling distribution approximates the chi-square distributions quite adequately.

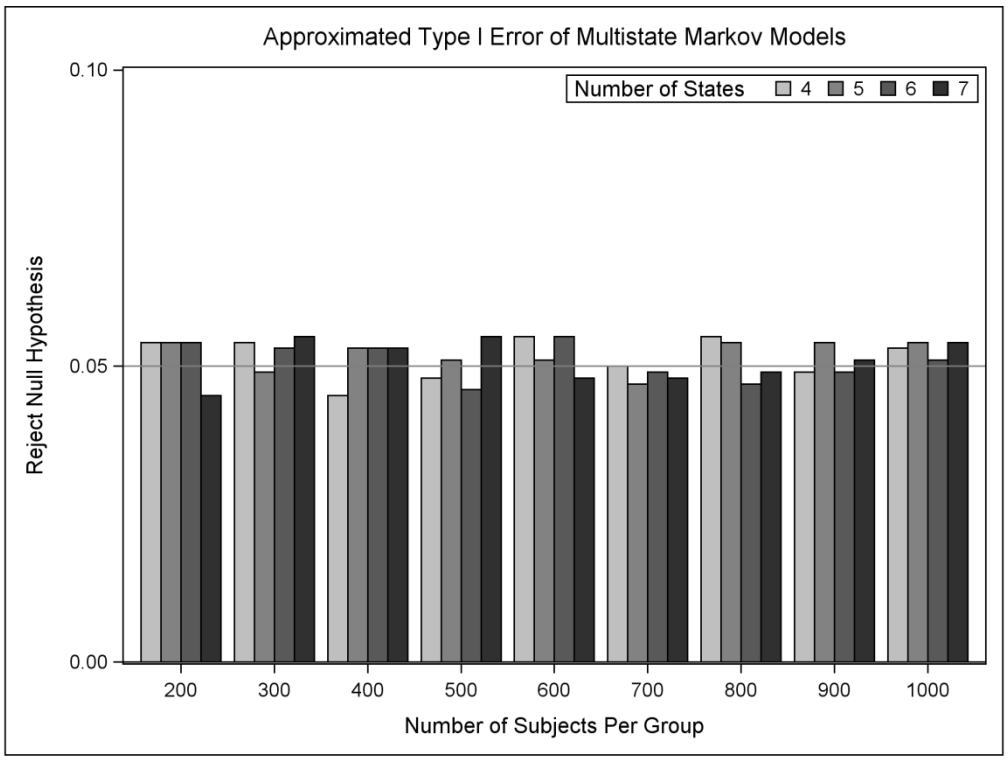


Figure 3.4: Graph of approximated type I error probabilities for models with three follow-up visits based on 1,000 simulations.

For considering power we need to set alternative hypotheses. There are many potential alternative hypotheses so we consider two scenarios that are clinically relevant. In the first scenario, transition probabilities are assigned such that the only difference between treatment groups is in the transition from mRS = 3 to mRS = 2. The results with three and six follow-up visits are displayed in Figures 3.5a and 3.5b, respectively. The transition probabilities assigned for these simulations are presented in Appendix 3B. The results indicate that for a seven-state model with three follow-up visits, approximately 500 subjects are needed in each group to obtain 80% power. There is a marginal increase in power when states mRS = 4 and mRS = 5 are combined in the six-state model. When mRS = 2 and mRS = 3 are combined for the original five-state model we see an extreme

decrease in power. This is expected because the model was misspecified. The only difference between treatment groups was in the transition from mRS = 3 to mRS = 2 so when these two states are combined, there are virtually no differences to detect. The same phenomenon is observed in the four-state model because the difference is still lost from aggregating mRS = 2 and mRS = 3. If we consider the fact that the difference lies between those two states and instead collapse mRS = 0 and mRS = 1 in the alternative five-state model (5*) then we see another marginal increase in power. The observed increases in power are expected because there are no differences in the two groups in the aggregated states and there are fewer parameters to estimate in the model.

Figure 3.5b displays the approximated power in the scenario where only the transition from mRS = 3 to mRS = 2 differs, now with six follow-up visits instead of three. The results for the models with six follow-up visits are similar to those in the models with three follow-up visits, except that the power is significantly increased. The power for the seven-, six- and five-state* model are all very similar. Each of these models requires approximately 150 subjects in each group to obtain 80% power. When mRS = 2 and mRS = 3 are combined in the five-state model (and subsequently in the four-state model), there is an extreme loss of power, as previously observed.

The results of the power simulations in the second scenario, where the treatment effect exists for all transitions, are displayed in Figures 3.5c and 3.5d. The assumed transition probabilities are described in Appendix 3C. The approximate power for the three follow-up visit case is displayed in Figure 3.5c. In the six- and seven- state model, the iteration to obtain the estimates do not converge for sample sizes as small as 200.

There are negligible differences in power between each of the models. Since there are differences in all transitions, there will be some loss of power by aggregating states. However, there is a gain in power with fewer parameters in a reduced model. These two facts lead to very minimal change in power. For any given model with three follow-up visits in this scenario, approximately 600 subjects are needed per group to attain 80% power.

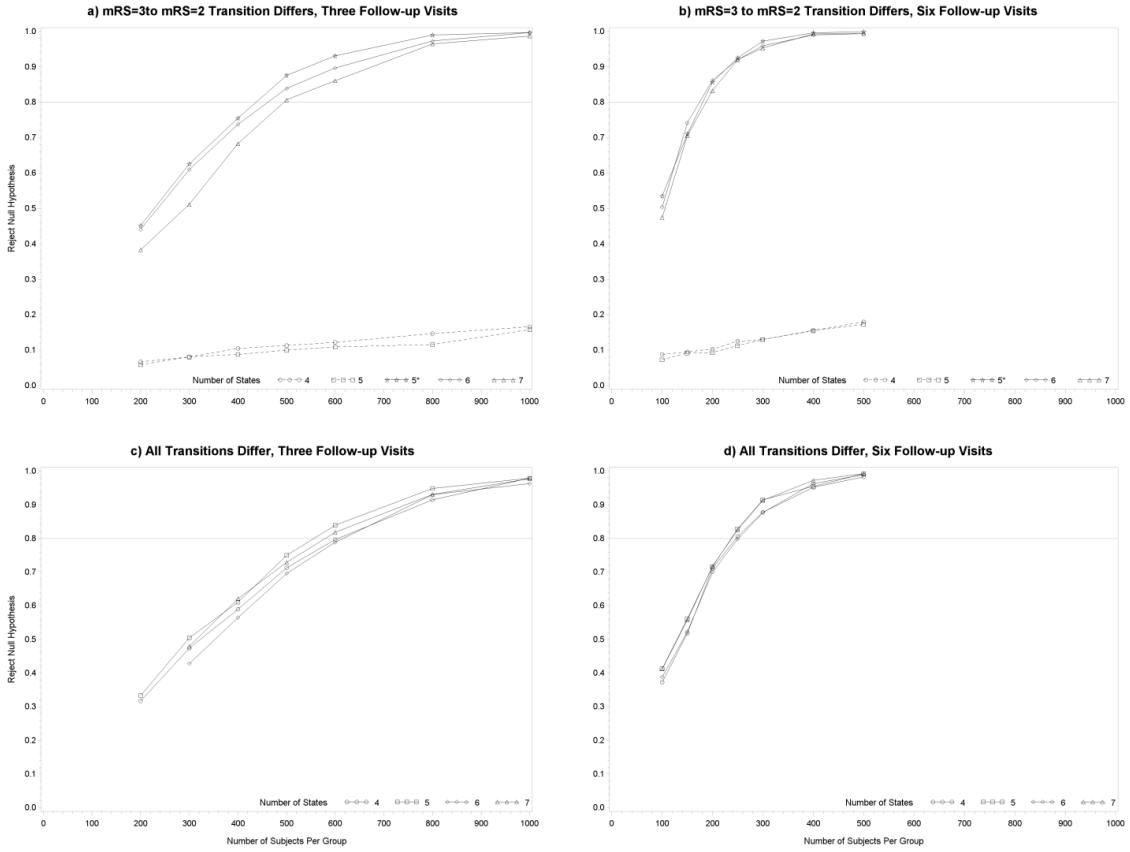


Figure 3.5: Graph of approximated power based on 1,000 simulations.

Figure 3.5d displays the approximated power where all assumed transition probabilities differ between groups and the number of follow-up visits is increased from three to six. As observed in the first scenario, the results from the models with six follow-

up visits are similar to those from the models with three follow-up visits, with a significant increase in power. The increase in power is expected since there are twice as many observation per subject contributing to the estimation of the model parameters. In this case, approximately 250 subjects are needed per group to reach 80% power.

Table 3.5. Comparison of power for models with three follow-up visits.

				Multistate Markov models				
Scenario	N per group	Logistic regression	van Elteren	7-state	6-state	5-state	4-state	3-state
2 to 1 only	200	0.257	0.117	-	-	-	0.206	0.256
	300	0.371	0.138	-	-	0.544	0.283	0.349
	400	0.469	0.193	0.635	0.663	0.670	0.381	0.468
	500	0.573	0.222	0.753	0.771	0.787	0.499	0.544
	600	0.646	0.224	0.837	0.848	0.870	0.567	0.626
	800	0.743	0.267	0.926	0.939	0.956	0.703	0.760
	1000	0.853	0.334	0.974	0.981	0.989	0.815	0.868
All shifts	200	0.735	0.741	-	-	0.405	0.358	0.470
	300	0.893	0.898	0.582	0.629	0.563	0.508	0.643
	400	0.960	0.970	0.694	0.763	0.723	0.615	0.765
	500	0.988	0.983	0.811	0.863	0.811	0.733	0.879
	600	0.996	0.997	0.875	0.943	0.897	0.814	0.933
	800	1.000	1.000	0.961	0.982	0.971	0.911	0.977
	1000	1.000	1.000	0.989	0.998	0.990	0.975	0.996

The approximated power from the models displayed in Figure 3.5 is compared with that from repeated logistic regression in Tables 3.5 and 3.6. Table 3.5 lists the power for the models with three follow-up visits and Table 3.6 lists the power for the models with six follow-up visits. Repeated logistic regression was performed using the dichotomized mRS scores, where scores of 0 or 1 were defined as successes and scores greater than 1 were defined as failures.

Table 3.6. Comparison of power for models with six follow-up visits.								
				Multistate Markov models				
Scenario	N per group	Logistic regression	van Elteren	7-state	6-state	5-state	4-state	3-state
2 to 1 only	100	0.119	0.075	-	-	0.474	0.288	0.321
	150	0.188	0.098	-	0.622	0.629	0.417	0.463
	200	0.257	0.117	0.732	0.734	0.776	0.546	0.578
	300	0.371	0.138	0.871	0.889	0.920	0.714	0.784
	400	0.469	0.193	0.957	0.961	0.967	0.860	0.906
	500	0.573	0.222	0.983	0.990	0.991	0.936	0.947
	600	0.646	0.224	0.994	0.992	1.000	0.997	0.982
All shifts	50	0.246	0.241	-	-	-	0.210	0.326
	75	0.350	0.357	-	-	0.419	0.358	0.485
	100	0.463	0.476	-	-	0.540	0.434	0.584
	125	0.523	0.519	0.579	0.673	0.659	0.521	0.725
	150	0.602	0.642	0.688	0.754	0.732	0.598	0.806
	175	0.687	0.690	0.735	0.827	0.794	0.662	0.867
	200	0.735	0.741	0.809	0.901	0.847	0.751	0.915
	300	0.893	0.898	0.945	0.975	0.968	0.906	0.984
	400	0.960	0.970	0.985	0.994	0.992	0.981	0.997

When only one assigned transition probability differs between groups, in correctly specified models, the multistate Markov model requires significantly fewer subjects than the repeated logistic regression model to be adequately powered. When the multistate model is misspecified, the repeated logistic regression is more powerful. When all assumed transition probabilities differ between groups, the repeated logistic regression requires fewer subjects per group to reach 80% power. When there are three follow-up visits, the repeated logistic regression model only requires about 300 subjects per group to be adequately powered, compared to 600 in the multistate model. In the six follow-up

visit case, approximately 150 subjects are needed per group compared to 250 in the multistate Markov model.

3.4 Summary and Discussion

The mRS, one of the most commonly used outcome measures in acute stroke trials, is ordinal but is often dichotomized for analysis. The loss of information from dichotomizing the ordinal variable was examined in this article. In addition, despite the availability of multiple mRS scores over time in many trials, a single measurement is often chosen for primary analysis. The additional information available from the longitudinal data could add further efficiency to the analysis. Multistate Markov modeling is presented here as an alternative analytic approach for ordinal outcomes collected longitudinally. The multistate Markov model describes how a process moves between states over time, which is desirable because it lends itself to clinically relevant interpretations.

In this paper, we have considered time-homogenous continuous Markov multistate models for mRS outcome data observed in phase III acute stroke trials. Simulations demonstrated that the desired type I error probability is preserved for the likelihood ratio test comparing a multistate Markov model including treatment to one without. Power was examined for two different clinically relevant scenarios. The two scenarios represented two diverse instances where a treatment effect exists. In the first scenario all of the assigned transition probabilities were the same for the two treatment groups except the transition from mRS = 3 to mRS = 2. The assigned treatment

probabilities in the second scenario differed between the groups for all transitions, representing a positive treatment effect for all shifts.

The key findings of the simulation studies could be summarized as follows:

- When the only difference between the treatment groups in assigned transition probabilities is from mRS = 3 to mRS =2,
    - misspecification of the five-state (and four-state) multistate model drastically decreases power as this masks the only difference between groups, the transition from mRS = 3 to mRS = 2
    - the multistate model yielding the highest power is the 5-state* model where mRS = 4 and mRS = 5 are combined, as well as mRS = 0 and mRS = 1
    - power is not drastically different for the seven- six- or five-state* Markov model
    - the multistate model, when correctly specified, is more powerful than repeated logistic regression
- When all assigned transition probabilities differ between groups,
    - power is essentially equal for all four multistate Markov models considered
    - the repeated logistic regression models are more powerful than the multistate Markov models
- For both scenarios, and all combinations of states considered, increasing the number of follow-up visits from three to six drastically increased power.

We considered a case where two distinct states were combined to examine the effects of misspecification. It is important to note that for a process that is truly Markov on I states, a reduced-state model will not satisfy the Markov property [76]. The sojourn time will be non-exponential for the merged states and bias can be expected through the misspecification. This highlights the importance of correctly specifying models when using the multistate Markov approach. A modified version of Akaike's criterion could aid in model selection [77].

We conclude that multistate Markov modeling can be a more efficient approach to analysis of mRS data from acute stroke trials. There are situations where dichotomization might not lose efficiency and may be more powerful than the multistate Markov model. Depending on the observed data structure, either technique could be more powerful. In every model, however, increasing the number of follow-up visits from three to six dramatically improved the power to detect a treatment difference.

A limitation of this study is the computational intensity required to run the simulations. For the scenarios with a larger number of states, the time required to complete the simulations was lengthy. Because of the time these simulations take, each was only repeated 1,000 times. Larger simulation studies, say with 10,000 runs rather than 1,000, would improve the precision on the estimates of the operating characteristics. A second limitation of this study is the lack of effect size measurement. In order to quantify an effect size, we would need to be able to define what outcome would be of interest. For example, some previous studies have considered a 10% difference in proportion of good outcome, where good outcome is defined by a dichotomized mRS

scale. Quantification of the effect is not straightforward when using Markov multistate modeling. This is a practical question to consider in the future.

A future direction of this work could be to compare the results of multistate Markov modeling to repeated cumulative logistic regression. At the time of submission the authors could not find any publications where longitudinal proportional odds models or adjacent categories logit models were applied to mRS data. Interesting issues arise about how to handle the proportional odds assumption and how to compare models when the assumption fails. This may be a useful extension of the analysis of longitudinal mRS data.

3.5 Appendices

Appendix 3A: assumed transition probabilities for type I error simulations

In this appendix, we present Tables 3.1A-3.4A, which show the probabilities used to determine the trajectories for the subjects in the type I error simulation study.

Table 3.1A. Assumed transition probabilities for the seven-state model.

	Probabilities for time = 1						
	0	1	2	3	4	5	6
	0.1000	0.1200	0.1300	0.1500	0.2300	0.1400	0.1300
	Conditional Transition Probabilities for time > 1						
	0	1	2	3	4	5	6
0	0.8000	0.1700	0.0200	0.0050	0.0030	0.0010	0.0010
1	0.2000	0.6800	0.0800	0.0200	0.0100	0.0050	0.0050
2	0.0500	0.2800	0.5400	0.1100	0.0100	0.0010	0.0090
3	0.0200	0.0800	0.2200	0.6000	0.0600	0.0050	0.0150
4	0.0050	0.0150	0.0600	0.2300	0.6000	0.0700	0.0200
5	0.0005	0.0070	0.0075	0.0450	0.2800	0.4800	0.1800

Table 3.2A. Assumed transition probabilities for the six-state model.

	Probabilities for time = 1					
	0	1	2	3	4/5	6
	0.100	0.120	0.130	0.150	0.370	0.130
	Conditional Transition Probabilities for time > 1					
	0	1	2	3	4/5	6
0	0.800	0.170	0.020	0.005	0.004	0.001
1	0.200	0.680	0.080	0.020	0.015	0.005
2	0.050	0.280	0.540	0.110	0.011	0.009
3	0.020	0.080	0.220	0.600	0.065	0.015
4/5	0.003	0.011	0.034	0.138	0.714	0.100

Table 3.3A. Assumed transition probabilities for the five-state model.					
	Probabilities for time = 1				
	0	1	2/3	4/5	6
	0.100	0.120	0.280	0.370	0.130
Conditional Transition Probabilities for time > 1					
	0	1	2/3	4/5	6
0	0.800	0.170	0.025	0.004	0.001
1	0.200	0.680	0.100	0.015	0.005
2/3	0.035	0.180	0.735	0.038	0.012
4/5	0.003	0.011	0.172	0.714	0.100

Table 3.4A. Assumed transition probabilities for the four-state model.				
	Probabilities for time = 1			
	0/1	2/3	4/5	6
	0.220	0.280	0.370	0.130
Conditional Transition Probabilities for time > 1				
	0/1	2/3	4/5	6
0/1	0.924	0.063	0.010	0.003
2/3	0.215	0.735	0.038	0.012
4/5	0.014	0.172	0.714	0.100

Appendix 3B: assumed transition probabilities for the scenario with only one differing assumed transition

In this appendix, we present Tables 3.1B-3.4B, which show the probabilities used to determine the trajectories for the subjects in the simulation study to approximate power when the treatment effect exists for only one transition (from mRS = 3 to mRS = 2).

Table 3.1B. Assumed transition probabilities for the seven-state model.

	Probabilities for time = 1 (control)						
	0	1	2	3	4	5	6
	0.110 (0.090)	0.130 (0.110)	0.150 (0.110)	0.110 (0.190)	0.230 (0.230)	0.140 (0.140)	0.130 (0.130)
	Conditional transition probabilities for time > 1 (control)						
	0	1	2	3	4	5	6
0	0.800 (0.800)	0.170 (0.170)	0.020 (0.020)	0.005 (0.005)	0.003 (0.003)	0.001 (0.001)	0.001 (0.001)
1	0.200 (0.200)	0.680 (0.680)	0.080 (0.080)	0.020 (0.020)	0.010 (0.010)	0.005 (0.005)	0.005 (0.005)
2	0.050 (0.050)	0.280 (0.280)	0.540 (0.540)	0.110 (0.110)	0.010 (0.010)	0.001 (0.001)	0.009 (0.009)
3	0.030 (0.010)	0.110 (0.050)	0.300 (0.140)	0.480 (0.720)	0.060 (0.060)	0.005 (0.005)	0.015 (0.015)
4	0.005 (0.005)	0.015 (0.015)	0.060 (0.060)	0.230 (0.230)	0.600 (0.600)	0.070 (0.070)	0.020 (0.020)
5	.0005 (.0005)	0.007 (0.007)	.0075 (.0075)	0.045 (0.045)	0.280 (0.280)	0.480 (0.480)	0.180 (0.180)

Table 3.2B. Assumed transition probabilities for the six-state model.						
	Probabilities for time = 1 (control)					
	0	1	2	3	4/5	6
	0.110 (0.090)	0.130 (0.110)	0.150 (0.110)	0.110 (0.190)	0.370 (0.370)	0.130 (0.130)
	Conditional transition probabilities for time > 1 (control)					
	0	1	2	3	4/5	6
0	0.800 (0.800)	0.170 (0.170)	0.020 (0.020)	0.005 (0.005)	0.004 (0.004)	0.001 (0.001)
1	0.200 (0.200)	0.680 (0.680)	0.080 (0.080)	0.020 (0.020)	0.015 (0.015)	0.005 (0.005)
2	0.050 (0.050)	0.280 (0.280)	0.540 (0.540)	0.110 (0.110)	0.011 (0.011)	0.009 (0.009)
3	0.030 (0.010)	0.110 (0.050)	0.300 (0.140)	0.480 (0.720)	0.065 (0.065)	0.015 (0.015)
4/5	0.003 (0.003)	0.011 (0.011)	0.034 (0.034)	0.138 (0.138)	0.714 (0.714)	0.100 (0.100)

Table 3.3B. Assumed transition probabilities for the five-state model.					
	Probabilities for time = 1 (control)				
	0	1	2/3	4/5	6
	0.110 (0.090)	0.130 (0.110)	0.260 (.300)	0.370 (.370)	0.130 (0.130)
	Conditional transition probabilities for time > 1 (control)				
	0	1	2/3	4/5	6
0	0.800 (0.800)	0.170 (0.170)	0.025 (0.025)	0.004 (0.004)	0.001 (0.001)
1	0.200 (0.200)	0.680 (0.680)	0.100 (0.100)	0.015 (0.015)	0.005 (0.005)
2/3	0.040 (0.030)	0.195 (0.165)	0.715 (0.755)	0.038 (0.038)	0.012 (0.012)
4/5	0.003 (0.003)	0.011 (0.011)	0.172 (0.172)	0.714 (0.714)	0.100 (0.100)

Table 3.4B. Assumed transition probabilities for the four-state model.				
	Probabilities for time = 1 (control)			
	0/1	2/3	4/5	6
	0.120 (0.100)	0.150 (0.120)	0.230 (0.280)	0.500 (0.500)
	Conditional transition probabilities for time > 1 (control)			
	0/1	2/3	4/5	6
0/1	0.924 (0.924)	0.063 (0.063)	0.010 (0.010)	0.003 (0.003)
2/3	0.235 (0.195)	0.715 (0.755)	0.038 (0.038)	0.012 (0.012)
4/5	0.014 (0.014)	0.172 (0.172)	0.714 (0.714)	0.100 (0.100)

Appendix 3C: assumed transition probabilities for scenario with global treatment effect

In this appendix, we present Tables 3.1C-3.4C, which show the probabilities used to determine the trajectories for the subjects in the simulation study to approximate power when the treatment effect exists for all transitions.

Table 3.1C. Assumed transition probabilities for the seven-state model.

	Probabilities for time = 1 (control)						
	0	1	2	3	4	5	6
	0.150 (0.100)	0.150 (0.100)	0.140 (0.100)	0.100 (0.150)	0.200 (0.250)	0.200 (0.230)	0.050 (0.070)
	Conditional transition probabilities for time > 1 (control)						
	0	1	2	3	4	5	6
0	0.800 (0.720)	0.186 (0.230)	0.010 (0.042)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
1	0.200 (0.160)	0.693 (0.676)	0.080 (0.120)	0.001 (0.002)	0.005 (0.010)	0.001 (0.002)	0.020 (0.030)
2	0.050 (0.030)	0.280 (0.220)	0.509 (0.528)	0.100 (0.140)	0.020 (0.030)	0.001 (0.002)	0.040 (0.050)
3	0.020 (0.010)	0.130 (0.080)	0.230 (0.200)	0.560 (0.620)	0.040 (0.050)	0.010 (0.020)	0.010 (0.020)
4	0.010 (0.005)	0.030 (0.020)	0.060 (0.050)	0.250 (0.200)	0.510 (0.565)	0.070 (0.080)	0.070 (0.080)
5	0.002 (0.001)	0.010 (0.005)	0.010 (0.005)	0.070 (0.050)	0.250 (0.200)	0.408 (0.439)	0.250 (0.300)

Table 3.2C. Assumed transition probabilities for the six-state model.						
	Probabilities for time = 1 (control)					
	0	1	2	3	4/5	6
	0.150 (0.100)	0.150 (0.100)	0.140 (0.100)	0.100 (0.150)	0.400 (0.480)	0.060 (0.070)
	Conditional transition probabilities for time > 1 (control)					
	0	1	2	3	4/5	6
0	0.800 (0.720)	0.186 (0.230)	0.010 (0.042)	0.001 (0.002)	0.002 (0.004)	0.001 (0.002)
1	0.200 (0.160)	0.693 (0.676)	0.080 (0.120)	0.001 (0.002)	0.006 (0.012)	0.020 (0.030)
2	0.050 (0.030)	0.280 (0.220)	0.509 (0.528)	0.100 (0.140)	0.021 (0.032)	0.040 (0.050)
3	0.020 (0.010)	0.130 (0.080)	0.230 (0.200)	0.560 (0.620)	0.050 (0.070)	0.010 (0.020)
4/5	0.006 (0.003)	0.020 (0.013)	0.035 (0.028)	0.160 (0.125)	0.619 (0.641)	0.160 (0.190)

Table 3.3C. Assumed transition probabilities for the five-state model.					
	Probabilities for time = 1 (control)				
	0	1	2/3	4/5	6
	0.150 (0.100)	0.150 (0.100)	0.240 (0.250)	0.400 (0.480)	0.060 (0.070)
	Conditional transition probabilities for time > 1 (control)				
	0	1	2/3	4/5	6
0	0.800 (0.720)	0.186 (0.230)	0.011 (0.044)	0.002 (0.004)	0.001 (0.002)
1	0.200 (0.160)	0.693 (0.676)	0.081 (0.122)	0.006 (0.012)	0.020 (0.030)
2/3	0.035 (0.020)	0.205 (0.150)	0.700 (0.745)	0.035 (0.050)	0.025 (0.035)
4/5	0.006 (0.003)	0.020 (0.013)	0.195 (0.153)	0.619 (0.641)	0.160 (0.190)

Table 3.4C. Assumed transition probabilities for the four-state model.				
	Probabilities for time = 1 (control)			
	0/1	2/3	4/5	6
	0.300 (0.200)	0.240 (0.250)	0.400 (0.480)	0.060 (0.070)
	Conditional transition probabilities for time > 1 (control)			
	0/1	2/3	4/5	6
0/1	0.939 (0.893)	0.046 (0.083)	0.004 (0.008)	0.011 (0.016)
2/3	0.240 (0.170)	0.700 (0.745)	0.035 (0.050)	0.025 (0.035)
4/5	0.026 (0.016)	0.195 (0.153)	0.619 (0.641)	0.160 (0.190)

Appendix 3D: plots of p-values and test-statistics from type I error simulation study

In this appendix, we present Figures 3.1D and 3.2D, which display the distribution of the p-values and test-statistics from the likelihood ratio tests from the type I error simulation study.

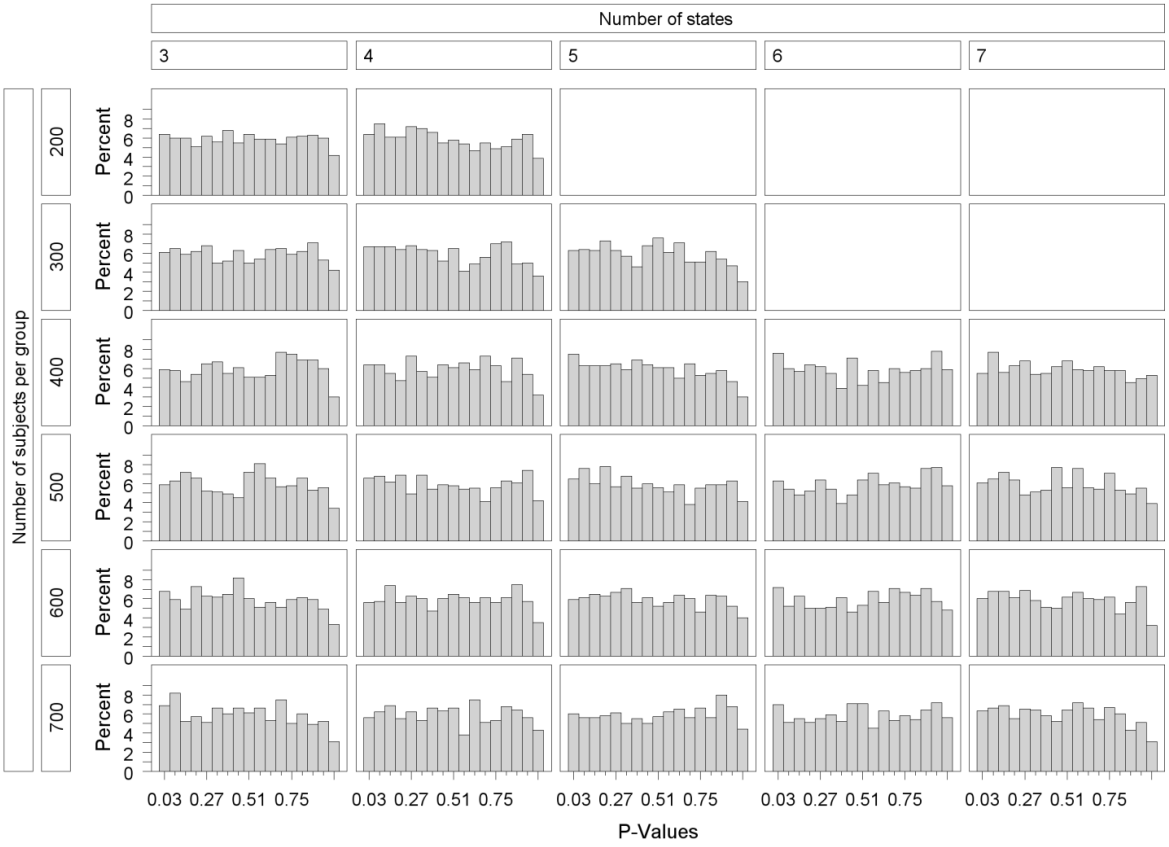


Figure 3.1D. Distribution of p-values from the likelihood ratio tests calculated in the type I error simulation study.

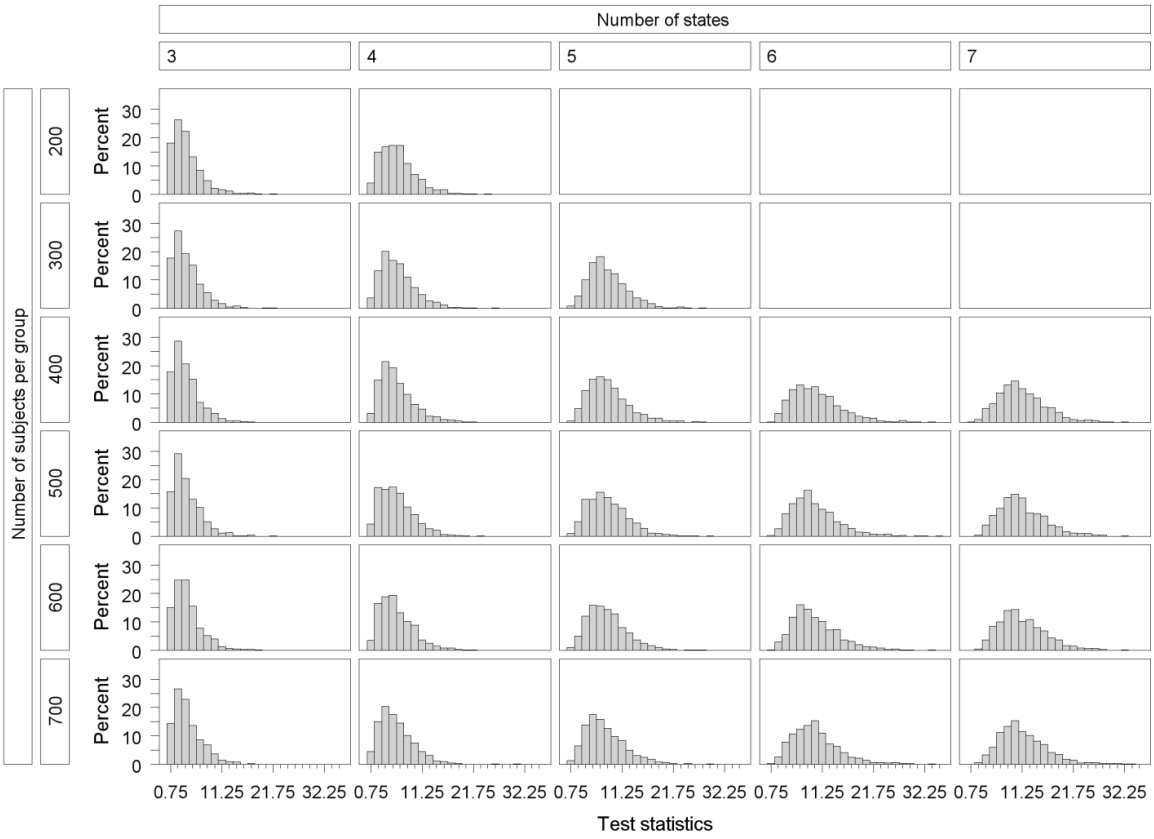


Figure 3.2D. Distribution of test-statistics from the likelihood ratio tests calculated in the type I error simulation study.

4 Original Manuscript 2

Title: Treatment effect on ordinal functional outcome using piecewise multistate Markov model with unobservable baseline: An application to the modified Rankin Scale

Authors: Christy Cassarly[1], Renee' H. Martin[1], Marc Chimowitz[2], Edsel A. Peña[3], Viswanathan Ramakrishnan[1], Yuko Y. Palesch[1]

Affiliation:

[1] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC

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

[3] Department of Statistics, University of South Carolina, Columbia, SC

Submission Status: submitted to *Statistics in Medicine*

Abstract

In clinical trials, longitudinally assessed ordinal outcomes are commonly dichotomized and only the final measure is used for primary analysis, partly for ease of clinical interpretation. Dichotomization of the ordinal scale and failure to utilize the repeated measures can reduce statistical power. Additionally, in a certain emergent settings, the same measure cannot be assessed at baseline prior to treatment. For such a data set, a piecewise-constant multistate Markov model that incorporates a latent model for the unobserved baseline measure is proposed. These models can be useful in analyzing disease history data and are advantageous in clinical applications where a disease process naturally moves through increasing stages of severity. Two examples are provided using

acute stroke clinical trials data. Conclusions drawn in this paper are consistent with those from the primary analysis for treatment effect in both of the motivating examples. Use of these models allows for a more refined examination of treatment effect and describes the movement between health states from baseline to follow-up visits which may provide more clinical insight into the treatment effect.

Keywords

longitudinal ordinal outcome; piecewise multistate models; panel data; modified Rankin Scale

4.1 Introduction

Outcomes on an ordinal scale is quite common in clinical trials [78]. It is also common in these trials to analyze the data using a dichotomized version of the ordinal measure. For instance, in treatments of acute stroke, often the modified Rankin Scale (mRS) is used as the primary outcome. The mRS measures functional ability using a 7-point ordinal scale (Table 4.1). Although, mRS is measured on this 7-point scale, for testing primary hypotheses of interest, it is commonly dichotomized either by collapsing into $\{0,1\}$ vs $\{2,3,4,5,6\}$ or $\{0,1,2\}$ vs $\{3,4,5,6\}$. Loss of information, when such ordinal outcomes are collapsed into a dichotomy, has been studied and shown to result in reduction of statistical power [79]. In stroke trials, alternative analytic methods on the observed ordinal scale are gaining attention [1].

Table 4.1. Modified Rankin Scale.

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities but able to look after own affairs without assistance
3	Moderate disability requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent, and requiring constant nursing care and attention
6	Dead

Another common approach in most longitudinal trials that measure mRS over time, is to use only the final measure made at 90 days, ignoring the earlier measurements

because the outcome status at 90 days is considered to be the only measurement of clinical relevance. Moreover, typically, the ultimate goal of a treatment or therapeutic action is to achieve improvement in patient health compared to their baseline measure. To accurately quantify improvement, a measure of the outcome at baseline would be ideal. Conditions with sudden onset, such as traumatic brain injury, stroke and status epilepticus often require immediate attention and treatment, posing a challenge to assess baseline outcome measures for clinical trials that may also lack practical meaning based on patient status. Specific to acute stroke trials the mRS is not obtainable at baseline [80]. The current method of addressing this is to adjust for the severity of the condition (disease) at baseline.

In this manuscript a continuous-time non-homogeneous Markov process is proposed as an alternative to study the evolution of acute onset diseases. Of specific interest is exploration of potential differences in transition rates between two treatment groups. Using this method, it is possible to analyze treatment effects in the observed ordinal scale and incorporate data measured longitudinally. In addition, after treatment, since much of the progression or recovery experienced by acute ischemic stroke patients is expected to occur early, on a method for predicting the baseline mRS state is proposed [81]. This baseline mRS may then be utilized in a model that more fully characterizes the evolution of disease over time.

The paper is organized as follows. Two motivating examples of large acute stroke therapy trials are described in Section 4.2. Homogeneous and piecewise-constant multistate Markov models (MSMMs) are introduced. In Section 4.3, the baseline

estimation procedure is described and demonstrated and piecewise-constant MSMMs using the estimated baseline scores for the motivating examples are fit. Concluding remarks are presented in Section 4.4.

4.2 Methods

Methods are developed and motivated through two phase III acute stroke therapy trials, namely the National Institute of Neurologic Disorders and Stroke tissue plasminogen activator (NINDS tPA) Stroke Study Part 2 [38] and the Albumin in Acute Stroke (ALIAS) Trial [82].

The NINDS tPA Stroke Study Part 2 was designed to compare intravenous tPA versus placebo in subjects with acute ischemic stroke using a global test statistic [38]. The global test statistic simultaneously tested for treatment effect in four correlated outcomes (mRS, Barthel Index, Glasgow outcome scale and National Institutes of Health Stroke Scale). The Barthel Index is a simple index of independence that scores the ability of patients to care for themselves [40]. Patients that can perform all activities assessed with complete independence are given a score of 100. The Glasgow outcome scale is a global assessment of function that ranges from 1 indicating good recovery to 5, death [41]. The National Institutes of Health Stroke Scale (NIHSS) is a 42-point scale that measures neurologic deficit where 0 indicates normal function [83].

A total of 624 patients were enrolled in NINDS tPA, with 312 in each treatment group. The primary analysis showed a significant global test score for the four dichotomized outcomes as well as for the dichotomized mRS alone at 90 days [39]. In

addition to the primary outcome assessment at 90 days, the mRS was collected at 7-10 days, 180 days and 360 days from randomization for each subject.

The ALIAS Trial was designed to compare 25% human serum albumin and saline in patients with acute ischemic stroke [82]. Part 1 of the trial was suspended after enrolling 434 subjects due to safety concerns of albumin [84]. Part 2 of the trial was slightly redesigned with unblinded safety analysis and enrolled 841 subjects. The analysis of both Parts 1 and 2, as well as the two combined, showed a lack of treatment effect on primary and secondary outcomes, including the dichotomized mRS at 90 days [85]. In addition to the primary outcome assessment at 90 days, the mRS was also collected at 30 days, 180 days, 270 days and 360 days from randomization for each subject.

4.2.1 Multistate Markov Models

MSMMs in continuous-time has been used to model the course of many diseases [4]. The MSMMs incorporate longitudinal ordinal data and can provide clinically relevant summary statistics to describe covariate effects, including sojourn times and transition rates. These models are advantageous in clinical applications, where a disease process naturally moves through increasing stages of severity [4]. Homogeneous continuous-time MSMMs, where the transition rates are assumed to be constant over time, have been used to analyze various diseases [86-88].

The assumption of homogeneity is not always realistic. With acute onset diseases, the rapid nature of onset and intervention likely characterize a process that changes quickly early on and tapers off after the initial acute recovery stage. The transition rates of the remainder of the longitudinal disease process, after the initial burst of rapid

movement post-baseline, are likely to differ from those observed in the acute phase. In this case, a non-homogeneous model with piecewise constant intensity rates can be used.

Multistate Markov modeling requires that the Markov property holds for the observed data. Consider a system with a finite state-space $\Omega = \{1, 2, 3, ..., I\}$, where I represents the number of states. Let $X(s+t)$ be a discrete random variable that indicates the state occupied by the system at time $s+t$. Let $\mathrm{F}_{X(s)} = \{X(v) : v \le s\}$ which denotes all information pertaining to the history of X up to time s [64]. The series of observations has the Markov property if the conditional distribution of $X(s+t)$ given $\mathrm{F}_{X(s)}$, satisfies

$$P\left\{X(s+t) = j \mid \mathrm{F}_{X(s)}, X(s) = i\right\} = P\left\{X(s+t) = j \mid X(s) = i\right\} = p_{ij}(s,t),\ i, j \in \Omega. \quad (4.1)$$

In other words, the present state depends only on the immediately preceding observation and not on the ones before it. In the context of clinical trials, the state of the system is the health state of one individual.

Though the ordinal outcome in acute therapy clinical trials is observed at discrete times, the disease process is continuous, where progression or recovery can occur at any time. Continuous-time MSMMs can analyze this type of data, known as panel data, as long as the sampling times are considered to be noninformative [4]. In the clinical trial setting where the observation scheme is fixed in advance, this assumption is valid. If, however, a subject visited a clinic because of a change in symptoms and the outcome was collected, the sampling time would be informative and could bias inference. These continuous-time MSMMs are flexible enough to model panel data with noninformative

sampling times but also include exact time of death, which is commonly observed in clinical trials.

A common assumption of continuous-time MSMMs is that of homogeneity, where transition probabilities remain constant over time. When homogeneity is assumed, the transition probabilities, $p_{ij}(t)$, are defined as

$$p_{ij}(t) = P\{X(s+t) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\}. \tag{4.2}$$

Since this expression does not depend on s, the transition from state i to state j on a time interval of length t has the same probability at any time. The $p_{ij}(t)$ are elements of the transition probability matrix, $\mathbf{P}(t)$.

The movement of a subject between states is described by λ_{ij}, the transition intensities:

$$\lambda_{ij} = \lim_{\delta t \to 0} \frac{P(X(\delta t) = j \mid X(0) = i)}{\delta t}, \text{ for } i \neq j. \tag{4.3}$$

The intensities represent the instantaneous rate of moving from state i to state $j \neq i$ and form the generator matrix, $\mathbf{\Lambda}$, whose rows sum to zero and the diagonal entries are $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$. The transition probability matrix $\mathbf{P}(t)$ can be solved by taking a matrix exponential of $\mathbf{\Lambda}$ scaled by the time interval,

$$\mathbf{P}(t) = e^{\mathbf{t\Lambda}} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{\Lambda}^k}{k!} \tag{4.4}$$

where $\mathbf{\Lambda}^k$ is the k$^{\text{th}}$ power of the generator matrix $\mathbf{\Lambda}$.

Suppose X is observed over $t_0 < t_1 < t_2 < ... < t_M$. Let $i_0, i_1, i_2, ..., i_M$ be the observed states over these time points. Then, the associated likelihood function is

$$\begin{aligned} L(\Lambda \mid x) &= P(\mathrm{X}_0 = i_0) \prod_{j=i}^{M} P(X_{t_j} = i_j \mid X_{t_{j-1}} = i_{j-1}) \\ &= P(\mathrm{X}_0 = i_0) \prod_{j=i}^{M} P(X_{t_j - t_{j-1}} = i_j \mid X_0 = i_{j-1}) \\ &= P(\mathrm{X}_0 = i_0) \prod_{j=i}^{M} P_{i_{j-1} i_j}(t_j - t_{j-1}) \, . \end{aligned} \tag{4.5}$$

Using equations (4.4) in (4.5), the likelihood reduces to

$$L(\Lambda \mid x) = P_{i0} \prod_{j=1}^{M} \left(\sum_{k=0}^{\infty} \frac{(t_j - t_{j-1})^k (\Lambda^k)_{i_{j-1} i_j}}{k!} \right), \tag{4.6}$$

where P_{i0} is the initial probability that the process is at i_0.

4.2.2 Piecewise-constant Multistate Markov Models

In the case of ischemic stroke occurrence and treatment, subjects can get better or worse very quickly. For this reason, the time homogeneity assumption is expected to fail for the first transition, from the estimated baseline to the first observed outcome. Therefore, assumption of homogeneity is relaxed and a non-homogeneous model is considered.

The MSMM for panel data can be extended to accommodate piecewise-constant intensity matrices for the non-homogeneous case [59]. Here, the transition probability functions are dependent on s and the transition matrix function is $\mathbf{P}(s,t) = (p_{ij}(s,t))$. The transition intensity functions are now defined by

$$\lambda_{ij}(t) = \lim_{\delta t \to 0} \frac{P(X(\mathrm{t}+\delta t) = j \mid X(\mathrm{t}) = i)}{\delta t},\ i, j \in S,\ i \neq j\,. \tag{4.7}$$

The time-homogeneity assumption can be tested by using a likelihood ratio test to compare the time-homogenous model to piecewise-constant models with different cutoff times.

The effect of a covariate, specifically treatment, is incorporated into the model as transition intensity functions [69]. Let z be a vector of observed covariates then

$$\lambda_{ij}(t;z) = \lambda_{ij}(t)\exp(z'\beta_{ij}(t)),\ \ \lambda_{ij}(t) > 0 \tag{4.8}$$

where $\beta_{ij}(t)$ is the parameter vector associated with the covariate vector z in the transition between states i and j in time t. The transitional rates are represented by $\lambda_{ij}(t;z)$ at time t for the subjects with vector z.

4.2.3 Latent Baseline Estimation

To consider the full evolution of ischemic stroke over time, an estimated baseline functional outcome is needed because baseline mRS is not obtainable in the acute setting. While functional outcome is not available at baseline, many other measures that are correlated with functionality are available. An estimation procedure using baseline characteristics known to be highly correlated with the mRS was developed.

As a preliminary step, to summarize information from numerous baseline measurements considered clinically relevant for functional outcome in ischemic stroke patients, data reduction was performed using a Principal Components Analysis (PCA). The items included in the PCA were age, baseline glucose, time from stroke onset to

randomization, the Alberta Stroke Program Early CT score (ASPECTS), and NIHSS. Age and time from stroke onset to randomization are two very well established predictors of outcome in acute stroke. Another baseline characteristic associated with poor outcome is "stress" hyperglycemia [89]. This hyperglycemia can be quantified using acute post-stroke glucose levels. The ASPECTS is also a strong prognostic indicator of outcome [90], which is a 10-point quantitative topographic CT scan score, where a normal scan receives a score of 10 [91]. For each defined region of the brain, a point is subtracted if there is evidence of ischemic change. The NIHSS is commonly used to measure baseline stroke severity. The individual items of the scale are presented in Appendix 4A. Baseline NIHSS is known to strongly predict outcome in acute stroke therapy trials [92]. Although total score is typically used for indicating stroke severity, each item of the scale was used in the PCA individually in order to more efficiently assess the contribution of each facet of the scale.

After reducing the data to fewer PCA's sextiles (six categories because the seventh category, namely mRS = 6 corresponds to death) based on the joint distributions of the PCA's will be used to define the baseline states of the individuals. Then the MSMM likelihood ratio tests will be used to compare treatment effects. However, in this likelihood ratio test, the uncertainty of the estimated baseline states has to be considered. This is achieved through bootstrap approach, using which an empirical distribution was derived to determine the p-values. The steps used in this non-parametric bootstrap approach are as follows:

1. Sample with replacement from the original dataset 1,000 times.

2. Use the baseline estimation procedure from the original data on each of the 1,000 bootstrapped samples (fixing the number of significant components as well as the variables used in each of the component score calculation).
3. Fit the piecewise MSMM to each resample and obtain the test statistic.
4. Compare the original test statistic to the new bootstrap distribution, made up of the 1,000 test statistics from the bootstrap samples.
5. The bootstrap p-value is calculated by finding the proportion of bootstrap samples in which the test statistics is larger than or equal to the one calculated from the original sample.

4.3 Results

Detailed descriptions of the PCA for each trial are presented in Appendix 4B. For both NINDS tPA and ALIAS, most of the variability was adequately explained by the first two components. Thus, component scores were calculated for Components 1 and 2 for each trial. Larger values on the component scores were expected to be associated with worse functional outcome. To assign the baseline mRS state, each of the component scores were divided into sextiles with equal probability. In Figure 4.1, the joint distribution of the two discretized scores are shown were used to assign values of mRS = 0 to mRS = 5 with equal probability (subjects cannot be dead at baseline, so no one was assigned an mRS = 6). There is uncertainty in the assignment of baseline states which needs to be accounted for in the hypothesis tests comparing treatments.

In the NINDS tPA trial, several subjects either did not have available CT scans or the scans were not of sufficient quality to obtain ASPECTS (16/624). Scores for the

second component could not be calculated and subsequently the baseline mRS could not be estimated. These subjects were not excluded from the analysis unless they died in the first week of follow-up, leaving them with only one observed mRS score (5/624). Therefore, the total number of subjects included in the MSMM was 619, where 320 were randomized to receive tPA and 323 to receive placebo.

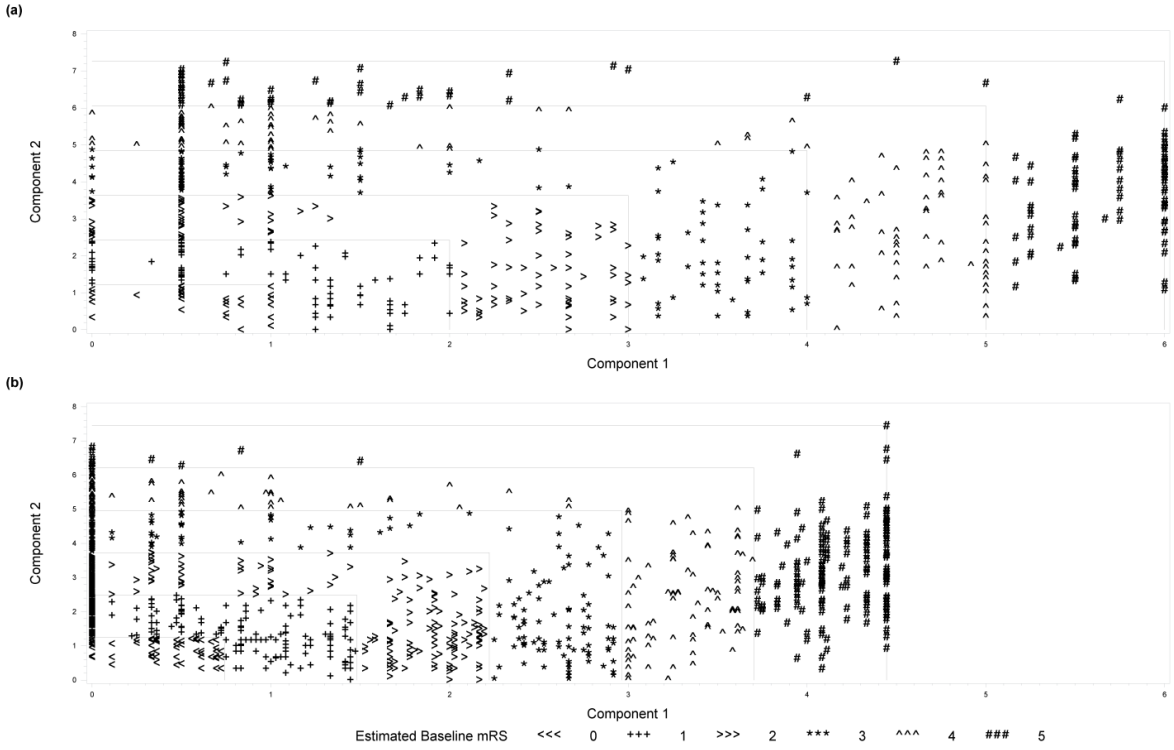


Figure 4.1: Baseline mRS score from summed standardized component scores for: (a) NINDS tPA and (b) ALIAS.

4.3.1 The Longitudinal Data

In the ALIAS Trial, a small number of subjects withdrew consent or were lost-to follow-up prior to the 30 day visit and had only one observation available, the estimated baseline mRS (17/1275). Excluding these subjects, a total of 1258 were included in the MSMM, where 628 were randomized to receive albumin and 630 to receive placebo. In

Appendix 4C, baseline characteristics of the 17 excluded subjects are summarized alongside those of the subjects that were not excluded. No notable differences exist between the groups and thus the exclusion of these subjects should be inconsequential in the analysis.

4.3.2 Development of MSMM for the NINDS tPA Study and ALIAS

In order to analyze panel data with continuous-time Markov chains it is important to consider which transitions can realistically occur in continuous time. When states represent severity it is assumed that in order for a subject to transition from one state to another non-adjacent state they also transition through the intermediate states. Thus, a reduced transition intensity matrix should be estimated, where non-adjacent state transitions are fixed to equal zero. The exception is when a state represents death. The reality is that a subject can die from any state.

In practice, if there is not enough information from the data, on a certain transition rate, more transition intensities may need to be set to zero [4]. State tables display counts of the pairs of transitions between states in successive observation times and summarize them in frequency tables of previous state against current state. These state tables can be used to identify counts that are too few to model.

The state table of all aggregate transtions from the NINDS tPA data (baseline to 360 days) is displayed in Table 4.2. Even though it is possible for subjects to die from any state, it is highly unlikely to occur from states 1, 2, 3 or 4. This is not surprising as subjects are only observed over the course of one year. The relatively healthy subjects have a low risk of death. Therefore more constraints are required for this model. When

the table is stratified by time for the piecewise model and again by treatment, the frequency of death from the healthier states (1, 2, 3 and 4) is too small to estimate the transition intensities. As a result, the original model was modified to no longer allow death from any state. Constraints were imposed such that death was only allowable from states 5 and 6.

Table 4.2. NINDS tPA state table.

	To (state j)							
	1	2	3	4	5	6	7	Total
From (state i)								
1	206	52	9	3	5	0	2	277
2	87	245	39	12	23	7	8	421
3	23	73	85	35	32	14	10	272
4	13	41	62	148	52	25	10	351
5	11	26	27	84	187	53	41	429
6	4	8	7	25	83	142	83	352
Total	344	445	229	307	382	241	154	2102

Similarly, when the state table was examined for ALIAS, small counts were observed for death from states 1, 2 and 3. The reduced allowable transition matrix for ALIAS fixed the intensities from these states to death to equal 0, only allowing the intensities from 4, 5 and 6 to death to be estimated.

4.3.3 Analysis of the NINDS tPA Study and ALIAS

The entries of the transition intensity matrices were estimated by applying the maximum-likelihood method and accounting for the two constant intervals partitioned at time $t = .333$ (representative of 7-10 days on a month-long interval) in the NINDS tPA trial and time $t = 1$ (representative of 30 days) in ALIAS. Analysis was performed using the *msm* package in R using the Broyden-Fletcher-Goldfarb-Shanno method [4].

For the NINDS tPA trial, using a likelihood ratio test to compare the piecewise model including treatment (-2LL = 6251) to one without treatment (-2LL = 6282), the model with treatment was preferred, indicating a statistically significant treatment effect on the transition rates ($p = 0.002$, $df = 12$). It is interesting to note that in models excluding estimated baseline mRS, the comparison of a model with treatment (-2LL = 4033) adjusting for baseline NIHSS to one without treatment (-2LL = 4053) failed to detect a statistically significant treatment effect ($p = 0.053$, $df = 12$).

In ALIAS, using a likelihood ratio test to compare the piecewise model including treatment (-2LL = 12730) to one without (-2LL = 12745), the reduced model was preferred, indicating no significant treatment effect on the transition rates ($p = 0.29$, $df = 13$), confirming the results from the primary analysis of the trial [43].

For NINDS tPA, after comparing the empirical distribution to the observed test statistic of 31.18, the p-value from the bootstrap procedure was 0.04 which is larger than the observed p-value of 0.002 but still indicative of a significant treatment effect. For ALIAS, the p-value calculated from the bootstrap procedure, where the test statistic from the analysis of the original sample was 15.25, was 0.49 which is also larger than the observed p-value of 0.29 but results in the same conclusion that there is no a significant treatment effect.

Table 4.3 shows the transition intensities estimated in the piecewise-constant model for the NINDS tPA Stroke Study. Transition rates differ between the placebo and tPA groups. The differences between groups can also be presented using hazard ratios

(Table 4.4). A hazard ratio greater than one indicates that the rate of transition is higher in the treatment group.

The only statistically significant hazard ratio was from state 2 to state 3 (HR = 0.51). This can be interpreted as the most significant impact of tPA is to reduce the hazard of transitioning from mRS = 1 to mRS = 2. The other hazard ratios, although not statistically significant, suggest a trend of tPA reducing the hazard of negative transitions.

Table 4.3. Maximum-likelihood estimates of transition rates among states.

	Placebo		tPA	
Transition	$0 \leq t \leq 0.333$	$t > 0.333$	$0 \leq t \leq 0.333$	$t > 0.333$
1→2	10.83 (2.36, 49.71)	0.09 (0.06, 0.14)	7.91 (1.89, 33.06)	0.07 (0.05, 0.09)
2→3	19.30 (6.07, 61.37)	0.09 (0.06, 0.15)	9.75 (3.15, 30.22)	0.05 (0.03, 0.07)
3→4	12.84 (6.53, 25.28)	0.19 (0.12, 0.32)	16.59 (7.09, 38.86)	0.25 (0.14, 0.44)
4→5	24.03 (10.50, 54.98)	0.12 (0.08, 0.19)	16.07 (6.71, 38.50)	0.08 (0.05, 0.13)
5→6	2.97 (1.92, 4.59)	0.05 (0.04, 0.10)	2.63 (1.70, 4.07)	0.05 (0.03, 0.09)
5→7	0.48 (0.27, 0.86)	0.05 (0.04, 0.07)	0.46 (0.26, 0.83)	0.05 (0.03, 0.07)
6→7	0.50 (0.28, 0.88)	0.12 (0.09, 0.18)	0.60 (0.35, 1.02)	0.15 (0.10, 0.21)
2→1	7.53 (1.64, 34.55)	0.08 (0.05, 0.11)	7.54 (1.85, 30.72)	0.08 (0.06, 0.10)
3→2	18.06 (5.49, 59.44)	0.23 (0.16, 0.32)	20.59 (6.43, 65.91)	0.26 (0.18, 0.36)
4→3	14.15 (6.74, 29.71)	0.22 (0.15, 0.32)	14.06 (6.09, 32.46)	0.22 (0.15, 0.33)
5→4	6.29 (2.59, 15.26)	0.20 (0.15, 0.27)	6.86 (2.86, 16.46)	0.22 (0.16, 0.30)
6→5	3.08 (2.01, 4.72)	0.23 (0.17, 0.32)	2.46 (1.61, 3.77)	0.18 (0.13, 0.26)

Table 4.4. Hazard ratios (95% CI).

1→2	0.73 (0.43, 1.23)
2→3	0.51 (0.28, 0.90)
3→4	1.29 (0.66, 2.52)
4→5	0.67 (0.38, 1.17)
5→6	0.88 (0.53, 1.48)
5→7	0.96 (0.55, 1.66)
6→7	1.20 (0.75, 1.92)
2→1	1.00 (0.62, 1.62)
3→2	1.14 (0.71, 1.83)
4→3	0.99 (0.59, 1.68)
5→4	1.09 (0.72, 1.65)
6→5	0.80 (0.52, 1.23)

4.4 Summary and Discussion

Dichotomization of ordinal outcomes is common but results in a loss of information and can reduce statistical power. Some patients with severe disability at baseline may never have the potential to achieve success as defined by the dichotomy. Thus, the prognostic heterogeneity of subjects does not allow for potential equal contribution to the estimation of treatment effect for all subjects with a dichotomized outcome [9].

A number of alternative methods for ordinal outcome data have received attention in recent years [93]. Linear regression and analysis of variance have been suggested where the ordinal outcome is treated as a continuous variable. Summary statistics from these models do not have straightforward interpretations because non-integer values from ordinal scales do not have a clear meaning.

A number of ordinal analyses have also been suggested. Ordinal logistic regression, under the assumption of proportional odds, assumes an identical effect of the predictors for each cumulative probability [45]. If the proportional odds assumption holds, statistical power can be increased as compared to analysis using a strict dichotomy. The score test for assessing the proportional odds assumption, however, is anticonservative. If the assumption fails, this analysis could mask important effects at one end of the ordinal outcome [46]. The partial proportional odds model relaxes this assumption and includes a term that allows the odds ratios to increase proportional to the outcome scale [46]. Alternatively, the cumulative logit model allows for the calculation of odds ratios for each adjacent category of response in relation to covariates and does not require the proportional odds assumption [45]. One drawback of both the partial

proportional odds and cumulative logit models is that they can require a larger sample size to be adequately powered.

The sliding dichotomy is another alternative method for the analysis of ordinal outcomes. It allows for the definition of success to vary based on patient-specific baseline prognostic variables while maintaining a dichotomized outcome, however, there are no guidelines for selection of number of prognostic groups nor cut points for those groups [48]. Poor selection of these groups could lead to a reduction in power. Furthermore, while the sliding dichotomy allows for baseline severity adjusted outcome, it still ignores any non-successful transitions [94].

The Cochran-Mantel Haenszel (CMH) shift test can also be used to analyze the distribution of ordinal data [7]. This test can show whether a treatment causes a significant shift toward good outcome. Shift analysis can account for ordered categories, has no distributional assumptions and is easy to implement. However, it is not feasible for large scale clinical trials with non-simple randomization schemes because it can only accommodate a limited number of covariates. There are also no summary statistics that appeal to a clinical audience so proportional odds logistic regression is often used in conjunction with the CMH test to provide an estimate of treatment effect [44]. In addition, shift analysis assumes that a treatment effect exists only in one direction, where only benefit is considered, not harm.

An approach to transform the mRS into a patient-centered outcome measure was recently proposed [53]. The chosen patient-centered outcome measure was utility, which is the desirability of a specific health outcome to a patient [54]. The utility weights were

derived for each level of the mRS by averaging utility values derived in two studies, using two different methods. Analysis using the utility weights is straightforward as the weights have already been defined and the utility-weighted mRS is analyzed using a *t* test. Though this method is easy to implement and provides greater statistical power, it is based on only two populations and may not be representative of patients in other locations. In addition, the utility values do not have as clear of an interpretation as some other analysis methods.

None of the previously mentioned methods utilize the repeated measures even though outcome is collected over time. In fact, a literature search for repeated measures analysis of acute stroke trial data only returned one article where a generalized estimating equations approach was used for repeated measures analysis [11]. This approach only considered the dichotomized outcomes from the NINDS tPA study. The work presented here is the first known study of the repeated measures acute stroke therapy data using the ordinal scale.

The results presented in this manuscript are the first to estimate a missing baseline ordinal outcome for use in a MSMM. In the case of ischemic stroke occurrence and treatment much of the progression or recovery experienced by a patient is expected to occur early. Functional outcome measures are not suitable at baseline and as a result, functional changes over time from baseline cannot be measured. Therefore, latent estimation the functional baseline was warranted, allowing for inclusion of an informative transition from baseline to first follow-up to be included in a longitudinal model.

Using the longitudinal data, including the estimated baseline, this work showed that there are differences in the rate of transitions between the treatment and placebo groups in the NINDS tPA trial, confirming the results of the primary analysis while allowing for examination of the effect on all adjacent-state transitions. The ability to estimate the intensities for all adjacent-state transitions allowed for examination of the most significant effect of treatment. Specifically, it was determined that the most significant impact of tPA is reduction the hazard of transitioning from mRS = 1 to mRS = 2. In addition, the conclusion of no treatment effect in the ALIAS Trial data was also consistent with the primary analysis from that trial.

In the MSMM of the NINDS tPA data, where the estimated baseline was not included and the model adjusted for baseline NIHSS instead, as is done in most other types of analysis of the mRS, the effect of treatment was only marginally significant. Thus, it seems as though inclusion of the estimated baseline mRS improved the ability to detect a treatment effect. It is hypothesized that the inclusion of the latently estimated baseline mRS is improves the model because of the acute nature of ischemic stroke therapy and the expected early recovery and disease progression directly following treatment.

The MSMM can incorporate longitudinal ordinal data and provide clinically interpretable summary statistics to describe covariate effects on all transition rates and sojourn times. Estimation of transition rates can describe treatment effect in a much finer gradient than modeling collapsed ordinal scale allowing for a more comprehensive understanding of differences in the effect of treatment. The MSMM also allows for

specific hypotheses to be tested-- for example a likelihood ratio test could be used to test whether the effect of treatment is the same for all forward transitions and backward transitions [59].

Another benefit of these models is the potentially decreased sample size needed to detect a treatment effect. A previous simulation study indicated that the power of MSMMs applied to acute onset clinical trial data was significantly increased when the number of follow-up visits was increased [93]. Future trials could collect the ordinal outcome more frequently over the course of follow-up, increasing the power to detect differences using this modeling technique.

A limitation of MSMMs is that they are computationally intensive, especially when using bootstrapping to obtain the bootstrap empirical distribution. In addition, use of these models requires a priori decisions about the transitions that can realistically occur, which may be a data driven decision. Interpretation of the full model could potentially be overwhelming, as there are many parameters that describe the effect of one covariate; however, the model also allows for testing whether the effect of a covariate is the same for certain transitions, which could reduce the number of parameters. The flexibility to estimate the full or reduced model allows a number of clinical questions to be answered using one approach.

Future directions could include more complex methods for baseline estimation. For example, a Bayesian PCA could be used in alternative baseline estimation procedure. This method could potentially address the uncertainty of the assigned baseline scores via extraction of the posterior distributions for the component scores [95]. In addition, the

methods presented for estimation could be extended to cases where more than two components were used in the estimation procedure.

Acknowledgements

Supported by grant U01-NS061861-05 from the National Institute of Neurological Disorders and Stroke.

4.5 Appendices

Appendix 4A: modified NIHSS summary

In this appendix, Table 4.1A is presented, which shows the 15 items of the NIHSS. The form for recording the data contains detailed instructions for use of the scale.

Table 4.1A. The Modified National Institutes of Health Stroke Scale Summary.		
1A	Level of Consciousness	0 = Alert 1 = Not alert, obtunded 3 = Unresponsive
1B	LOC Questions	0 = Answers both correctly 1 = Answers one correctly 2 = Answers neither correctly
1C	LOC Commands	0 = Performs both tasks correctly 1 = Performs one task correctly 2 = Performs neither task
2	Best Gaze	0 = Normal 1 = Partial gaze palsy 2 = Total gaze palsy
3	Visual	0 = No visual loss 1 = Partial hemianopia 2 = Complete hemianopia 3 = Bilateral hemianopia
4	Facial Palsy	0 = Normal 1 = Minor paralysis 2 = Partial paralysis 3 = Complete paralysis
5	Motor Arm a. Left b. Right	0 = No drift 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement
6	Motor Leg c. Left d. Right	0 = No drift 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement
7	Limb Ataxia	0 = Absent 1 = Present in one limb 2 = Present in two limbs
8	Sensory	0 = Normal 1 = Mild to moderate sensory loss 2 = Severe to total sensory loss
9	Best Language	0 = No aphasia, normal 1 = Mild to moderate aphasia 2 = Severe aphasia 3 = Mute or global aphasia
10	Dysarthria	0 = Normal 1 = Mild to moderate 2 = Severe
11	Extinction and Inattention (Neglect)	0 = No abnormality 1 = Mild 2 = Severe

Appendix 4B: data reduction using principal component analysis

When applied to a set of variables, PCA can group correlated variables into a smaller set of composite variables (principal components). The resulting linear combinations of variables account for as much variability in the data as possible and can be used to calculate component scores. A number of methods exist for computation of component scores. Non-refined methods are simple, easy to compute and easy to interpret, while refined methods are more complex and exact [96]. Refined computation methods are generally less stable across samples; hence, a non-refined method is implemented for the motivating data sets. Of the non-refined methods, the summation of standardized variables is preferred when the standard deviations of the raw data vary widely, as was found in the NINDS tPA and ALIAS data.

PCA was used to group measures of severity (individual items of the NIHSS) and other baseline variables known to be associated with functional ability (age, baseline glucose, time from stroke onset to randomization and ASPECTS score) into components and calculate component scores. For the analysis using PCA, variables were reformatted so that the direction of effect was consistent across all candidate variable and standardized scales were used. Component scores from PCA are more intuitive if the expected relationship with the variables and outcome is in the same direction. For example, increased age increases the risk of a negative outcome. All of the variables included in the PCA are positively correlated with bad outcome except ASPECTS. A smaller ASPECTS is predictive of negative outcome. For interpretation, the direction of the ASPECTS scale was reformatted such that 10 represented the worst and 0 normal.

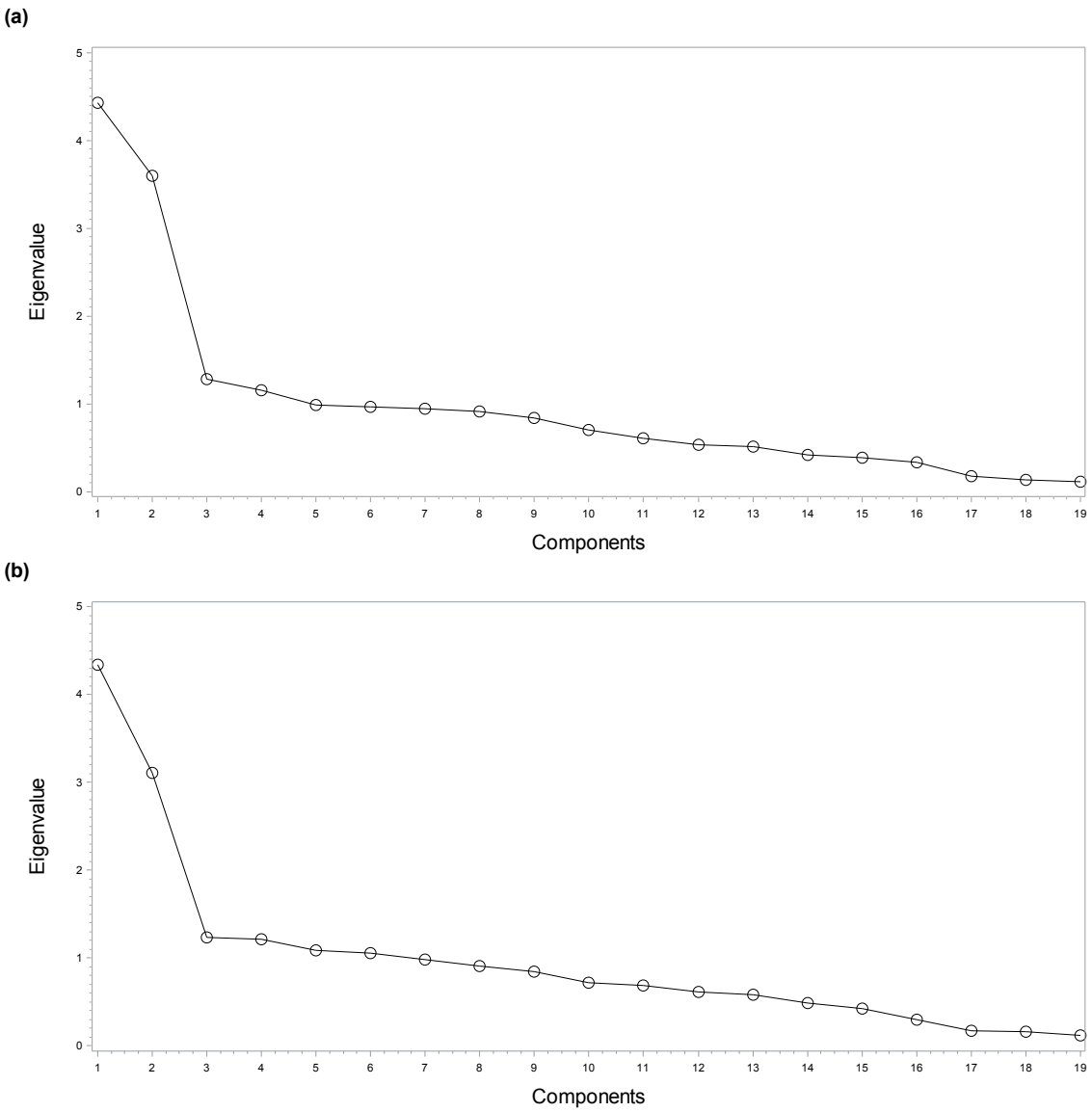


Figure 4.1B. Scree plot of eigenvalues from PCA of baseline variables from: (a) NINDS tPA and (b) ALIAS.

In the process of determining the number of components to retain from the PCA analysis, the eigenvalues of the correlation matrix were visually assessed using scree plots (Figure 4.1B). From this illustration it can be seen that the line flattens after the second component in the scree plot for both NINDS tPA (a) and ALIAS (b) and thus the first two components were found to adequately explain most of the variability. In order to

assess how strongly the components and variables are related, the component loadings for retained components were examined. Component loadings can be interpreted as the correlation between an observed variable and a component. A general rule of thumb deems components with loadings of at least |.4| highly explanatory [97].

The loadings for the first two components are displayed in Table 4.1B. Based on the criteria outlined above, baseline glucose, time from onset to randomization, and individual NIHSS items Q1A and Q07 were not included in the component score calculation for NINDS tPA. Age, baseline glucose, time from onset to randomization, and individual NIHSS items Q07 and Q10 were not included in the component score calculation for ALIAS.

Table 4.1B**.** Factor loadings on components based on PCA.

	NINDS tPA Components		ALIAS Components	
	1	2	1	2
Age	0.15	-0.03	0.16	0.13
Baseline glucose	0.10	-0.04	0.06	-0.02
Stroke onset to randomization	-0.08	-0.06	0.03	-0.04
NIHSS item				
Q1A – Level of Consciousness (LOC)	0.38	0.39	0.32	0.42
Q1B – LOC Questions	0.78	-0.18	0.80	-0.01
Q1C – LOC Commands	0.75	-0.01	0.78	0.14
Q02 – Best Gaze	0.37	0.66	0.20	0.71
Q03 – Visual	0.36	0.61	0.25	0.64
Q04 – Facial Palsy	0.22	0.50	0.01	0.48
Q5A – Motor Arm Left	-0.39	0.77	-0.63	0.58
Q5B – Motor Arm Right	0.84	-0.20	0.82	-0.01
Q6A – Motor Leg Left	-0.25	0.74	-0.53	0.56
Q6B – Motor Leg Right	0.80	-0.11	0.76	0.03
Q07 – Limb Ataxia	-0.20	-0.21	-0.15	-0.24
Q08 – Sensory	0.20	0.60	-0.02	0.54
Q09 – Best Language	0.86	-0.19	0.87	-0.03
Q10 – Dysarthria	0.60	0.04	0.35	0.22
Q11 – Extinction and Inattention	0.15	0.70	-0.09	0.68
ASPECTS	0.15	0.49	0.15	0.48

Each of the variables deemed important with the cutoff of >|.4| on the first two components were standardized to range [0, 1]. The standardized values of variables that loaded high on each component were summed to calculate the component scores as follows:

- NINDS tPA Component Score 1 =

  Q1B + Q1C + Q5B + Q6B + Q09 + Q10

- NINDS tPA Component Score 2 =

  Q02 + Q03 + Q04 + Q5A + Q6A + Q08 + Q11 + ASPECTS

- ALIAS Component Score 1 =

  Q1B + Q1C + Q5B + Q6B + Q09

- ALIAS Component Score 2 =

  Q1A + Q02 + Q03 + Q04 + Q5A + Q6A + Q08 + Q11 + ASPECTS.

These component scores were then used in the main manuscript to assign baseline mRS state.

Appendix 4C: baseline characteristics of ALIAS subjects included and excluded from MSMM analysis

In this appendix, Table 4.1C is presented with baseline characteristics of the 17 ALIAS subjects excluded from the analysis because of withdrawn consent or lost-to-follow-up prior to the 30 day visit in combination with all mRS scores missing post-baseline.

Table 4.1C. Baseline Characteristics of the Subjects.			
Characteristic		Included (N = 1258)	Excluded (N = 17)
Age [mean (SD)]		66.1 (13.6)	62.6 (15.3)
Male sex [n (%)]		674 (53.6)	10 (58.8)
Race [n (%)]	White	969 (77.0)	13 (76.5)
	Black	197 (15.7)	3 (17.7)
	Asian	60 (4.8)	0 (0.0)
	American Indian/Alaska Native/First Nations People	7 (0.6)	0 (0.0)
	Native Hawaiian or Pacific islander	3 (0.2)	0 (0.0)
	Multiple, Other, or Unknown	1 (0.1)	0 (0.0)
Ethnic group [n (%)]	Non-Hispanic/Latino	1147 (91.2)	15 (88.2)
	Hispanic/Latino	60 (4.8)	2 (11.8)
	Unknown	51 (4.1)	0 (0.0)
Medical history [n (%)]	Hypertension	911 (72.4)	10 (58.8)
	Atrial fibrillation	257 (20.4)	2 (11.8)
	Past congestive heart failure	55 (4.4)	1 (5.9)
	Past myocardial infarction	155 (12.3)	5 (29.4)
	Past stroke	238 (18.9)	6 (35.3)
	Past transient ischaemic attack	157 (12.5)	2 (11.8)
	Diabetes mellitus	261 (20.8)	4 (23.5)
	Hyperlipidemia	554 (44.0)	8 (47.1)
	Peripheral vascular disease	75 (6.0)	1 (5.9)
Baseline NIHSS score [median, (inter-quartile range)]		11 (8 - 17)	9 (7 – 13)
Baseline ASPECTS score	Baseline ASPECTS > 7 [n/N, (%)]	932/1238 (75.3)	15/16 (93.8)
Clinical findings [mean (SD)]	Systolic blood pressure, mm Hg	156.9 (29.0)	157.3 (31.3)
	Plasma glucose, mmol/L	7.4 (3.2)	7.9 (3.3)
	Creatinine, µmol/L	90.2 (25.9)	92.7 (27.4)

5 Original Manuscript 3

Title: A comparison of multistate Markov modeling with contemporary outcome analysis in acute stroke trial data

Authors: Christy Cassarly[1], Renee' Martin[1], Marc Chimowitz[2], Edsel Peña[3], Viswanathan Ramakrishnan[1], Yuko Palesch[1]

Affiliation:

[1] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC

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

[3] Department of Statistics, University of South Carolina, Columbia, SC

Submission Status: drafted for submission to *Stroke*

Abstract

Background and Purpose – Historically, ordinal measures of functional outcome have been dichotomized for the primary analysis in trials of acute stroke therapy. A number of alternative methods to analyze the ordinal scales have been proposed, with an emphasis on maintaining the ordinal structure as much as possible. In addition, despite the availability of longitudinal outcome data in many trials, the primary analysis consists of a single endpoint. Inclusion of information about the course of disease progression allows for a more complete understanding of the treatment effect.

Methods – Multistate Markov modeling, which allows for the full ordinal scale to be analyzed longitudinally, is compared with previously suggested analytic techniques for

the ordinal modified Rankin Scale (dichotomous-logistic regression; continuous-linear regression; ordinal- shift analysis, proportional odds model, partial proportional odds model, adjacent categories logit model; sliding dichotomy; utility weights; repeated measures). Each of the methods is used to re-analyze the National Institute of Neurological Diseases and Stroke tissue plasminogen activator study.

Results – All methods detected a statistically significant treatment effect except the multistate Markov model without predicted baseline (p=0.053). The multistate Markov model allows for a more refined examination of treatment effect and describes the movement between modified Rankin Scale states over time which may provide more clinical insight into the treatment effect.

Conclusions – Multistate Markov models are feasible and desirable in describing treatment effect in acute stroke therapy trials. Future trials could increase power to detect a treatment effect using these models by collecting the outcome more frequently.

Keywords

acute stroke; outcomes; randomized controlled trials; statistical analysis

5.1 Introduction

A number of potential explanations for the failure of most acute stroke therapy trials to show efficacy have been discussed, including differences in preclinical and clinical models, inappropriate inclusion criteria, and poor methodological and statistical standards [6]. Specifically, there has been a recent emphasis on exploring alternative outcomes and analytic methods for stroke therapy trials.

The modified Rankin Scale (mRS) is the most commonly chosen primary outcome measure in clinical trials of acute stroke therapy [2]. Despite the ordinality of the outcome measure, many trials have dichotomized the mRS for the primary analysis [7]. In general, ignoring these differences and dichotomizing does not allow for examination of the treatment effect at finer gradients of the scale and can result in a loss of statistical power [1]. Any reduction in power may result in failure to find a clinically meaningful treatment effect during analysis of the data. The mRS should be analyzed in such a way that maintains the original structure of the scale as much as possible, using continuous or ordinal approaches [1, 3].

A number of alternative methods have been proposed to improve the analysis of the mRS. Some trials have analyzed the mRS as a continuous outcome, utilizing t-tests or linear regression [71]. Other trials have used the Cochran-Mantel Haenszel (CMH) shift test to analyze the distribution of the mRS, where the primary outcome is a favorable shift toward better functional outcome [7]. Ordinal logistic regression has also been proposed and applied in re-analysis of stroke trial data [46]. The proportional odds model (POM) has been used but the test for the proportional odds assumption is not well-

powered. In cases where the assumption was not justifiable, the partial proportion odds model (PPOM) or the adjacent categories logit (ACAT) has been used [46]. A popular alternative to continuous, ordinal and strict dichotomous analysis is responder analysis or the sliding dichotomy, where the definition of success is allowed to vary depending on baseline severity [48]. Most recently, a utility weighted mRS (UW-mRS) was derived to provide a patient centered metric of the degree of benefit or harm of a treatment that can be analyzed with a t-test or linear regression [53, 98].

A drawback of the outcome measures and analytic strategies listed above is that each analyzes data from a single endpoint, commonly the 90 day outcome, for the primary analysis despite the availability of repeated response measures collected over the course of longitudinal follow-up. Inclusion of information about the course of disease progression, using the longitudinal data, allows for a more comprehensive understanding of the benefit of a treatment [99]. None of the previously mentioned methods have utilized the repeated measures data. A literature search for repeated measures analysis of acute stroke trial data returned only two articles where a generalized estimating equations approach was used for repeated measures analysis of the mRS [11, 58].

Most recently, the Multistate Markov model (MSMM) was proposed for analysis of the mRS [93, 100]. The MSMM analyzes repeated measures data with ordinal outcomes. These types of models describe how a subject moves between a series of disease states over time, which is desirable in the description of disease processes that naturally move through increasing stages of severity [59]. Results suggest that the MSMM can be a more efficient approach than dichotomized methods used to analyze the

mRS data in some scenarios [93]. MSMMs can provide a better clinical understanding of the disease process since the information from the entire course of the disease is used to estimate the parameters of the model.

The purpose of this article is to demonstrate the MSMM as an approach for analysis of the mRS. The MSMM and the alternative methods listed above will be used to re-analyze the National Institute of Neurological Disorders and Stroke (NINDS) tissue plasminogen activator (t-PA) trial data. The results from each of the analytical analysis approaches will be compared with the results using the MSMM approach.

5.2 Materials and Methods

5.2.1 Trial Data

The seminal NINDS t-PA trial showed a consistently significant effect of t-PA using a global test of four outcomes (Barthel Index, mRS, Glasgow Outcome Scale and NIHSS) in the analysis of the primary outcome at 90 days post-stroke [39]. In addition to the 90 day primary outcome assessment, the mRS was also collected at 7-10 days, 180 days and 360 days from randomization.

Acute stroke requires immediate attention and treatment, posing a challenge to assess baseline outcome measures for clinical trials. Thus, the mRS is not obtainable at baseline and most often analysis is adjusted for baseline severity using the NIHSS [80]. Much of the progression or recovery experienced by a patient suffering from an acute onset disease is expected to occur early on. Moreover, typically, the goal of a treatment or therapeutic action is improvement in patient health compared to their baseline measure. To accurately quantify improvement, a measure of the outcome at baseline is ideal.

A prediction procedure using principal component analysis (PCA) for data reduction of baseline variables known to be correlated with functional ability was previously described [100]. Briefly, PCA is a statistical data reduction method that, when applied to a large set of variables, can group correlated variables into a smaller set of important composite variables, or components. The PCA grouped measures of severity (individual items of the NIHSS) and other baseline variables known to be associated with functional ability (age, baseline glucose, time from stroke onset to randomization and the Alberta Stroke Program Early CT score) into components to calculate component scores. The resulting component scores were used to assign the latent baseline mRS score and thus creating a comparable baseline mRS for analysis purposes.

5.2.2 Multistate Markov Models

In this paper, continuous-time MSMMs are used to describe the progression and recovery between mRS levels, or the disease states, over time. The main assumption of the MSMM is that the probabilities governing the transition between states only depend on the current state occupied by an individual, and not on previous disease history.

Death (mRS = 6) is known as an absorbing state because transitions out of this state cannot occur and mRS scores of 0 to 5 are examples of transient states, where transitions are allowed between the states. The data from the NINDS t-PA trial were observed at arbitrary times that were specified in advance so exact times of state transitions are unknown. Data of this type, observations of a continuous process at discrete times, are called panel data. Because the underlying disease process is continuous, where progression or recovery can happen at any time, it is assumed that in

order for a subject to transition from one state to a non-adjacent state they also transition through the intermediate states [4]. Thus, the general MSMM for panel data only estimates adjacent state transitions and transitions to death from any state. The allowable transitions between transient and absorbing states for the general model of the mRS are illustrated in Figure 5.1, where arrows indicate the allowed transitions between states.

MSMMs of panel data are governed by transition intensities that depend on time and individual level or time-dependent covariates. The transition intensities represent the instantaneous risk of transition between two mRS scores. Commonly, the transition intensities are assumed to be constant over time but this is often an unrealistic assumption. If the assumption fails, a model with piecewise-constant transition intensities can be used. This allows for the transition intensity matrices to change at breakpoints, remaining constant between the breakpoints. In addition to transition intensities, transition probabilities can also be estimated based on the observed transition rates using maximum likelihood estimation [4]. When modeling covariates in a MSMM, hazard ratios can be estimated that correspond to the effect of a covariate on the transition intensities.

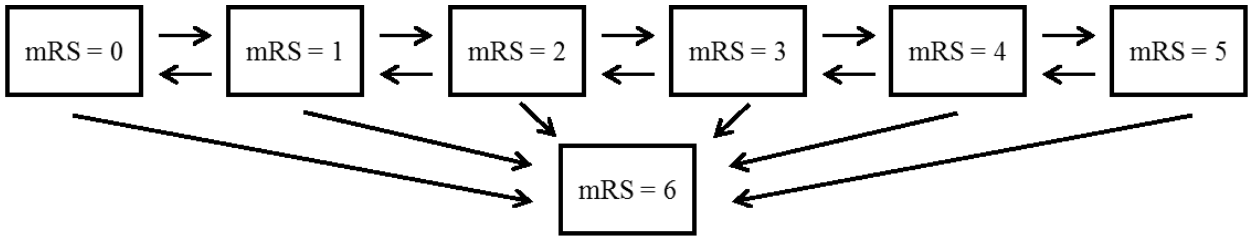


Figure 5.1. General MSMM for panel observed mRS data.

Likelihood ratio test (LRT) statistics are used to compare nested MSMMs. A reduced model is nested in a more complex model if all of the terms in the smaller model occur in the larger model. If two nested models are compared and the test is significant, then the more complex model fits the data better than the reduced model. LRTs are used to determine the significance of covariates and to compare models with constant transition intensities to ones with piecewise-constant intensities.

5.2.3 Statistical Analysis and Assumptions

For analyses using a fixed dichotomy, favorable outcome was defined as mRS ≤ 1, as was done in the primary paper [38]. The PPOM includes an additional term using a second parameter that allows for the ORs to increase proportional to the outcome scale. This PPOM, the restricted PPOM, is used when there is a linear deviation from the proportional odds assumption required for the POM, which is true of the NINDS t-PA data [46]. For the sliding dichotomy analysis, favorable outcome was defined to be consistent with previous re-analysis of the NINDS t-PA data where mRS = 0 for subjects with mild stroke (NIHSS < 7), mRS ≤ 1 for subjects with moderate stroke (NIHSS = 8-14) and mRS ≤ 2 (NIHSS > 14) [101]. The UW-mRS values were derived by averaging patient centered and person-tradeoff studies and are reported by Chaisinanunkul et al [53].

The ACAT and MSMMs were fit in R statistical software version 3.3.0 using the *VGAM* and *msm* packages, respectively. All other analysis was completed in SAS 9.4. When appropriate, analyses were adjusted for baseline NIHSS, which is known to be highly predictive of outcome [92]. The model using responder analysis as well as the

MSMM with predicted baseline mRS did not include baseline NIHSS because baseline severity is already accounted for. The shift analysis was also not adjusted for baseline severity as the test does not accommodate continuous covariates. Shift analysis for the NINDS t-PA data was previously repeated for different stratifications of the NIHSS and the results are reported elsewhere [7].

5.3 Results

The analysis presented in this section is based on 619 subjects that had mRS scores recorded at 90 days. The raw 90 day mRS outcome distributions for the placebo and t-PA groups are presented in Table 5.1. There are slight differences in the results presented in this section compared with other re-analyses of the trial because the raw observed values are used rather than the intent to treat imputation.

Table 5.1. NINDS t-PA 90 day mRS Counts (%).

	0	1	2	3	4	5	6	Total
Control	33 (5.3)	50 (8.1)	37 (6.0)	45 (7.3)	61 (9.9)	21 (3.4)	63 (10.2)	310
Treatment	57 (9.2)	74 (12.0)	23 (3.7)	40 (6.5)	42 (6.8)	19 (3.1)	54 (8.7)	309
Total	90	124	60	85	103	40	117	

Figures 5.2 and 5.3 display mRS scores over time for the control and t-PA groups, respectively. These plots, called Sankey plots, show the percentage of subjects with each mRS score at each follow-up visit as well as the change in the number of subjects with each score over time [74]. The longitudinal bar chart shows the percentage of subjects with each mRS score at each visit. The bands connecting the bars, or the links, represent

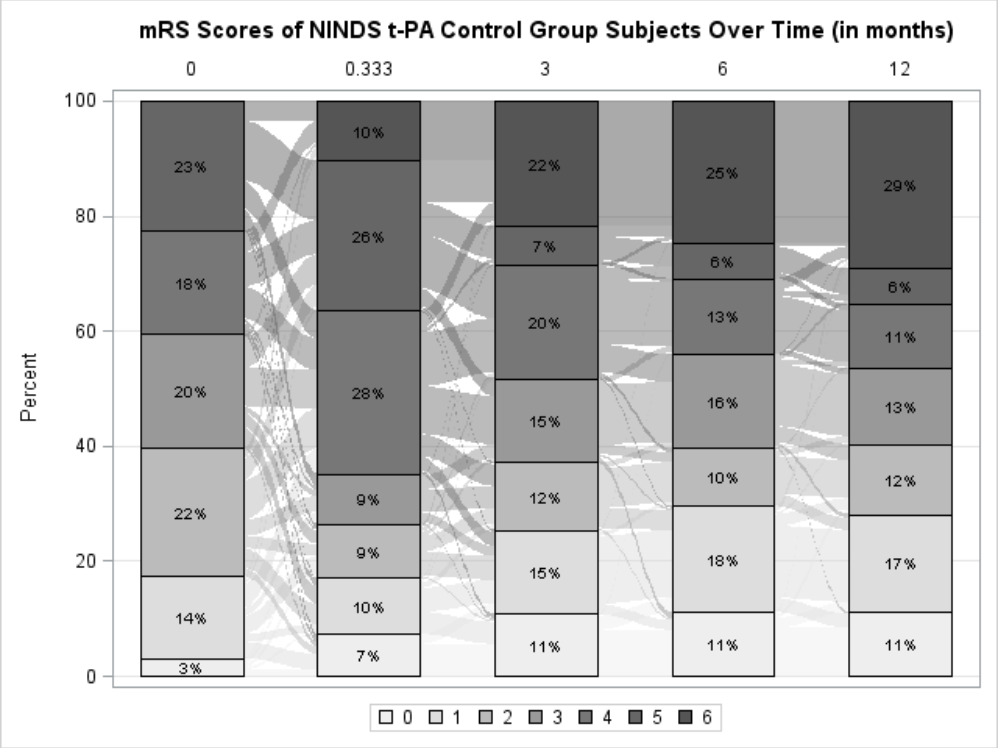


Figure 5.2. Sankey plot of NINDS t-PA control group subjects mRS scores over time.

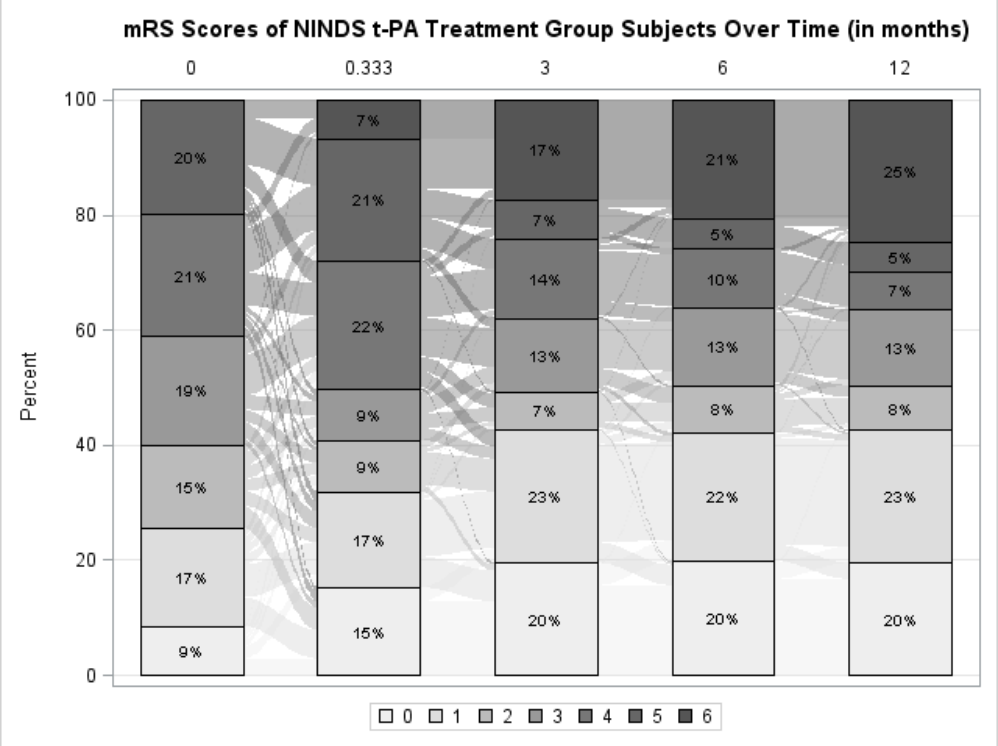


Figure 5.3. Sankey plot of NINDS t-PA treatment group subjects mRS scores over time.

the change in the number of subjects in each state, over time. A thicker link is indicative of more subjects transitioning between the two states. In the 90 day mRS alone, there are differences between the groups across the entire ordinal scale that are ignored in a traditional dichotomized analysis. The use of one follow-up visit also results in a loss of information as there are differences in the distribution of the mRS as well as the transition rates over the entire follow-up period. Additionally, the inclusion of the predicted baseline mRS allows one to observe the differences in the transition rates between treatment groups in the crucial window immediately following randomization and during the acute treatment phase. All of these differences can be measured and described using MSMMs and are not accounted for using other ordinal data analysis methods.

In the general MSMM (Figure 5.1), some of the parameters estimated were close to zero. Specifically, the transition intensities to death from mRS = $\{0,1,2,3\}$ were all very small. When there is not enough information from the data on certain transition rates, more intensities may need to be set to zero [4]. Thus, the general model was reduced, no longer allowing death from any state. Constraints were imposed such that death is only allowable from mRS = 4 or mRS = 5.

The results from all methods are presented in Table 5.2. The results are consistent with previously reported re-analyses of the NINDS t-PA data with minor, insignificant differences in estimates due to the adjustment for the NIHSS and the use of the raw mRS data versus intent to treat [11, 46, 53, 94, 102]. Table 5.3 presents a review of the interpretation of the summary statistics obtained from each of the methods of analysis.

Table 5.2. Results from previously used methods for analysis of the mRS.				
Method	Outcome Measure	Summary Statistic	(95% CI)	p
Logistic regression	mRS at 90 d (0-1 vs. 2-6)	OR = 2.04	(1.39, 2.99)	0.0003
Linear regression	mRS at 90 d (continuous)	Diff. in means = 0.50	-	0.0073
Shift analysis	mRS at 90 d	-	-	0.0017
POM	mRS at 90 d	OR = 1.41	(1.01, 1.81)	0.0172
PPOM (linear trend)	mRS at 90 d 1-6 vs. 0 2-6 vs. 0-1 3-6 vs. 0-2 4-6 vs. 0-3 5-6 vs. 0-4 6 vs. 0-5	OR = 1.88 1.67 1.48 1.31 1.16 1.03	 (1.14, 2.61) (1.12, 2.21) (1.05, 1.90) (0.93, 1.69) (0.77, 1.55) (0.61, 1.45)	0.0017
ACAT	mRS at 90 d 1 vs. 0 2 vs. 1 3 vs. 2 4 vs. 3 5 vs. 4 6 vs. 5	OR = 1.12 2.35 0.70 1.30 0.79 1.08	 (0.64, 1.97) (1.25, 4.44) (0.36, 1.38) (0.73, 2.32) (0.38, 1.66) (0.52, 2.21)	0.0163
Logistic regression of sliding dichotomy	mRS at 90 d (0 if NIHSS is 1-7, 0-1 if 8-14 and 0-2 if >14)	OR = 1.61	(1.13, 2.28)	0.0080
Linear regression of UW-mRS	UW-mRS at 90d	Diff. in means = 0.08	-	0.0175
Repeated measures GEE	mRS at 7-10, 90, 180 and 360 d (0-1 vs. 2-6)	OR = 1.89	(1.36, 2.63)	0.0002
Repeated measures GEE (with baseline)	Predicted mRS at baseline and mRS at 7-10, 90, 180 and 360 d (0-1 vs. 2-6)	OR = 1.78	(1.33, 2.38)	0.0001
MSMM (without baseline)	mRS at 7-10, 90, 180 and 360 d 0→1 1→2 2→3 3→4 4→5 4→6 5→6 1→0 2→1 3→2 4→3 5→4	Hazard Ratio = 0.72 0.46 3.04 0.71 0.90 0.98 1.69 0.99 1.03 1.58 0.99 0.58	 (0.40, 1.30) (0.23, 0.93) (0.98, 9.41) (0.34, 1.49) (0.36, 2.23) (0.50, 1.91) (0.97, 2.95) (0.60, 1.64) (0.63, 1.70) (0.64, 3.92) (0.64, 1.53) (0.32, 1.05)	0.0533

Piecewise MSMM (with baseline)	Predicted mRS at baseline and mRS 7-10, 90, 180 and 360 d	Hazard Ratio =		0.0018
	0→1	0.73	(0.43, 1.23)	
	1→2	0.51	(0.28, 0.90)	
	2→3	1.29	(0.66, 2.52)	
	3→4	0.67	(0.38, 1.17)	
	4→5	0.88	(0.53, 1.48)	
	4→6	0.96	(0.55, 1.66)	
	5→6	1.20	(0.75, 1.92)	
	1→0	1.00	(0.62, 1.62)	
	2→1	1.14	(0.71, 1.83)	
	3→2	0.99	(0.59, 1.68)	
	4→3	1.09	(0.72, 1.65)	
	5→4	0.80	(0.52, 1.23)	

The results of the MSMM are presented as hazard ratios that estimate the effect of the covariate on transition intensities. A hazard ratio above one signifies a positive association between treatment and the rate of transition, whereas a hazard ratio of one implies no effect.

In the MSMM with baseline mRS, treatment with t-PA significantly reduced the transition intensity between mRS = 1 and mRS = 2 with a hazard ratio of 0.51 (95% CI: 0.28, 0.90). None of the other hazard ratios were significantly different from one. This finding is consistent with the results of the ACAT model where the only significant OR is the one comparing mRS category 2 to mRS category 1. The conclusion drawn from the ACAT is that the most relevant impact of t-PA is to reduce the odds of observing a category 2 versus a category 1 at 90 days [46]. The results from the MSMM allow for a more refined conclusion- the most relevant impact of t-PA is to reduce the hazard of transitioning from mRS category 1 to mRS category 2. Therefore, the t-PA is more protective of worsening from category 1 rather than promoting improvement from category 2, which is a distinction that cannot be made from the ACAT results.

Table 5.3. Summary of statistics obtained from each type of analysis of the mRS.		
Method	Statistic(s)	Interpretation
Logistic regression	OR	The odds of good outcome in the treatment group versus placebo
Linear regression	Difference of means	Improvement of the average mRS score in patients that received treatment
Shift analysis	Probability value (no effect size or OR)	The treatment group shifted in a favorable direction toward a better mRS score versus placebo
POM	Summary odds ratio	The odds of a lower mRS the treatment group versus placebo
PPOM	ORs for six possible dichotomizations of mRS	Treatment has a significant benefit for certain definitions of good outcome
ACAT	ORs the six adjacent categories of response	The treatment group is more likely to have smaller mRS for certain adjacent mRS scores
Logistic regression of sliding dichotomy	OR	The odds of good outcome (defined by baseline severity) in the treatment group versus placebo
Linear regression of UW-mRS	Difference of mean utility scores	Improvement of the average utility score in patients that received treatment
Repeated measures GEE (dichotomized)	OR	The odds of good outcome over the 12-month period in the treatment group versus placebo
MSMM	Hazard ratios for each allowable transition	The hazard (instantaneous risk) of transitioning from one mRS state to another in the treatment group versus placebo

5.4 Discussion

It is not realistic to choose one analytic method that is most appropriate for the mRS for all studies because the efficiency varies depending on the expected distribution of the treatment effect [3]. In general, ordinal approaches are more efficient when treatment effects are distributed over the entire outcome range or when the distribution of treatment effect could not be prespecified [3]. Therefore, it is important to know what the expected

result of intervention is in the design and sample size calculation stage of a trial. In comparison, dichotomous approaches are more efficient than ordinal approaches when treatment effects cluster at single-state transitions and can be specified in advance [102]. However, it is uncommon for clustering to be predictable. If the clustering cannot be predicted, an ordinal approach should be used.

In this paper, the dichotomized methods were found to be most statistically efficient with respect to power for the NINDS t-PA trial, and inclusion of predicted baseline mRS improved the ability to detect a treatment effect in the repeated measures analysis. The treatment effect clustered at the transition from mRS category 1 to mRS category 2. If limited information were available in the planning stages for this trial to confidently predict that the treatment effect would be clustered at that transition, it would have been worthwhile to consider an ordinal approach. Acute stroke trials are challenging to conduct as there are few centers that can recruit many patients in the early time window required for treatment [2]. Because of the low recruitment rate and cost associated with conducting acute stroke trials, inefficient statistical tests must be avoided to protect from being underpowered.

Of the approaches that do not rely on the strict dichotomy, the PPOM and MSMM with predicted baseline were the most efficient. The PPOM and MSMM were found to be more efficient than linear regression, responder analysis, shift analysis and the UW-mRS for analysis of the NINDS t-PA data. The PPOM is represented by ORs for the six possible dichotomizations of the mRS. The first three ORs are significantly different

from one indicating that treatment has a significant benefit whether 0, 0-1 or 0-2 is defined as good outcome.

Construction of MSMMs provides a more comprehensive view of the disease process and allow for exploration of how covariates affect the movement of the process. The obvious benefit to using the MSMM is the ability to handle progression and recovery simultaneously by estimating transition rates for both. Because of this, the MSMM allows for identification of where the treatment effect has the greater impact. Here the effect of treatment was greatest in reducing the hazard of transitioning from mRS category 1 to mRS category 2. A more clear understanding of the effect of treatment could also be beneficial in identifying characteristics of subjects that are more likely to benefit or experience harm from a therapy.

Another benefit of MSMMs is the potential for decreased sample size. The power of MSMMs applied to acute onset clinical trial data was shown to increase significantly when the number of follow-up visits was increased [93]. Future trials could collect the mRS more frequently, increasing the power to detect differences using this modeling technique. This would be a more cost-effective than recruiting more subjects to increase power as the telephone assessment of stroke disability with the mRS is reliable in comparison with a face-to-face assessment [103].

The MSMM results in a more comprehensive understanding of treatment effect; however it also increases the difficultly to determine the sample size to adequately power a study using this analysis. Without a summary statistic of effect size, the implementation of these models in the analysis of the primary outcome in trials requires quite a bit of

foresight on the expected distribution of the effect of a therapy or treatment. However, once the distribution of expected treatment effect has been specified simulation-based power analysis for these models is straightforward. Another limitation of MSMMs is the increasingly computationally intensive nature as covariates and time-varying intensities are added to the models.

Future directions of this research may include development of a software package to automate the baseline mRS prediction. The package could include more complex methods for estimation, potentially Bayesian PCA. Another feature of the package could be assistance with data manipulation required to use the *msm* package in R to fit the MSMMs (eg. wide to long format and incorporating exact time of death). In addition, a package could be developed to streamline the simulation-based power analyses.

Acknowledgements

This work was supported by National Institute of Health/NINDS Grant Number: U01NS06186105.

Disclosures

None.

6 Overall Discussion

6.1 Specific Aims Revisited

The aims of this dissertation are:

1. To explore the operating characteristics (type I error and power) of MSMMs compared with repeated logistic regression used to analyze sparsely populated repeated measures ordinal data.
2. To develop a MSMM approach with piecewise-constant transition intensities that incorporates a latent baseline state.
3. Analyze acute stroke therapy trial data using the methods developed in Aim 2 and compare the results with those from alternative methods previously suggested for the analysis of the mRS.

6.2 Summary and Conclusions

This work focuses on the use of MSMMs to analyze sparsely populated and longitudinally collected ordinal data. The mRS score from acute stroke therapy trials was the motivating example. To determine whether MSMMs were feasible as an analytic method for sparsely populated ordinal data, the operating characteristics are investigated using simulation studies in Aim 1. Results indicate that MSMMs can be a more efficient approach than repeated measures logistic regression to analyze sparsely populated ordinal data. There are also situations where dichotomization might not lose efficiency and may be more powerful than the MSMM. Depending on the observed data structure and treatment effect distribution, either method could be more powerful. Results also show

that increasing the number of follow-up visits can dramatically improve power to detect a treatment difference. Thus, we recommend that future acute stroke therapy trials collect the mRS more frequently, increasing power to detect treatment group differences. Increasing the frequency of outcome collection could be more cost-effective than recruiting more subjects since the telephone assessment of the mRS is as reliable as the face-to-face assessment [103].

Given that the MSMM is an approach that could realistically analyze sparsely populated ordinal data, a latent baseline estimation procedure is developed in Aim 2. Methods that analyze only one time point have traditionally adjusted for baseline severity using the NIHSS score because the mRS score is not available at baseline. When modeling data longitudinally, the transition from baseline to first follow-up is important because much of the progression or recovery experienced by a patient suffering from an acute onset disease is expected to occur early on. Inclusion of the latent baseline in a piecewise-constant MSMM improves efficiency to detect a treatment effect as compared to the MSMM without baseline that adjusted for baseline NIHSS.

In the application to the NINDS t-PA trial in Aim 3, the MSMM with baseline has proven to be an efficient method of analysis, as compared to many of the other popular methods for the mRS. While dichotomized analysis is the most powerful for this particular data set, for most trials, prediction of the clustering of treatment effect *a priori* is not realistic. If the treatment effect clustering is predicted incorrectly, the dichotomized statistical test becomes inefficient. For trials where the treatment effect is expected to be distributed over the entire range of the outcome or when the clustering cannot be

predicted, methods that analyze the full ordinal scale should be used [3]. The MSMM with estimated baseline is comparable in efficiency to the PPOM and the two methods outperform the other ordinal methods. The MSMM allows for direct identification of where the treatment effect is most significant and the PPOM does not.

The goal of this dissertation is to lay the foundation for the use of MSMMs in practice to analyze ordinal data, specifically data from acute stroke therapy trials. We conclude, from the example presented, that the MSMM with latent baseline mRS is as efficient, if not more, than other methods currently used to analyze acute stroke therapy trial data. A limitation of this work is that comparison of efficiency of other methods is only done for data from one trial. The "best" method for analysis of the mRS will change depending on the distribution of the treatment effect. Thus, future work should consider data from other trials to better understand the comparative efficiency of the MSMM.

This work supports the use of MSMMs for acute stroke therapy trial but immediate implementation of these models for the primary analysis of new studies is likely not feasible because of lack of readily available software to design a study that uses MSMM analysis. The MSMM is a great tool to identify the finer details of the treatment effect but the complexity of the model makes determination of sample size needed to be adequately powered to detect a treatment effect difficult. Future work would explore how to appropriately power a new study using MSMM. Without a summary statistic of effect size, the implementation of these models in the analysis of the primary outcome in trials requires foresight on the expected distribution of the effect of a therapy or treatment.

However, once the distribution of expected treatment effect has been specified simulation-based power analysis for these models is straightforward.

Additional future directions of this research include development of a software package to automate the baseline mRS estimation. Another feature of the package could be assistance with data manipulation required to use the *msm* package in R to fit the MSMMs. Finally, the package could also contain functions to streamline the simulation-based power analysis that was used in Aim 1.

List of References

1. The Optimising Analysis of Stroke Trials (OAST) Collaboration. Can We Improve the Statistical Analysis of Stroke Trials? Statistical Reanalysis of Functional Outcomes in Stroke Trials. *Stroke* **2007; 38(6)**: pp 1911-1915, DOI 10.1161/strokeaha.106.474080.
2. Saver JL. Optimal End Points for Acute Stroke Therapy Trials: Best Ways to Measure Treatment Effects of Drugs and Devices. *Stroke* **2011; 42(8)**: pp 2356-2362, DOI 10.1161/strokeaha.111.619122.
3. Bath PMW, Lees KR, Schellinger PD, et al. Statistical Analysis of the Primary Outcome in Acute Stroke Trials. *Stroke* **2012; 43(4)**: pp 1171-1178, DOI 10.1161/strokeaha.111.641456.
4. Jackson C. Multi-State Models for Panel Data: The msm Package for R. *J Stat Softw* **2011; 38(8)**: pp 28, DOI 10.18637/jss.v038.i08.
5. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart Disease and Stroke Statistics—2012 Update. *A Report From the American Heart Association* **2012; 125(1)**: pp e2-e220, DOI 10.1161/CIR.0b013e31823ac046.
6. Mergenthaler P, Meisel A. Do Stroke Models Model Stroke? *Dis Model Mech* **2012; 5(6)**: pp 718-725, DOI 10.1242/dmm.010033.
7. Savitz SI, Lew R, Bluhmki E, et al. Shift Analysis Versus Dichotomization of the Modified Rankin Scale Outcome Scores in the NINDS and ECASS-II Trials. *Stroke* **2007; 38(12)**: pp 3205-3212, DOI 10.1161/strokeaha.107.489351.

8. Keller DM. IMS-III: Endovascular Terapy Benefits Severe Stroke, 2014. http://www.medscape.com/viewarticle/834005.

9. Price M, Hertzberg V, Wright DW. Does the sliding dichotomy result in higher powered clinical trials for stroke and traumatic brain injury research? *Clinical Trials* **2013; 10(6)**: pp 924-934, DOI 10.1177/1740774512458601.

10. Committee for Proprietary Medicinal Products. Points to Consider on Clinical Investigation of Medicinal Products for the Treatment of Acute Stroke. *The European Agency for the Evaluation of Medicinal Products* **2001**.

11. Feng W, Vasquez G, Suri MFK, et al. Repeated-Measures Analysis of the National Institute of Neurological Disorders and Stroke rt-PA Stroke Trial. *J Stroke Cerebrovasc Dis* **2011; 20(3)**: pp 241-246, DOI 10.1016/j.jstrokecerebrovasdis.2010.01.003.

12. Aalen OO. Armitage lecture 2010: Understanding treatment effects: the value of integrating longitudinal data and survival analysis. *Stat Med* **2012; 31(18)**: pp 1903-1917, DOI 10.1002/sim.5324.

13. Alessandrino EP, Della Porta MG, Malcovati L, et al. Optimal timing of allogeneic hematopoietic stem cell transplantation in patients with myelodysplastic syndrome. *American Journal of Hematology* **2013; 88(7)**: pp 581-588, DOI 10.1002/ajh.23458.

14. Allen EJ, Farewell VT. Pragmatic Analysis of Longitudinal Data on Disease Activity in Systemic Lupus Erythematosus. *Communications in Statistics- Theory and Methods* **2009; 38(18)**: pp 3369-3388, DOI 10.1080/03610920902972335.

15. Batina NG, Crnich CJ, Anderson DF, et al. Models to predict prevalence and transition dynamics of methicillin-resistant Staphylococcus aureus in community nursing homes *American Journal of Infection Control* **2016; 5(1)**: pp 507-514, DOI 10.1016/j.ajic.2015.10.042.

16. Cao Q, Postmus D, Hillege HL, et al. Probability Elicitation to Inform Early Health Economic Evaluations of New Medical Technologies: A Case Study in Heart Failure Disease Management. *Value in Health* **2013; 16(4)**: pp 529-536, DOI 10.1016/j.jval.2013.02.008

17. Chauvel O, Lacombe K, Bonnard P, et al. Risk factors for acute liver enzyme abnormalities in HIV-hepatitis B virus-coinfected patients on antiretroviral therapy. *Antiviral Therapy* **2007; 12(7)**: pp 1115-1126.

18. Chui M. Nonparametric Simultaneous Modelling of Operative Mortality and Long-Term Survival after Coronary Artery Bypass Surgery. *Simon Fraser University* **2002**, DOI 1541561615.

19. Combescure C, Chanez P, Daures J-P, et al. Assessment of variations in control of asthma over time. *Eur Respir J* **2003; 22**: pp 298-304, DOI 10.1183/09031936.03.00081102.

20. Elbasha EE, Szucs T, Chaudhary MA, et al. Cost-effectiveness of Raltegravir in Antiretroviral Treatment-Experienced HIV-1–Infected Patients in Switzerland. *HIV Clinical Trials* **2009; 10(4)**: pp 233-253, DOI 10.1310/hct1004-233.

21. Gangnon RE, Lee KE, Klein BK, et al. Effect of the y402h variant in the complement factor h gene on the incidence and progression of age-related

macular degeneration: Results from multistate models applied to the beaver dam eye study. *Archives of Ophthalmology* **2012; 130(9)**: pp 1169-1176, DOI 10.1001/archophthalmol.2012.693.

22. Garcia ME, Lee A, Neuhaus J, et al. Diabetes Mellitus as a Risk Factor for Development of Depressive Symptoms in a Population-Based Cohort of Older Mexican Americans. *Journal of the American Geriatrics Society* **2016; 64(3)**: pp 619-624, DOI 10.1111/jgs.14019.

23. Haeussler K, van den Hout A, Baio G. A dynamic Bayesian Markov model for health economic evaluations of interventions against infectious diseases. *Cornell University Library* **2016**.

24. Hanly JG, Su L, Urowitz MB, et al. A Longitudinal Analysis of Outcomes of Lupus Nephritis in an International Inception Cohort Using a Multistate Model Approach. *Arthritis & Rheumatology* **2016; 68(8)**: pp 1932-1944, DOI 10.1002/art.39674.

25. Jackson DJ, Evans MD, Gangnon RE, et al. Evidence for a Causal Relationship between Allergic Sensitization and Rhinovirus Wheezing in Early Life. *American Journal of Respiratory and Critical Care Medicine* **2012; 185(3)**: pp 281-285, DOI 10.1164/rccm.201104-0660OC.

26. Jambarsang S, Baghban AA, Nazari SSH, et al. Effect of baseline CD4 count on efficacy of highly-active antiretroviral therapy for HIV patients. *International Journal of Biology, Pharmacy and Allied Sciences* **2015; e4(11)**: pp 700-713.

27. Joutard X, Paraponaris A, Teyssier LS, et al. Continuous-Time Markov Model for Transitions Between Employment and Non-Employment: The Impact of a Cancer Diagnosis. *Annals of Economics and Statistics* **2012; (107/108)**: pp 239-265, DOI 10.2307/23646578.

28. Liu W-J, Lee L-T, Yen M-F, et al. Assessing progression and efficacy of treatment for diabetic retinopathy following the proliferative pathway to blindness: implication for diabetic retionpathy screeing in Taiwan. *Diabetic Medicine* **2003; 20(9)**: pp 727-733, DOI 10.1046/j.1464-5491.2003.01019.x.

29. Ndumbi P, Gillis J, Raboud JM, et al. Clinical implications of altered T-cell homeostasis in treated HIV patients enrolled in a large observational cohort. *AIDS* **2013; 27(18)**: pp 2863-2872, DOI 10.1097/01.aids.0000432471.84497.

30. Nunez J, Nunez E, Bayes-Genis A, et al. Long-term serial kinetics of N-terminal pro B-type natriuretic peptide and carbohydrate antigen 125 for mortality risk prediction following acute heart failure. *European Heart Journal: Acute Cardiovascular Care* **2016**, DOI 10.1177/2048872616649757.

31. Raiche M, Hebert R, Dubois M-F, et al. Yearly transitions of disability profiles in older people living at home. *Archieves of Gerontology and Geriatrics* **2012; 55(2)**: pp 399-405, DOI 10.1016/j.archger.2011.12.007.

32. Raiche M, Hebert R, Dubois M-F, et al. Covariates of Disability-Profile Transitions in Older People Living at Home. *Journal of Biosciences and Medicines* **2014; 2(3)**: pp 25-36, DOI 10.4236/jbm.2014.23005.

33. Rodriguez-Girondo M, de Una-Alvarez J. A nonparametric test for Markovianity in the illness-death model. *Stat Med* **2012; 31(30)**: pp 4416-4427, DOI 10.1002/sim.5619.

34. Saint-Pierre P, Bourdain A, Chanez P, et al. Are overweight asthmatics more difficult to control? *Allergy* **2006; 61(1)**: pp 79-84, DOI 10.1111/j.1398-9995.2005.00953.

35. Tung T-H, Chen S-J, Shih H-C, et al. Assessing the Natural Course of Diabetic Retinopathy: A Population-Based Study in Kinmen, Taiwan. *Opthalmic Epidemiology* **2006; 13(5)**: pp 327-333, DOI 10.1080/09286580600826637.

36. Zhang S-K, Kang L-N, Chang IJ, et al. The Natural History of Cervical Cancer in Chinese Women: Results from an 11-Year Follow-Up Study in China Using a Multistate Model *Cancer Epidemiol Biomarkers Prev* **2014; 23(7)**: pp 1298-1305, DOI 10.1158/1055-9965.

37. Gangnon RE, Lee K, E, Klein BEK, et al. Misclassification can explain most apparent regression of age-related macular degeneration: results from multistate models with misclassification. *Invest Opthalmol Vis Sci* **2014; 55**: pp 1780-1786, DOI 10.1167/iovs.13-12375.

38. The National Institute of Neurological Disorders Stroke rt-PA Stroke Study Group. Tissue Plasminogen Activator for Acute Ischemic Stroke. *N Engl J Med* **1995; 333(24)**: pp 1581-1588, DOI 10.1056/nejm199512143332401.

39. Tilley BC, Marler J, Geller NL, et al. 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(11)**: pp 2136-2142, DOI 10.1161/01.str.27.11.2136.

40. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Md State Med J* **1965; 14**: pp 61-65.

41. Jennett B, Bond M. Assessment of outcome after severe brain damage. *Lancet* **1975; 305**: pp 480-484.

42. Ginsberg MD, Palesch YY, Martin RH, et al. The Albumin in Acute Stroke (ALIAS) Multicenter Clinical Trial. *Stroke* **2011; 42(1)**: pp 119-127, DOI 10.1161/strokeaha.110.596072.

43. Ginsberg MD, Palesch YY, Hill MD, et al. High-dose albumin treatment for acute ischaemic stroke (ALIAS) part 2: a randomised, double-blind, phase 3, placebo-controlled trial. *The Lancet Neurology* **2013; 12(11)**: pp 1049-1058, DOI 10.1016/S1474-4422(13)70223-0.

44. Lees KR, Davalos A, Davis SM, et al. Additional Outcomes and Subgroup Analyses of NXY-059 for Acute Ischemic Stroke in the SAINT I Trial. *Stroke* **2006; 37(12)**: pp 2970-2978, DOI 10.1161/01.str.0000249410.91473.44.

45. Agresti A. Modelling ordered categorical data: recent advances and future challenges. *Stat Med* **1999; 18(17-18)**: pp 2191-2207.

46. DeSantis SM, Lazaridis C, Palesch Y, et al. Regression Analysis of Ordinal Stroke Clinical Trial Outcomes: An Application to the NINDS t-PA Trial. *Int J Stroke* **2014; 9(2)**: pp 226-231, DOI 10.1111/ijs.12052.

47. Lall R, Campbell MJ, Walters SJ, et al. A review of ordinal regression models applied on health-related quality of life assessments. *Stat Methods Med Res* **2002; 11(1)**: pp 49-67, DOI 10.1191/0962280202sm271ra.

48. Garofolo KM, Yeatts SD, Ramakrishnan V, et al. The effect of covariate adjustment for baseline severity in acute stroke clinical trials with responder analysis outcomes. *Trials* **2013; 14**: pp 98-98, DOI 10.1186/1745-6215-14-98.

49. Garofolo KM. The Analysis of Acute Stroke Clinical Trials with Responder Analysis Outcomes. *Medical University of South Carolina* **2012**.

50. Young FB, Lees KR, Weir CJ. Improving Trial Power Through use of Prognosis-Adjusted End Points. *Stroke* **2005; 36(3)**: pp 597-601, DOI 10.1161/01.str.0000154856.42135.85.

51. Price M. Issues in Causal Inference and Applications to Public Health. *Emory* **2009**.

52. Lees KR, Bath PMW, Schellinger PD, et al. Contemporary Outcome Measures in Acute Stroke Research: Choice of Primary Outcome Measure. *Stroke* **2012; 43(4)**: pp 1163-1170, DOI 10.1161/strokeaha.111.641423.

53. Chaisinanunkul N, Adeoye O, Lewis RJ, et al. Adopting a Patient-Centered Approach to Primary Outcome Analysis of Acute Stroke Trials Using a Utility-Weighted Modified Rankin Scale. *Stroke* **2015; 46(8)**: pp 2238-2243, DOI 10.1161/strokeaha.114.008547.

54. Feeny D. Response to Lenert and Kaplan: A Utility Approach to the Assessment of Health-Related Quality of Life. *Medical Care* **2000; 38(9)**: pp 151-154.

55. Rivero-Arias O, Ouellet M, Gray A, et al. Mapping the Modified Rankin Scale (mRS) Measurement into the Generic EuroQol (EQ-5D) Health Outcome. *Medical Decision Making* **2010; 30(3)**: pp 341-354, DOI 10.1177/0272989X09349961.

56. Hong K-S, Saver JL. Quantifying the Value of Stroke Disability Outcomes. *Stroke* **2009; 40(12)**: pp 3828-3833, DOI 10.1161/strokeaha.109.561365.

57. Yang B. Analyzing Ordinal Repeated Measures Data Using SAS®. *Eli Lilly and Company* **2006**.

58. Palesch YY, Yeatts SD, Tomsick TA, et al. Twelve-Month Clinical and Quality-of-Life Outcomes in the Interventional Management of Stroke III Trial. *Stroke* **2015; 46(5)**: pp 1321-1327, DOI 10.1161/strokeaha.115.009180.

59. Jackson C. *Multi-state modelling with R: the* `msm` *package.* MRC Biostatistics Unit: Cambridge, U.K., 2016.

60. Gani R, Nixon RM, Hughes S, et al. Estimating the rates of disability progression in people with active relapsing-remitting multiple sclerosis. *Journal of Medical Economics* **2007; 10(2)**: pp 79-89, DOI 10.3111/200710079089.

61. Mdala I, Olsen I, Haffajee AD, et al. Comparing clinical attachment level and pocket depth for predicting periodontal disease progression in healthy sites of patients with chronic periodontis using multi-state Markov models. *J Clin Periodontol* **2014; 41(9)**: pp 837-845, DOI 10.1111/jcpe.12278.

62. Shirley KE, Small DS, Lynch KG, et al. Hidden Markov Models for Alcoholism Treatment Trial Data. *The Annals of Applied Statistics* **2010; 4(1)**: pp 366-395, DOI 10.1214/09-aoas282.

63. Gharoodi ZR, Ganjali M, Berridge D. A Transition Model for Ordinal Response Data with Random Dropout: An Application to the Fluvoxamine Data. *J Biopharm Stat* **2009; 19(4)**: pp 658-671, DOI 10.1080/10543400902964100.

64. Chiang CL. *An Introduction to Stochastic Processes and Their Applications.* Robert E. Krieger Publishing Co.: New York, 1980.

65. Titman AC. Model diagnostics in multi-state models of biological systems. In Model diagnostics in multi-state models of biological systems, Editor (ed)^(eds). Fitzwilliam College: City, 2007.

66. Jackson CH. Multi-state models for panel data: the msm package for R. *Journal of Statistical Software* **2011; 38(8)**: pp 1-29.

67. Anderson DF. Continuous Time Markov Chains, 2011. http://www.math.wisc.edu/~anderson/605F11/Notes/StochBioChapter6.pdf.

68. Kalbfleisch JD, Lawless JF. The Analysis of Panel Data Under a Markov Assumption. *J Am Stat Assoc* **1985; 80(392)**: pp 863-871.

69. Pérez-Ocón R, Ruiz-Castro JE, Gámiz-Pérez ML. Non-homogeneous Markov models in the analysis of survival after breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **2001; 50(1)**: pp 111-124, DOI 10.1111/1467-9876.00223.

70. Mandel M, Mercier F, Exkert B, et al. Estimating Time to Disease Progression Comparing Transition Models and Survival Models- an Analysis of Multiple Sclerosis Data. *Biometrics* **2013; 69(1)**: pp 225-234, DOI 10.1111/biom.12002.

71. Nunn A, Bath PMW, Gray LJ. Analysis of the Modified Rankin Scale in Randomised Controlled Trials of Acute Ischaemic Stroke: A Systematic Review. *Stroke Res Treat* **2016**, DOI 10.1155/2016/9482876.

72. Titman AC. Model diagnostics in multi-state models of biological systems. *Fitzwilliam College* **2007**.

73. Dai Y-H. A Perfect Example for the BFGS Method. *State Key Laboratory of Scientific and Engineering Computing* **1997**.

74. Rosanbalm S. Getting Sankey with Bar Charts. *PharmaSUG* **2015**.

75. Broderick JP, Palesch YY, Demchuk AM, et al. Endovascular Therapy after Intravenous t-PA versus t-PA Alone for Stroke. *N Engl J Med* **2013; 368(10)**: pp 893-903, DOI 10.1056/NEJMoa1214300.

76. Regnier ED, Shechter SM. State-space size considerations for disease-progresssion models. *Stat Med* **2013; 32(22)**: pp 3862-3880, DOI 10.1002/sim.5808.

77. Thom HH, Jackson CH, Commenges D, et al. State Selection in Markov Models for Panel Data with Application to Psoriatic Arthritis. *Stat Med* **2015; 34(16)**: pp 2456-2475, DOI 10.1002/sim.6460.

78. Roozenbeek B, Lingsma HF, Perel P, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Critical Care* **2011; 15(3)**, DOI 10.1186/cc10240.

79. Sankey SS, Weissfeld LA. A study of the effect of dichotomizing ordinal data upon modeling. *Communications in Statistics - Simulation and Computation* **1998; 27(4)**: pp 871-887, DOI 10.1080/03610919808813515.

80. Bruno A, Saha C, Williams LS. Percent Change on the National Institutes of Health Stroke Scale: A Useful Acute Stroke Outcome Measure. *J Stroke Cerebrovasc Dis* **2009; 18(1)**: pp 56-59, DOI 10.1016/j.jstrokecerebrovasdis.2008.09.002.

81. Alexandrov AV, Demchuk AM, Felberg RA, et al. High Rate of Complete Recanalization and Dramatic Clinical Recovery During tPA Infusion When Continuously Monitored With 2-MHz Transcranial Doppler Monitoring. *Stroke* **2000; 31(3)**: pp 610-614, DOI 10.1161/01.str.31.3.610.

82. Ginsberg MD, Palesch YY, Hill MD. The ALIAS (ALbumin In Acute Stroke) Phase III randomized multicentre clinical trial: design and progress report. *Biochemical Society Transactions* **2006; 34(6)**: pp 1323-1326, DOI 10.1042/bst0341323.

83. Lyden P, Brott T, Tilley B, et al. Improved reliability of the NIH Stroke Scale using video training. *Stroke* **1994; 25**: pp 2220-2226.

84. Hill MD, Martin RH, Palesch YY, et al. The Albumin in Acute Stroke Part 1 Trial. *An Exploratory Efficacy Analysis* **2011; 42(6)**: pp 1621-1625, DOI 10.1161/strokeaha.110.610980.

85. Martin RH, Yeatts SD, Hill MD, et al. ALIAS (Albumin in Acute Ischemic Stroke) Trials. *Analysis of the Combined Data From Parts 1 and 2* **2016; 47(9)**: pp 2355-2359, DOI 10.1161/strokeaha.116.012825.

86. Aalen OO, Farewell V, De Angelis D, et al. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Stat Med* **1997; 16(19)**: pp 2191-2210.

87. Keiding N, Klein JP, Horowitz MM. Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine* **2001; 20(12)**: pp 1871-1885, DOI 10.1002/sim.810.

88. Marshall G, Jones RH. Multi-state models and diabetic retinopathy. *Statistics in Medicine* **1995; 14(18)**: pp 1975-1983, DOI 10.1002/sim.4780141804.

89. Capes SE, Hunt D, Malmberg K, et al. Stress Hyperglycemia and Prognosis of Stroke in Nondiabetic and Diabetic Patients. *A Systematic Overview* **2001; 32(10)**: pp 2426-2432, DOI 10.1161/hs1001.096194.

90. Hill MD, Demchuk AM, Goyal M, et al. Alberta Stroke Program Early Computed Tomography Score to Select Patients for Endovascular Treatment. *Interventional Management of Stroke (IMS)-III Trial* **2014; 45(2)**: pp 444-449, DOI 10.1161/strokeaha.113.003580.

91. Pexman JHW, Barber PA, Hill MD, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke. *American Journal of Neuroradiology* **2001; 22(8)**: pp 1534-1542.

92. Adams HP, Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology* **1999; 53(1)**: pp 126, DOI 10.1212/wnl.53.1.126.

93. Cassarly C, Martin RH, Chimowitz M, et al. Assessing type I error and power of multistate Markov models for panel data-A simulation study. *Commun Stat Simul Comput* **2016**, DOI 10.1080/03610918.2016.1222425.

94. Saver JL. Novel End Point Analytic Techniques and Interpreting Shifts Across the Entire Range of Outcome Scales in Acute Stroke Trials. *Stroke* **2007; 38(11)**: pp 3055-3062, DOI 10.1161/strokeaha.107.488536.

95. Smycka J, Keil P. Bayesian PCA, 2015. http://www.petrkeil.com/?p=2316.

96. DiStefano C, Zhu M, Mindrila D. Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evaluation* **2009; 14(20)**.

97. Stevens JP. *Applied Multivariate Statistics for the Social Sciences* (5 edn). Taylor & Francis Group: New York, NY, 2009.

98. Goyal M, Menon BK, van Zwam WH, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet* **2016; 387(10029)**: pp 1723-1731, DOI 10.1016/S0140-6736(16)00163-X.

99. Goldie FC, Fulton RL, Dawson J, et al. Exploration of time-course combinations of outcome scales for use in a global test of stroke recovery. *Int J Stroke* **2014; 9(6)**: pp 755-758, DOI 10.1111/ijs.12131.

100. Cassarly C, Martin RH, Chimowitz M, et al. Treatment effect on ordinal functional outcome using piecewise multistate Markov model with unobservable baseline: An application to the modified Rankin Scale *Stat Med* **2016; [UNDER REVIEW]**.

101. 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(2)**: pp 414-416, DOI 10.1161/01.STR.0000254580.39297.3c.

102. Saver JL, Gornbein J. Treatment effects for which shift or binary analyses are advantageous in acute stroke trials. *Neurology* **2009; 72(15)**: pp 1310-1315, DOI 10.1212/01.wnl.0000341308.73506.b7.

103. Savio K, Pietra GLD, Oddone E, et al. Reliability of the modified Rankin Scale applied by telephone. *Neurology International* **2013; 5(1)**: pp e2, DOI 10.4081/ni.2013.e2.