Statistical methodology for analyzing ordinal outcomes of Traumatic Brain Injury

A Thesis submitted by

Raaj Kishore Biswas

For the award of

Master of Science (Research)

2017

Abstract

Statistical analysis of data for treatment of traumatic brain injury (TBI) from randomized clinical trials (RCTs) regularly fails to identify statistically significant changes in patient condition. Patient outcome is typically measured on an ordinal scale which is then analyzed with statistical methods that lack sensitivity to detect changes across all measured outcomes, have restrictive assumptions, or lack adequate statistical power. The conventional binary regression model, the proportional odds model, partial proportional odds model, and the continuation ratio model are four standard methods applied in the analysis of ordinal variables, traditionally deemed effective in a number of cases. To overcome their known deficiencies in some scenarios, the sliding dichotomy model was recently developed to accurately analyze the changes in patient condition across ordinal scales and has had several productive applications in specific cases.

This study compares the sample size, type I error rate and power among these models. This study attempts to detect the consistency among the contemporary models and also the weakness of the sliding dichotomy model in controlling the type I error rate. A few recommendations for handling ordinal variables in applied research are also proposed.

This study used data from Corticosteroid Randomisation after Significant Head Injury (CRASH), a baseline observed data set consisting of 10,008 patients, as the primary data set. Varying the sample sizes, the number of covariates, the band size of the sliding dichotomy approach and randomizing the treatment effect created different scenarios. A number of possible contexts that might occur in practical clinical trials was simulated to try and test the applicability of the models. For each scenario, the effect on statistical power and type I error rates of the models was assessed. Another supplementary primary data set, already collected from Bangladesh, was applied to compare the two data sets. Apart from these two, we tested two other non-clinical trial data sets to assess the models' application in the field of public health.

Although previous studies have suggested that smaller samples sizes can maintain desired power for some applications of the sliding dichotomy model, the results of this study indicate that consideration of the type I error rates does not encourage this approach, due to the risk of false-positive inferences from application of this method. The model could not even maintain the error rate even when the sample size was high (over 1000); often times the type I error rate

were higher than 5%.

Inconsistent results were observed from all the models applied to different data sets. These inconsistencies across all the ordinal methods suggest that researchers may find value in evaluating multiple methods and using goodness of fit statistics to help report and interpret results, and also encourages use of meta-analysis in some studies. However, this is not the best or ultimate solution to inconsistent performance of methods. Specific problems with the current methods were detected as part of this research and some potential solutions were outlined. Empirical studies with both clinical and non-clinical data are required to devise a model that can adequately balance the errors and statistical power, and have less (or no) restrictive assumptions.

Keywords: Clinical trial; Ordinal methods; Statistical assumptions; Binary logistic regression; Proportional odds model; Partial proportional odds model; Continuation ratio model; Sliding dichotomy model; Cumulative proportion of type I error.

Certification of Thesis

This thesis is the work of **Raaj Kishore Biswas** except where otherwise acknowledged, with the majority of the authorship of the papers presented as a Thesis by Publication undertaken by the student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Principal Supervisor: Rachel King

Associate Supervisor: Enamul Kabir

Student and supervisors signatures of endorsement are held at the University.

Contributions from Publication Co-authors

Two papers are shown in this thesis by publication.

Chapter 3.1: Statistical Analysis of Ordinal Outcomes: Comparison of the existing methods, *under peer review* in PLoS ONE (reference ID: PONE-D-17-28769).

Contribution: RK Biswas contributed 55% to this paper. E Kabir, R King and N Ananna are the co-authors.

Chapter 3.2: Effect of Sex and Age on Traumatic Brain Injury: A Geographical Comparative study, *accepted* in Archives of Public Health (reference ID: AOPH-D-17-00072).

Contribution: RK Biswas contributed 65% to this paper. R King and E Kabir are the co-authors.

Acknowledgments

I would like to acknowledge Dr Ohidul Islam Siddiqui, who gave me the initial idea and also informed me regarding the CRASH data set. I thank all MRC CRASH collaborators, who made their data free for Traumatic Brain Injury studies. I would like to thank Nurjahan Ananna as well for her assistance in attaining a primary data from *Ibrahim medical college*, Bangladesh. I appreciate the Demographic and Health Surveys (DHS) Program, who regularly surveys in low-income countries and provides free access to their data bank.

I would like to specially recognize *National Institute of Neurosciences & Hospital*, Dhaka, Bangladesh for giving me access to their patients' records. I would like to mention Dr Chandra Shekhar Balla, Junior Consultant and Dr Kalim Uddin, Neurosurgeon and consultant at National Institute of Neurosciences & Hospital for helping me amidst their busy schedule.

Dr Rachel King and Dr Enamul Kabir, my supervisors, deserve full credit for my research. Without their help, I would not have completed this thesis. Dr King's continuous persistence with my amateurish writing skills made me a better academic writer. Her patience in explaining my errors was exquisite. Dr Kabir is the reason why the continuous disappointments from various editors and reviewers never bothered me. He made my stay in Toowoomba a bliss. I perhaps had the best possible supervisory team.

Thanks to my co-authors for their contribution in the papers. They aided me with data collection, study design and critical manuscript editing: Rachel King, Enamul Kabir and Nurjahan Ananna.

I appreciate the assistance of English Angels program under *USQ Open Access College*, particularly Dr Marcus Harmes and Dr Barbara Harmes. They have helped me to improve my academic writing over the past 12 months. I acknowledge the brilliant cooperation from the *Faculty of Health, Engineering and Sciences* and *School of Agricultural, Computational and Environmental Sciences*. I would like to mention Professor Grant Daggard, who aided me in several paper submissions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis provides a rigorous assessment of past studies evaluating ordinal outcome variables, analyzes the deficiencies of the contemporary models from these studies, and provides a pathway to develop a robust methodology for clinical trials, particularly for Traumatic Brain Injury (TBI). Ordinal outcome variables, or variables with more than two ranked scales, are popular in many contemporary large scale studies as they are easily applied and interpreted by practitioners, with categories or levels reflecting treatment policies and protocols (Harrell, 2015). A number of statistical methods are available for analysis of such outcome variables. However, each of these methods have their own strengths and weaknesses and none has been accepted as providing consistent unambiguous interpretation, particularly in clinical trials.

Clinical trials of traumatic brain injury (TBI) have a disappointing history in terms of results, with a long track record of indecisive phase III trials (McHugh et al., 2010). The current methods in analyzing phase trials, where the ordinal outcome variables are used, are not robust enough to properly detect the improvements (or deteriorations) in the patients' outcome. A similar level of inadequacy has been observed in stroke trials (Bath et al., 2012). In recent years, the sliding dichotomy model was developed within the area of TBI research (Price et al., 2013) to specifically address this problem of low-power analysis leading to unsuccessful detection of treatment effects when the response variable is ordinal. As it has led to more decisive conclusions from clinical trials, the sliding dichotomy model has gained support among medical researchers, especially when the probability of a favorable outcome is high (Price et al., 2013).

The development processes of the traditional models are important to understand as pathways to design a new model to overcome their existing shortcomings. The chi-square test (Fisher, 1922) or Fisher's exact test (Upton, 1992) are common and reliable statistical tools

for applied studies. The chi-square test has limitations including over dependence on sample size and lack of sensitivity in the orders of outcome scales (Mann and Wald, 1942). The most commonly used traditional (fixed) dichotomous methods are the binary logistic regression and the proportional odds model; however, they also have limitations. For use with ordinal data these models first require that the outcome variables be converted into binary form to avoid complicated analysis (e.g. convergence error) and interpretation (Scott et al., 1997). Although mathematical flexibility and simpler interpretation validate the suitability of the binary linear logistic regression model (Cox, 1972; Hosmer Jr and Lemeshow, 2004), ranked data with more than two scale outcome variables are not efficiently analyzed in this process and valuable information is lost. The most popular method in analyzing ordinal outcomes is the proportional odds model, which is fairly straightforward to apply. However, it has a strict assumption: the proportional odds assumption or parallel assumption, an expectation of similar effects from the covariates for all levels of the outcome scales (Hemri et al., 2016). Another model often considered is the continuation ratio model, which is not flexible enough to incorporate variety of data types. These weaknesses in currently accepted methods paved the way to introduce a more robust methodology in analysis of clinical trial data: the sliding dichotomy model (Berge and Barer, 2002).

Previous studies have shown that the sliding dichotomy model has performed better in achieving higher statistical power with a limited sample size in comparison with other fixed dichotomous methods in certain scenarios (Price et al., 2013). Accepting these studies, it is a natural expectation that sliding dichotomy, in-line with the traditional methods, would also control the type I error rate (the false positives resulting from noise when there is no true effect). However, no assessment on the control of type I error rates of the sliding dichotomy model was previously performed and this study will do so. A method of analysis with a tendency to falsely detect significant effects is of as much, if not more, concern as those that fail to detect meaningful effects at all.

The focus of most social science and public health studies is not on the characteristics of the statistical methods but on the meaningful interpretation of results within the application area (Easterbrook et al., 1991; Stern and Simes, 1997; Sterne et al., 2001). In many cases the method chosen by a researcher may be based simply on the most frequently applied model within their specific discipline area or the availability of relevant software packages (Diez-Roux, 2000; Homer and Hirsch, 2006). As no particular method is unanimously accepted or endorsed for analyzing ordinal outcome data, the analysis approach can also often include the fitting of several different models to determine the best approach (Chu and Ghahramani, 2005). However,

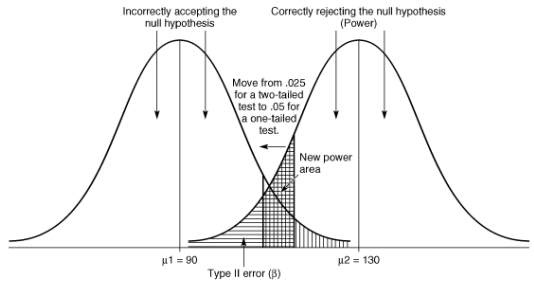
from an applied researcher's perspective, this assessment is often based on the results of the analysis and not necessarily on consideration of the process and limitations of the statistical methods themselves (Jüni et al., 2001; Morgan and Ziglio, 2007).

Traumatic brain injury (TBI) is a global public health problem. When an external force causes alteration to brain function, or there is other evidence of brain pathology, it is regarded as TBI (Menon et al., 2010). Each year, 100 per 100 000 persons in the United States alone sustain a TBI, resulting in hospitalization and potential life-long disability due to a complex variety of cognitive, physical, and emotional consequences in the aftermath of the original TBI event (Thurman et al., 1999; Wells et al., 2005). Schneier et al. (2006) showed that paediatric TBI of US children under 17 years accounted for more than $1 billion in total hospital costs in 2000. These figures show both the clinical and the economic need to study TBI.

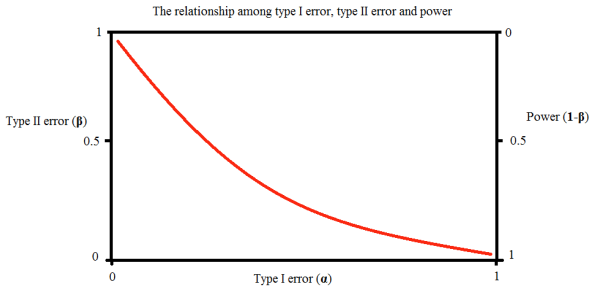
The choice of an appropriate primary-outcome measures for a clinical trial is the critical first step to minimize the potential of the type I or II error rates (Bagiella et al., 2010). Advances in the treatment of TBI require that success in treatment interventions be identifiable in clinical trial studies, so that they can be further improved and refined. At least 21 randomized clinical trials (RCTs) have been conducted from 1980 to 2007 in search of the clinical effectiveness of interventions to treat head injury. However, none of these trials has reported convincing 'positive' findings (Maas et al., 2007). As the traditional dichotomous methods were not performing satisfactorily, the concept of the sliding dichotomy model developed. This study critically analyzes this model to further contribute to the existing literature and each methods' mathematical model is detailed in section 2.2 .

1.1 Power and Type I error

Statistical power is a key parameter to measure the effectiveness of a mathematical (statistical) model. The power of any test of statistical significance is defined by the probability of correctly rejecting a false null hypothesis (Moore et al., 2012). A type I error is the inaccuracy that occurs when the null hypothesis (H_0) is rejected despite being true. Type II error is the exact opposite defined by the probability of accepting null hypothesis when it is false. Types I and II error rates are a zero-sum game for any given sample size; which means any model providing more protection against one type of error is bound to increase the rate of the other kind of error (Lieberman and Cunningham, 2009). Generally type I error is regarded as being more critical than type II.



(a) The Errors and power distribution in bell curves (Goos and Meintrup, 2016)



(b) Relationship direction among the errors and power

Figure 1.1: Type I error, type II error and power explained graphically

Statistical power $(1 - \beta)$ and the type I error rate (α) are related in such a way that when power is increased, α also increases (Blomberg, 2014). Hence a balance between these two is necessary to attain an 'optimal' robust model. For clinical studies, $p = 0.05$ is the generally accepted threshold for type I error rate (Jakobsen et al., 2014), which means a 5% probability of making a mistake is acceptable while rejecting a null hypothesis. However, this year Benjamin et al. (2017) proposed to change the default P-value threshold for statistical significance for claims of new discoveries from 0.05 to 0.005. This change has been proposed due to the increasing evidence concern over non-reproducible research claims of significant effects or relationships within the scientific community. Reducing the accepted P-value would likely reduce the number of significant findings and within each application of a statistical model the type I error rate would also be reduced. Use of appropriate primary-outcome measures (the difference in patient condition before and after treatment) in clinical trials (Bagiella et al., 2010) and their evaluating methods are important for minimizing the potential of type I or II error rates. Past studies with ordinal outcomes, particularly in clinical trials, have not been successful, detailed in section 1.5.

1.2 Traumatic brain injury (TBI)

Alteration in brain function takes a number of forms: any period of loss of or a decrease in level of consciousness (LOC); loss of memory for events immediately before (retrograde amnesia) or after the injury; or any alteration in mental state at the time of the injury (confusion, disorientation, slowed thinking) (Menon et al., 2010). TBI usually occurs as a result of bumps,

blows, or jolts to the head or penetration of the head in any manner which interrupts the regular function of the brain, although not all blows to the head cause TBI (Faul et al., 2010). Studies regarding the treatment of TBI have been increasing significantly in recent years due to the growing global frequency of TBI's.

The severity and frequency of TBI, as mentioned before, is overwhelming. TBI and its subsequent injuries result in more than 50,000 deaths yearly in the USA (Tiesman et al., 2011). Each year approximately 370,000 new cases of TBI are hospitalized in USA (Reeves et al., 2000) and the figure is more than 100,000 for Europe (Maas et al., 2012). Young people are the most common sufferers of TBI and often experience long-term disabilities, which take a toll on both the work force and the economy (Rutland-Brown et al., 2006). It is estimated that TBI results in yearly expenditure of $17 billion dollars in the United States alone (National Center for Injury Prevention and Control (US), 2003). Due to the absence of vital statistics, the impact of TBI's in low-income countries is not easily attainable. From one study conducted in Bangladesh, it was found that individuals with the highest risk of TBI were between 26 and 35 years of age and the second highest frequency occurred among individuals aged 16-25 years, which is similar to American statistics (Mondol et al., 2013). Motor vehicle related trauma accounts for two thirds of moderate and severe TBI in Australia (Khan et al., 2003). These statistics establish TBI as a global issue occurring both in industrialized and third-world nations acting as a "silent global epidemic" (Goldstein, 1990; Myburgh et al., 2008).

1.3 Phase III trials

The method of clinical trials is longstanding and was first introduced in Avicenna's *The Canon of Medicine* in 1025 CE, where he defined the rules for experiments and testing of drugs and treatments, including the use of control groups (Huff, 2003). Generally, in contemporary intervention studies, the investigators provide a 'new' medicine or method of treatment to one group and their changes are compared with a similar homogeneous group without any interventions to test the effectiveness of that treatment (Singh, 2011). Clinical trials are divided into six different stages, called phases. They are commonly referred to as phase 0, I, II, III, IV, and V (Rogers, 2009). Treatment for humans is experimental in the Phase III trials. This 'pre-marketing phase' consists of randomized controlled multi-centre trials on large patient groups (300 to 3,000 or more depending upon the disease/medical condition studied) and targets a conclusive assessment on the effectiveness of the 'new' drug or treatment, in comparison with a current 'best standard' treatment (Friedman et al., 2010). While not required in all cases,

it is typically expected that there be at least two successful phase III trials, demonstrating a drug's safety and efficacy, in order to obtain approval from the appropriate regulatory agencies such as FDA (USA) (Food and Drug Administration (FDA), 2014), or the EMA (European Union) (European Medicines Agency, 2003). We applied data sets of phase III trials as the accuracy of this stage is the most important for assessing the treatment effect on human subjects.

Many of these types of trials, particularly brain injury trials, measure outcomes on ordinal scales. Thus, for a practical clinical trial, it is important to apply ordinal statistical models to assess the trial outcome; this leads to section 1.4.

1.4 Problems associated with Common Statistical Methods Used for Ordinal Outcomes

Ordered categorical outcomes are popular in various fields of study; however, in many cases they are not precisely analyzed (Anderson and Philips, 1981). Sometimes continuous variables are converted into ordinal groups for easier interpretations and their application in various policies, particularly in the physical science and thus, the applications of the ordinal responses are increasing.

The conventional model, commonly known as the **binary logistic regression model**, requires data in a dichotomous (good v bad) form. This does not accord with everyday clinical practice because patient's condition is practically on a continuum either scale or ordinal. A patient's outcome following a severe head injury or a major stroke lies on such continuum and there is no such arbitrary boundary separating a 'bad' outcome from a 'good' outcome (Berge and Barer, 2002). Thus, application of the binary logistic model in such cases results in loss of valuable information, as the model lacks statistical sensitivity in detecting clinically relevant benefit (or harm) from an intervention, and resists sample size calculations (Murray et al., 2005). Hence, the conventional analysis, based on a fixed dichotomy of an ordered outcome scale, neither makes efficient use of the available information nor is it in tune with clinical practice.

The **Proportional Odds Model** (POM) is an efficient method, although it has some drawbacks. The main weakness of the proportional odds model is the parallel assumption, explained in section 2.2.2. It restricts the application of the PPM in various data sets when the assumption is not satisfied, despite attaining high statistical power (Lall et al., 2002). Nonetheless, the ordinal logistic model provides a robust estimate of treatment effect even when this

assumption is not met. Proportional odds modeling is applied in both head injury and stroke trials (Bolland et al., 1998). The proportional odds model explores ordering of the outcome scale and it sensitively detects a shift over the entire ordered scale (Ilodigwe et al., 2013). In some cases, this method has shown robust results compared to the sliding dichotomy model (Price et al., 2013). The **Partial Proportional Odds Model** (PPOM) is an improvement of the proportional odds model, because it does not have any strict odds assumption. However, lack of proper interpretation of data have made the model quite unpopular and it is rarely applied in clinical trials (Washington et al., 2010).

Another possibility, but again not the most widely applied, is the **Continuation Ratio Model** (CRM). This model compares the probability of a response with the probability of the responses in higher scales (Fagerland and Hosmer, 2016). This creates a specific application for the model. However, a number of variations of this model has been developed over the years (Shen et al., 2015). In this study, the constrained continuation ratio model, the most common one, is assessed and reported. These four models are the prominent four models applied in traditional ordinal outcome analysis.

1.5 Applications of models in practice

A range of approaches have been proposed to analyze ordinal responses in their original measured form. A rating experiment resulting in a compound model was proposed by Andrich (1979) Andrich (1979). Two types of estimators for the ordinal outcomes were proposed based on Clayton's (1974) simple odds-ratio statistics (Mocullagh, 1977). In 1981, for the first-time maximum likelihood estimation procedures were established for ordered variables (Anderson and Philips, 1981). It was shown that analysis of ordinal outcomes substantially increased the statistical power compared to the analysis of fixed dichotomous or binary outcomes (Roozenbeek et al., 2011). Farewell (1982) proposed a new class of models based on the introduction of variability of classification into the proportional hazards model. Greenland (1985) illustrated some extensions of logistic models to the modeling of probabilities of the ordinal responses. In all of these models, a number of constraints, or assumptions, were proposed to maintain the ranking of the ordinal outcomes and achieve a meaningful interpretation (Campbell et al., 1991). One of the earliest studies that focused on comparing the performance of existing methods concluded that "the ordering of the response was more intuitive than objective" and "the stereotype model may be more appropriate than the grouped continuous model for data where the ordering is in doubt" (Greenwood and Farewell, 1988). Anderson's (1981) ordinal logis-

tic model, a rank based method, was assumed to be the best-fitted model by these authors (Greenwood and Farewell, 1988). The ordinal least square regression method requires homogeneous variance in the ordinal outcome scales to ensure unbiased parameter estimates; however, variance estimates within ordinal scales are often inconsistent, violating this assumption and resulting in false conclusions (Lipsitz and Buoncristiani, 1994). Any strict assumptions required for an ordinal outcome analysis method can not only make model fitting more mathematically challenging, but may also limit model application as the scale distance in the ranked categories vary from data to data and thus, any uniform assumption may jeopardize the model fitness (Scott et al., 1997). Among the existing analysis techniques, the POM or ordinal logistic regression is the most popular for analyzing ordinal outcomes due to its ease of interpretation (Harrell, 2015; Hemri et al., 2016).

The POM is the favored method due to its ease of application and interpretation (Scott et al., 1997). However, the proportional odds assumption, otherwise known as the parallel lines assumption, can limit its suitability for analysis of many data sets (McCullagh, 1980). The odds assumption is relaxed for the PPOM (Peterson and Harrell Jr, 1990); however, the interpretation of this model is not as straight forward as the POM. For some forms of ordinal data, current membership within any category requires that individuals must have passed through stages leading to their current level, and for these specific types of data the CRM is considered the most appropriate method of analysis (Hardin et al., 2007). The SDM has recently been promoted for demonstrating higher statistical power and providing easily interpretable parameters (Murray et al., 2005); however, subjectivity in choosing the fitted model and untested control over the type I error rate has also led to concerns about the reliability of this method (Price et al., 2013).

Some studies have compared the available methods for ordinal outcome variables within respective disciplines. Scott et al. (1997) demonstrated, based on a public health data set, that the application of the binary logistic model resulted in a substantial loss of information and that the chi-square test was unreasonable for such analysis . Comparing the POM and the CRM, the authors suggested that the POM was more suitable for ranked data analysis for clinical and epidemiological studies (Scott et al., 1997). Fixed dichotomous approaches were also found to be statistically inefficient and inappropriate methods in the analysis of stroke trial data (Bath et al., 2012) and neurological disease trials (Roozenbeek et al., 2009a,b). O'Connell and Liu (2011) applied residual analysis to compare between the POM and the PPOM using an education data set. They suggested that researchers should apply both the ordinary least square (OLS) and the binary logistic models to locate the influential or unusual cases in the data. Furthermore,

the authors also emphasized that the case-wise fit to the POM can be challenging and vigorous testing of assumptions is required, as the existence of outliers can significantly influence the analysis (O'Connell and Liu, 2011). Analysis of data from a clinical trial measuring traumatic brain injury, showed that the SDM allowed sample sizes to be reduced by up to 40% without the loss of any statistical power. Although the POM gave modest additional gains alongside the SDM, it was noted by these authors that the strict proportional odds model assumption may limit its application (McHugh et al., 2010). A comparison of the POM, PPOM and the multinomial regression model in the analysis of crash injury severity data concluded that the PPOM performed better with fewer strict assumptions to be met (Mooradian et al., 2013). Ananth and Kleinbaum (1997) applied six different ordinal models to a perinatal database and compared their goodness of fits and recommended sensitivity analysis along with assumptions testing prior to the fitting of any model to ordinal data.

Ordinal analyses are recommended for analyzing the numerological diseases in clinical trials, where adjustment of the covariates and the enrollment criteria tends to provide more statistical power (Roozenbeek et al., 2009a,b). As mentioned before, the fixed dichotomous approach is statistically inefficient in such trials. In these cases, the ordinal approaches like the POM, CRM or SDM are recommended (Bath et al., 2012). None of the models could uniformly fit every type of data, which creates a literature gap. Due to the lack of single robust model, meta-analysis is increasing in popularity.

Table 1.1: Strengths and weakness of the commonly used ordinal models

Models	Strengths	Weaknesses
Binary Logistic	Robust for binary outcome Easy to apply Simple interpretation No major assumption	Cannot incorporate ranked outcomes Limited information interpreted
Proportional Odds	Incorporate rank order outcomes One odds ratio for all levels Simple interpretation	Strict Proportional odds assumption Cannot adjust to higher variation among ordinal scales
Continuation Ratio	No major assumption Higher power than POM in specific scenarios	Application is restricted to special scenarios Interpretation can be tricky sometimes
Sliding Dichotomy	Appropriate for clinical trials Requires less sample to attain high power Better goodness of fit in most studies compared to POM or CRM	Subjective judgment while creating dichotomous bands Type I error rate untested

The Sliding Dichotomy Model (SDM) is the latest addition (Berge and Barer, 2002) to the ordinal model analysis of clinical trials, and shows much promise in analyzing clinical trial outcomes like Glasgow Outcome Scales (GOS). Murray et al. (2005) showed that fitting the SDM had an impact equivalent to more than doubling the sample size in phase III trials of head injury. The effective application of fitted values and formulating bands (explained in section 2.2.5) make the model more productive over the binary logistic model. The SDM does not require adherence to strict assumptions like the POM. Furthermore, this model is simpler to apply and statistically sound as well. Different studies have analyzed this approach and have achieved sample size reductions ranging from 14% to 50% (Young et al., 2005). Additionally, this model is assumed to provide the highest possible power and more robust results compared to the traditional methods in a number of scenarios, particularly when the probability of a favorable outcome is high (Price et al., 2013). For these reasons, the SDM is the

most recommended application in the head injury trials. The mathematical process underlying each method is explained in section 2.2.

In summary, the conventional binary logistic approach of analysis of a phase III trial in head injury takes an ordered scale, the GOS and collapses the scale to a binary outcome of favorable versus unfavorable. This step discards the steps of ranked outcome and limits statistical power (Murray et al., 2005). SDM also applies dichotomization but also attempts to retain information about differences between all ordinal levels. In addition, SDM helps to reduce the required sample size by the order of 40-50% compared to the traditional models, specifically the POM, for which it is only 25% (McHugh et al., 2010). The SDM attained superior statistical power and reduced sample sizes in post hoc analysis of some head injury trials (Murray et al., 2005). Cases do exist where the fixed dichotomy and the POM performed better than the SDM (Ilodigwe et al., 2013). There are differences among the binary logistic model, the POM, the CRM and the SDM in the magnitude and precision of the quantified odds ratios; however, more studies are required to reach a substantial conclusion regarding these differences in head injury trials (Ilodigwe et al., 2013). The current paradigm still requires a definitive method to help reach conclusive research outcomes. So in this study, the available models were evaluated and their statistical power along with error control algorithms were compared, thus providing a clear road map to develop a *robust* model for the future trials.

1.6 Objectives

The primary aim of this research is to evaluate a set of commonly used models for analysis of ordinal data when they are applied consistently to a common group of Traumatic Brain Injury (TBI) clinical trial data sets. The results and inferences drawn from each analysis, as well as the statistical performance of each model under several scenarios, will be considered in order to properly evaluate the known and suspected weaknesses of each model, particularly the recently proposed SDM which is assumed (in current literature) to overcome some of the problems associated with the more traditional models. The specific objectives are:

1. Apply the models in public health data sets and check if the results are consistent with use in analysis of clinical trials (section 3.1)

2. Compare the models' results in assessing the effect of two standard public health research covariates (Age, sex) on TBI when applied to two different data sets (including the data sets of Objective 1) (section 3.2)

3. Evaluate and compare the type I error control of the binary regression, the POM, and the SDM using the data sets of Objective 1 (section 3.3)

4. Identify and compare the relationship between power, sample size and type I error rate in ordinal outcomes of phase III trials of TBI (section 3.4)

5. Identify possible ways to develop the SDM or the POM (section 3.5)

The project is not intending to analyse or interpret the significance of the treatment covariate (age and sex) applied in each analysis, but rather apply them consistently as covariates to determine if their significance is substantially influenced by the model chosen by the researcher to be applied.

The first two objectives are detailed as papers in Chapter 3.1 and 3.2 respectively, and recent literature supporting the choice of age and sex as exemplar covariates is given in the introductions to each of these papers. These two objectives provide a detailed comparison of the potential inferences made from the application of the four models when applied to the same data sets. Although the SDM has been reported to perform well up till now, the following section (3.3) presents a quantitative analysis of the weaknesses of the SDM in maintain type I error rate. This analysis was further extended in section 3.4 with the power and sample size comparison. Therefore, the first four objectives provide a comprehensive assessment of the current literature gap; the comparative performance of the most popular models used for analysis of ordinal data in various scenarios for given data sets.

These results and assessment of performance then provides new information to help define the extent of the known and suspected weaknesses in the current models to help researchers (generally non-statisticians) make more informed choices about the most appropriate model for their context. The results from the first four objectives also inform suggested pathways for future improvement of the models and possible development of a new ordinal model.

Each objective is discussed in detail in the results and discussion section of each results section, and then the broader context of all results is discussed in Chapter 4 with a summary of the project. Altogether these five specific objectives provide a comprehensive exploration of a current literature gap, performance of the ordinal models in various scenarios, each model's specific weaknesses, and pathways to new development of ordinal outcome models.

Chapter 2

Methodology

The methodology chapter is divided into two sections. The first section (2.1) explains the data sets used in the thesis to assess the models. In the second section (2.2), the methods and their contextual relevance in the study is described.

2.1 Data Description

The data set applied in this study is from CRASH (Corticosteroid Randomisation after Significant Head Injury). The data set is a baseline observed data set consisting of 10,008 patients (ISRCTN74459797). This trial is one of the most recent randomized trials of corticosteroids used to treat head injury. Early results from the CRASH trial were published on 8 October 2004 (Lancet 2004; 364:1321-28) and the 6-month follow-up results in May 2005 (Lancet 2005; 365:1957-59) (Collaborators et al., 2005; CRASH, 2005a). The CRASH trial was a multi-centre collaboration, which includes data from various countries of Europe, Africa, South America, Asia and Oceania.

CRASH is a placebo controlled trial, with a large sample, for assessing the effects of a 48-hour infusion of corticosteroids on death or on neurological disability, among adults with head injury and some loss of consciousness. The total number of patients in the specific trial being discussed was 10,800. After removing cases with missing values in various variables, the final sample size forming the basis of this study was 7,236 patients.

In-hospital deaths, complications, and short-term recovery were recorded on the 'Early Outcome' form which can be completed entirely from the hospital notes - no extra tests were conducted. Long term recovery was assessed at six months using the Glasgow Outcome Scale (GOS), whifch assesses disability and handicap in major areas of life. The GOS was administered by a postal questionnaire (CRASH, 2005b), completed by the patient or a carer, or by

telephone interview. When the patient was admitted into the hospital, his/her condition was measured using the Glasgow Comma Scale (GCS), which has three important components: eye opening, verbal response and motor response (Freeman, 1987; Ivan and Bruce, 1982). Generally, GCS eye opening is measured from 1 to 4, where 1 indicates the best (spontaneous) condition of eyes and 4 indicates no response (cannot open) in a patient's eyes. The similar measurement scale is applied for both GCS verbal response (range 1 to 5) and GCS motor response (range 1 to 6). Thus, a patient having a total GCS score of 3 (1+1+1) is assumed to have no TBI and a patient with a score of 15 (4+5+6) is in the worst possible condition. Overall, the CRASH is a thorough trial and one of the largest for TBI in recent years. A treatment variable with two levels, 'Treatment A' and 'Treatment B', was generated and membership to treatment levels was assigned randomly among cases.

The outcome measure for TBI is the GOS. It has become the most widely used scale for assessing outcome after head injury and non-traumatic acute brain insults (Jennett et al., 1981). Many studies employ variations of this scale, often including additional sub-levels for each main level (Weir et al., 2012). In the CRASH data set, GOS was recorded as an 8-point scale, which has been simplified to the standard five-point scale for this analysis. The standard GOS is a five-point ordinal scale consisting of five main levels: Death (D), Vegetative State (VS), Severe Disability (SD), Moderate Disability (MD) and Good Recovery (GR) (Wright, 2011). In addition, the Death (D) and Vegetative State (VS) levels were merged into a single level (retaining the name Vegetative State (VS)) as the frequency of Death (D) were low in the data set. The stages for surviving patients are explained sequentially from the worst to the best condition.

1. *Vegetative Stage:* This stage is defined as conditions where patients show no evidence of meaningful responsiveness (Jennett and Plum, 1972).

2. *Severe Disability:* This stage is simply termed as 'conscious but dependent' (Teasdale et al., 1998) where the patients' need assistance in day to day life because when physical disability is severe after head injury there is almost always considerable mental deficit (Jennett et al., 1981).

3. *Moderate Disability:* In this section the patients are considered 'independent but disabled'. They can take care of themselves and live on their own.

4. *Good recovery:* This is the stage where patients almost return to normal life and social activities, although a chance of mild stress exists (Teasdale et al., 1998).

A primary data with 151 respondents were collected from National Institute of Nuero Sciences and Hospital, Sher-e-Bangla Nagar, Dhaka for understanding the depth of head injury data, specifically for GOS and GCS. This data contains the same variables as the CRASH data set.

Apart from these two TBI data sets, two other public health data were used to check the adequacy of the models in applied fields. One is a primary data from Ibrahim Medical College (BIRDEM), Bangladesh and the other is Bangladesh Demographic and Health Survey (BDHS), a nationwide health survey. These data sets are explained in their respective sections in results (Chapter 3).

2.2 Statistical models for ordinal outcomes

The five statistical methods applied to the data for this research are explained here along with the underlying mathematical models. No particular study design is reported in this chapter as the relevant study designs for each objective are detailed in the results (Chapter 3), as a section or within the relevant paper manuscript.

2.2.1 Binary Logistic Model

The conversion of an ordinal response variable with four scales (such as GOS) into a binary variable is most commonly achieved by merging levels 1 and 2 into 'unfavorable' and levels 3 and 4 into 'favorable'. The binary logistic model is then fitted with the relevant covariates. Even with the obvious loss of information, this method is widely applied because of its easy interpretation and mathematical flexibility (Hosmer Jr and Lemeshow, 2004). According to Harrell (2015), binary logistic model is defined, probability that $Y = 1$ given X, the value of the predictors:

$$\boldsymbol{Pr\{Y = 1|X\} = [1 + exp(-X\beta)]^{-1}} \tag{2.1}$$

where $\boldsymbol{x_i}$ is a vector measurement considered as the covariates and dummy variables corresponding to factor levels, $\boldsymbol{X\beta}$ represents $\boldsymbol{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k}$, and $\boldsymbol{\beta}$ is the parameter vector. In this study, the R function glm was applied for fitting the binary logistic model. The dichotomisation of the GOS outcome variable was achieved by combining Moderate Disability (MD) and Good Recovery (GR) as favourable outcomes and Severe Disability (SD) along with Vegetative State (VS) as unfavourable outcomes. The strengths and weaknesses of the model was explained in Table 1.1.

2.2.2 Proportional Odds Model

A member of the cumulative logistic regression family, the proportional odds model avoids the necessity for naïve or coarse dichotomization of the outcome variable, allowing for the analysis of the effect of covariates on all original ordinal levels (McCullagh, 2005; Liu and Mukherjee, 2008). Let a random variable be $\boldsymbol{Y}$ with $\boldsymbol{J}$ categories. The cumulative logit model is defined by Harrell (2015), for an outcome with levels $0, 1, 2, ..., k$:

$$\boldsymbol{Pr[Y \geq j|X]} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{exp[-(\alpha_j + X\beta)]}} \qquad j = 1, 2, ..., k. \tag{2.2}$$

where the $\boldsymbol{x}$'s are the covariates, $\boldsymbol{\alpha}$ is the overall intercept, and the $\boldsymbol{\beta}$'s are the unknown parameters. The *R* function *polr* was applied to fit this model. Although all the ordinal levels are retained for analysis, the model estimates a common odds ratio over all of the possible boundaries between levels of the ordinal outcome variable (Roozenbeek et al., 2011). This limitation is due to the proportional odds assumption, which states that the effects of the covariates $x_1, ..., x_{p-1}$ are the same for all levels of the outcome scale, or that the relationship between each pair of outcome groups is the same (Dobson and Barnett, 2008). This is rarely a reasonable assumption within TBI research. The proportional odds assumption is frequently violated. However, Senn and Julious (2009) advised that this issue should not be overstressed. They also accepted that two studies will reach separate conclusions if the cut points for the outcome variable (Y) are different. Many liberal tests for the assumption are available (Peterson and Harrell Jr, 1990). However, an additional limitation of the method is that the odds are strongly affected by sample size and the number of covariates (O'Connell and Liu, 2011). Table 1.1 showed the summary of this model's attributes and deficiencies.

2.2.3 Partial Proportional Odds Model

The PPOM, an extension of the POM, allows the odds assumption to be relaxed. It allows each covariate to vary over the outcome level with separate coefficients ($\boldsymbol{\beta_j}$) estimated for each level. The model can be defined as (Harrell, 2015; Mooradian et al., 2013):

$$\boldsymbol{Pr[Y \geq j|X]} = \frac{\boldsymbol{exp(\alpha_j + X_i\beta_j)}}{\mathbf{1} + \boldsymbol{exp(\alpha_j + X_i\beta_j)}}, \qquad j = 1, 2, ..., k. \tag{2.3}$$

The clear difference between the POM and this model is the conversion of $\boldsymbol{\beta}$ to $\boldsymbol{\beta_j}$. Stata package *gologit2* was applied to calculate the PPOM in this paper.

In this model, the coefficients of intermediate categories need careful interpretation as the sign of the parameters estimated for categories does not always determine the direction of the

effect (Washington et al., 2010). Sometimes it is not possible to come to an precise conclusion. Despite having a reduced burden of assumptions, this model is far from ideal. The advantages and disadvantages of the model was provided in Table 1.1.

2.2.4 Continuation Ratio Model

The CRM is generally applied when the stepwise ranks are given priority in ordinal analysis (Liu, 2010). When estimating the conditional probability of surpassing a category given that the individual is presently in that particular category, given $Y = 0, ..., k$, the CRM model can be defined as (Harrell, 2015),

$$\boldsymbol{Pr(Y = j|Y \geq j, X)} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{exp}[-(\boldsymbol{\theta_j} + \boldsymbol{X\gamma})]}, \quad j = 1, 2, ..., k. \tag{2.4}$$

where $\boldsymbol{x_1, x_2, ..., X_p}$ are a set of predictors, $\boldsymbol{\theta_0 \equiv 0}$, and $\boldsymbol{\theta_1, ..., \theta_{k-1}}$ are increments of intercept, and $\boldsymbol{\gamma}$ is the vector of regression coefficients. Package *vglm* from R was applied to fit this model.

Researchers of clinical trials have been hesitant to utilize adaptive designs; however, due to the complexity of design methods and lack of proper methods, the CRM was applied as an alternative to the traditional models (López et al., 2012). The CRM may provide improvement over the POM in clinical trial designs (Iasonos et al., 2011). However, Scott et al. (1997) showed the POM exhibited better goodness of fit than the CRM in clinical studies. The specific weaknesses and usefulnesses was summarized in Table 1.1.

2.2.5 Sliding Dichotomy Model

This model extends the application of the binary logistic model across multiple, successive dichotomizations within the one ordinal scale, to improve detection and interpretability of effects across all ordinal groups measured. Firstly, based on typical favorable and unfavorable dichotomization of the ordinal response, the binary logistic model is fitted with important covariates that affect the outcome scale and the fitted values are stored within the data set. The patients (cases) are then sorted in descending order of the fitted values and grouped into the maximum number of bands possible, where each band is formulated to represent a change in the scale of the outcome (Ilodigwe et al., 2013). Within the first band, and based on a four-point ordinal scale, scales 2, 3 and 4 of the original GOS are clustered separately (as 'favorable') from scale 1 (as 'unfavorable'). In the next band, scales 1 and 2 are merged separately

from scales 3 and 4. This process continues for all possible dichotomous cut points as displayed in Figure 2.1. Therefore, from an n point ordinal scale, a total of $n-1$ different bands are articulated and each band has its own reformed version of 'favorable' and 'unfavorable' outcomes (Figure 2.1). All of the 'favorable' and 'unfavorable' values from all of the bands are pooled as a new outcome variable so that the original k ordinal scales are transformed into a final, single binary scale (Mendelow et al., 2005). The traditional binary logistic model is then applied to fit the new binary outcome and attain the final result.

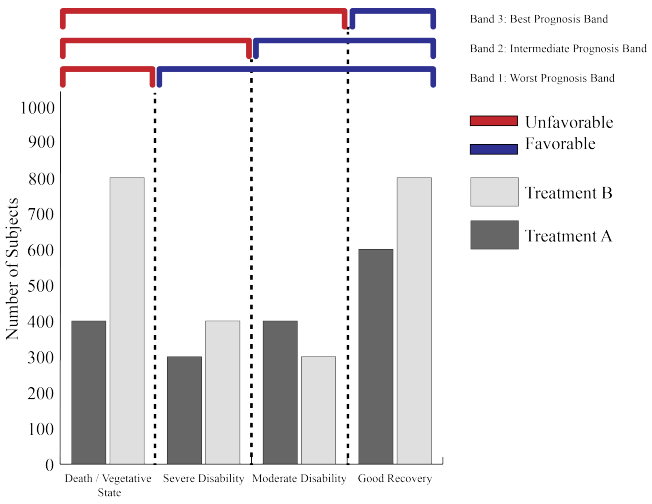


Figure 2.1: Sliding dichotomy explained graphically

The R function glm was used to fit this model. Three GCS variables from CRASH (i.e., eye opening, verbal response and motor response) were used as covariates in the initial model to estimate fitted values (based on standard coarse dichotomization of the GOS). Cases were sorted based on fitted values and three GOS bands were defined: 'Worst', 'Intermediate' and 'Best' prognosis bands (Figure 2.1). For the worst prognosis band, the favorable outcome scales were SD, MD, and GR; for the intermediate band, MD and GR and the only favorable outcome scale for the best prognosis band was GR. The robustness and limitations of the SDM was given in Table 1.1.

Chapter 3

Results

This chapter has five sections. The first two sections contain papers (one published and one under-review), each dealing with a different objective of the thesis. In section 3.3, the type I error analysis is discussed (the manuscript for this chapter is still in preparation and the work is therefore presented in a chapter form and not paper form), Section 3.4 contains the power analysis work and section 3.5 outlines considerations for new model development.

3.1 Statistical Analysis of Ordinal Outcomes: Comparison of the existing methods

This work is presented as a manuscript that is currently **under peer review** in *BMJ Open* (reference ID: bmjopen-2017-021145) titled 'Statistical Analysis of Ordinal Outcomes: Comparison of the existing methods'. RK Biswas initiated the idea and conducted the statistical analysis along with drafting the manuscript. He undertook most of the research work. N Ananna was part of the data collection, provided the ethical review and gave feedback regarding sampling process. E Kabir and R King supervised the whole process and critically edited the manuscript.

The first objective of the thesis (section 1.6) was assessed in this paper. The four models (POM, PPOM, CRM, SDM) were compared in five different covariate categories from a primary data from Bangladesh. Comparing AIC, BIC and $Pseudo - R^2$, we concluded that the SDM showed the best fitness among the four models. This paper further demonstrated the importance of considering multiple models for analyzing ordinal outcome variables, due to the want of a single robust model to fit all data sets to ordinal outcomes.

13 December 2017

Full title: Statistical Analysis of Ordinal Outcomes: Comparison of the existing methods

Short title: Statistical Analysis of Ordinal Outcomes

Raaj Kishore Biswas (corresponding),
School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Australia

Email: *RaajKishore.Biswas@usq.edu.au*

Telephone: +61 426 784 344, fax: N/A

Enamul Kabir
School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Australia

Email: *Enamul.Kabir@usq.edu.au*

Rachel King
School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Australia

Email: *Rachel.King@usq.edu.au*

Nurjahan Ananna
Ibrahim Medical College (BIRDEM, Bangladesh

Email: *nurjahansayed.ananna@gmail.com*

Word count: 3432

Reference count: 68

Abstract

Objective The contemporary statistical models applied to analyze ordinal variables have limitations such as strict model assumptions, uncontrolled type I error rates, lack of power etc. Data manipulation is sometimes necessary to come to a *favorable* and *meaningful* conclusion. This paper compares the proportional odds, partial proportional odds, continuation ratio, and sliding dichotomy model in detecting the rank order changes of an

ordinal outcome variable.
Methods Primary data, collected in a sub-urban region of Bangladesh, was used for the study. A range of covariates were divided into four categories and their effects on BMI were quantified separately using the four models. The significant variables were then fitted to the four models again. *AIC*, *BIC* and *Pseudo*$-R^2$ were calculated in all cases to determine the best-fitted model.
Results The four models did not consistently identify the same significant factors and the proportional odds model violated the proportional odds assumption. The sliding dichotomy model attained the lowest *AIC* and *BIC* along with the highest *Pseudo*$-R^2$ compared to the three other ordinal models.
Conclusions The assumptions for many ordinal models are strict and restrictive. Before analyzing any ordinal data, the assumptions of the models must first be satisfied. Inconsistencies in results across these methods suggests that researchers may find value in evaluating multiple methods and using Goodness-of-Fit indices to help report and interpret results. Analyzing sub-sets of covariates may limit the overfitting of the models. The initial significant factors and the most common significant factors across multiple models may then be included in the final model for interpretation.

Keywords: The proportional odds model, The partial proportional odds model, The continuation ratio model, The sliding dichotomy model, Ordinal variable

---

Strengths and limitations of this study

- Comparison among the application of contemporary ordinal models and the consistency in results obtained
- Model fitness was evaluated using *AIC*, *BIC* and *Pseudo*$-R^2$
- Simulation of multiple scenarios based on a higher initial sample size were not analyzed.
- Changes among the models' fitness statistics were not compared based on various sample sizes

---

Introduction

Ordered categorical variables are regularly used as outcome or response variables in various fields of study, particularly in the areas of social science and epidemiological research. (1) The analysis of variables comprised of more than two ranked scales are, in many cases, not robustly analyzed. (2) There are a number of statistical methods commonly applied to these types of data, however each have demonstrated limits to their efficacy, primarily related to test assumptions or features of the sample. (3) The limitations of these traditional approaches recently led to the development of a new method, the sliding dichotomy model, which itself has its own constraints and limitations. (4) This paper compares the effectiveness of the exiting traditional statistical methods regularly applied for analyzing ordinal dependent variables, with that of the sliding dichotomy method. The methods discussed in this paper are the Proportional Odds Model (POM), Partial Proportional Odds Model (PPOM),

Continuation Ratio Model (CRM), and Sliding Dichotomy Model (SDM).

The application of ordinal responses is increasing in medical and epidemiological studies; however, model selection remains speculative. (5) In many cases ordinal responses are preferred by practitioners and researchers as they allow interpretation across and between practically relevant categories or scales rather than along a continuum (or continuous variable) (1). However, from a statistical perspective: analysis of discrete data often requires larger sample sizes to make equivalent inferences to those that can be made based on continuous data; discrete and categorical data are generally considered less sensitive or less able to capture small variations; and partitioning of variance or identifying sources of variance is more difficult when the measurement scale has low sensitivity, which has led to limited options for statistical analyses of ordinal data. (6)

Some statistical methods can lead to further loss of information by requiring ordinal variables to be converted into binary form to simplify analysis and interpretation. (7) The resulting dichotomous outcome variables allow the application of common and reliable statistical tools such as the chi-square test or Fisher's exact test. (8, 9) The chi-square test has its own limitations, in particular not being flexible enough to adjustments and over dependence on sample size. (10) The binary linear logistic model offers a mathematically flexible model and ease of interpretation for researchers; however, ordinal information is still lost through the conversion to a binary outcome for analysis. (11, 12) The polytomous logistic model is applied to outcome variables with more than two levels; however, it cannot incorporate information on the order of categories which are fundamental to ordinal variables. (13) Conversion of an ordinal response to a binary form may also be performed naïvely assuming the distance between the consecutive scales are equal, which will give false results in most cases. (14)

A range of approaches has been proposed to analyze ordinal responses in their original measured form. A rating experiment resulting in a compound model was proposed by Andrich (1979). (15) Two types of estimators for the ordinal outcomes were proposed based on Clayton's (1974) simple odds-ratio statistics. (16) In 1981, for the first time maximum likelihood estimation procedures were established for ordered variables (2). It was shown that analysis of ordinal outcomes substantially increased the statistical power compared to the analysis of fixed dichotomous or binary outcomes. (17) Farewell (1982) proposed a new class of models based on the introduction of variability of classification into the proportional hazards model. (18) Greenland (1985) illustrated some extensions of logistic models to the modeling of probabilities of the ordinal responses. (19) In all of these models, a number of constraints, or assumptions, were proposed to maintain the ranking of the ordinal outcomes and achieve a meaningful interpretation. (20) One of the earliest studies that focused on comparing the performance of existing methods concluded that "the ordering of the response was more intuitive than objective" and "the stereotype model may be more appropriate than the grouped continuous model for data where the ordering is in doubt". (21) Anderson's (1981) ordinal logistic model, a rank-based method, was assumed to be the best-fitted model by these authors. (21) The ordinal least square regression method requires homogeneous variance in the ordinal outcome scales to ensure unbiased parameter

estimates; however, variance estimates within ordinal scales are often inconsistent, violating this assumption and resulting in false conclusions. (22) Any strict assumptions required for an ordinal outcome analysis method can not only make model fitting more mathematically challenging, but it may also limit model application as the scale distance in the ranked categories vary from data to data and thus, any uniform assumption may jeopardize the model fitness. (7) Among the existing analysis techniques, the POM or ordinal logistic regression is the most popular for analyzing ordinal outcomes due to its ease of interpretation. (5; 23)

The POM is the favored method due to its ease of application and interpretation. (7) However, the proportional odds assumption, otherwise known as the parallel lines assumption, can limit its suitability for analysis of many data sets. (24) The odds assumption is relaxed for the PPOM; (25) however, the interpretation of this model is not as straight forward as the POM. For some forms of ordinal data, current membership within any category requires that individuals must have passed through stages leading to their current level, and for these specific types of data, the CRM is considered the most appropriate method of analysis. (26) The SDM has recently been promoted for demonstrating higher statistical power and providing easily interpretable parameters; (27) however, subjectivity in choosing the fitted model and untested control over the type I error rate has also led to concerns about the reliability of this method. (4)

Some studies have compared the available methods for ordinal outcome variables within respective disciplines. Scott et al. (1997) demonstrated, based on a public health data set, that the application of the binary logistic model resulted in a substantial loss of information and that the chi-square test was unreasonable for such analysis. (7) Comparing the POM and the CRM, the authors suggested that the POM was more suitable for ranked data analysis for clinical and epidemiological studies. (7) Fixed dichotomous approaches were also found to be statistically inefficient and inappropriate methods in the analysis of stroke trial data (3) and neurological disease trials. (28; 29) O'Connell and Liu (2011) applied residual analysis to compare between the POM and the PPOM using an education data set. (30) They suggested that researchers should apply both the ordinary least square (OLS) and the binary logistic models to locate the influential or unusual cases in the data. Furthermore, the authors also emphasized that the case-wise fit to the POM can be challenging and vigorous testing of assumptions is required, as the existence of outliers can significantly influence the analysis. (30) Analysis of data from a clinical trial measuring traumatic brain injury showed that the SDM allowed sample sizes to be reduced by up to 40% without the loss of any statistical power. Although the POM gave modest additional gains alongside the SDM, it was noted by these authors that the strict proportional odds model assumption might limit its application. (31) A comparison of the POM, PPOM and the multinomial regression model in the analysis of crash injury severity data concluded that the PPOM performed better with fewer strict assumptions to be met. (32) Ananth and Kleinbaum (1997) applied six different ordinal models to a perinatal database, and compared their goodness-of-fit indices and recommended sensitivity analysis along with assumptions testing prior to the fitting of any model to ordinal data. (33)

The focus of most social science and epidemiological studies is not on the characteristics

of the statistical methods but the meaningful interpretation of results within the application area. (34; 35; 36) In many cases, the method chosen by a researcher may be based simply on the most frequently applied model within their specific discipline area or availability of relevant software packages. (37; 38) As no particular method is unanimously accepted or endorsed for analyzing ordinal outcome data, the analysis approach can also often include the fitting of several different models to determine the best approach. (39) However, from an applied researcher's perspective, this assessment is often based on the results of the analysis and not necessarily on consideration of the process and limitations of the statistical methods themselves. (40; 41)

This paper aims to compare the performance of four ordinal models when applied to a common data set, based on their identification of significant covariates and statistical measures of performance. The POM, PPOM, CRM and SDM are applied to a primary public health data set from Bangladesh to evaluate their performance. The objective was to assess the significant socio-economic factors that influenced Body Mass Index (BMI), where the ordinal BMI variable was categorized into four scales/groups namely underweight, normal, overweight and obese.

Methods

Data Description

The data collection was conducted by Ibrahim Medical College (BIRDEM), Bangladesh for assessing the health condition in the sub-urban areas of Bangladesh. The data were collected as a part of residential field site training (RFST) for 4th year medical students. 841 respondents were randomly sampled in the unions of Sreepur upazilla of Gazipur district. There was no missing data within the sample. BMI (kg/m^2) was measured as a continuous variable, which was later converted into four ordinal categories by the authors before analysis. These categories were defined based on the World Health Organization guidelines. (42) BMI under 18.5 kg/m^2 was considered 'underweight', between 18.5 to 24.9 kg/m^2 was 'normal', 24.9 to 30 kg/m^2 was 'overweight' and over 30 kg/m^2 was 'obese'. The effect of 12 independent socio-economic covariates were of interest; however, inclusion of all in each of the four methods would have over-populated each model. Therefore, covariates were categorized into four groups broadly representing individual status, family demography, social level and exterior health status (Table 1). The four categories of covariates were fitted to all of the models separately and the significant (5% level) variables were identified. All significant covariates from each category were then fitted to the models again as covariates. R (Version 3.4.0) and Stata (version 12.0) were used for analysis.

Table 1: Categorizing covariates for model formulation

No.	Covariates	Category	No.	Covariates	Category
1	Age	Individual Status	**7**	Social Class	Social Level
2	Sex		**8**	Education	
3	Marital Status		**9**	Occupation	Exterior Health Status
4	Number of family Member	Family Demography	**10**	Physical exercise	
5	Residence in Boyhood		**11**	Smoking Habit	
6	Residence for 3/4th or more life-time		**12**	Tobacco	

In the 'Individual Status' category age, sex and marital status were included, where sex had two scales (male and female) and marital status had four (unmarried, married, widowed and others). Age had three scales: young (below 25), adult (25 - 59) and old (over 59). Number of family members in 'Family demography' was a continuous variable. Residence in childhood was measured by categorizing the locality of the individual until s/he completed high school into urban, sub-urban, rural and other regions. The locality where an individual spent most (3/4 or more) of his/her time used the same categories as residence in childhood. Social class under 'Social Level' included the wealth quantiles: rich, upper middle, lower middle, poor and destitute. Education level was subdivided into illiterate, ability to ready only, write only, secondary or higher secondary education level and tertiary level. 'Exterior Health Status' status had four covariates: occupation (stressful, partially stressful, stress free), physical exercise (< 30 *min*, 30-60 min, 61-90 min and > 90 *min* per day), smoking habit (never, stopped more than 6 months, and current smoker) and tobacco use (yes and no). Tobacco use referred to the chewing of tobacco leaves, not the smoking of cigarettes.

Proportional OddsModel

A member of the cumulative logistic regression family, the POM is generally applied for ordinal outcome analysis, which is preferred over the dichotomous models. (43) Let a random variable be ***Y*** *with* ***J*** *categories*. The cumulative logit model is defined, (5) for an outcome with levels 0,1,2, . . ., k:

$$\Pr[Y \geq j \mid X] = \frac{1}{1 + exp[-(\alpha_j + X\beta)]}, \qquad j = 1, 2, \ldots, k$$

where the $\boldsymbol{x}$'s are the covariates, $\boldsymbol{\alpha}$ is the overall intercept, and the $\boldsymbol{\beta}$'s are the unknown parameters. The *R* function *polr* was applied to fit this model.

Although all the ordinal levels are retained for analysis, the model estimates a common odds ratio over all of the possible boundaries between levels of the ordinal outcome variable. (17) This limitation is due to the proportional odds assumption, which states that the effects of the covariates $\boldsymbol{x_1, \ldots, x_{p-1}}$ are the same for all levels of the outcome scale, or that the relationship between each pair of outcome groups is the same. This is rarely a reasonable assumption for any outcome variable and the proportional odds assumption is frequently violated. However, Senn and Julious (2009) advised, this issue should not be overstressed. (44) They also accepted that two studies would reach separate conclusions if the cut points for the outcome variable (*Y)* were different. Many liberal tests for the assumption are available. (25) However, an additional limitation of the method is that the odds are strongly affected by sample size and the number of covariates. (30)

Partial Proportional Odds Model

The PPOM, an extension of the POM, allows the odds assumption to be relaxed. It allows each covariate to vary over the outcome level with separate coefficients ($\boldsymbol{\beta_j}$) estimated for each level. The model can be defined as, (5; 32)

$$\Pr[Y \geq j \mid X] = \frac{exp(\alpha_j + X_i\beta_j)}{1 + exp(\alpha_j + X_i\beta_j)}, \qquad j = 1, 2, \ldots, k$$

The clear difference between the POM and this model is the conversion of $\boldsymbol{\beta}$ to $\boldsymbol{\beta_j}$. Stata package *gologit2* was applied to calculate the PPOM in this paper.

In this model, the coefficients of intermediate categories need careful interpretation, as the sign of the parameters estimated for categories does not always determine the direction of the effect. (45) Sometimes it is not possible to come to a precise conclusion. Despite having a reduced burden of assumptions, this model is far from ideal.

Continuation Ratio Model

The CRM is generally applied when the stepwise ranks are given priority in ordinal analysis. (46) When estimating the conditional probability of surpassing a category given that the individual is presently in that particular category, given $Y = 0, \ldots, k$, the CRM model can be defined as, (5)

$$\Pr[Y = j \mid Y \geq j, X] = \frac{1}{1 + exp[-(\theta_j + X\gamma)]}, \qquad j = 1, 2, \ldots, k$$

where $\boldsymbol{x_1, x_2, \ldots, x_p}$ are a set of predictors, $\boldsymbol{\theta_0 \equiv 0}$, and $\boldsymbol{\theta_1, \ldots, \theta_{k-1}}$ are increments of intercept, and $\boldsymbol{\gamma}$ is the vector of regression coefficients. Package *vglm* from R was applied to fit this model. Researchers of clinical trials have been hesitant to utilize adaptive designs; however, due to the complexity of some trial designs and a lack of analysis options for the resulting data from these trials, the CRM has been applied as an alternative to the traditional models. (47) Application of this model has had varying results, with some researchers finding improved fit of the CRM over the POM and others finding no improvement. (7, 48)

Sliding Dichotomy Model

The SDM extends the application of the binary logistic model across multiple, successive dichotomizations within the one ordinal scale, to improve detection and interpretability of effects across all ordinal groups measured. Firstly, based on typical favorable and unfavorable dichotomization of the ordinal response, the binary logistic model is fitted with the important covariates that affect the outcome scale and the fitted values are stored within the data set. The respondents (cases) are then sorted in descending order of the fitted values and grouped into the maximum number of bands possible, where each band is formulated to represent a change in the scale of the original ordinal outcome variable. (49) Within the first band, and based on a four-point ordinal scale, scales 2 (normal), 3 (overweight) and 4 (obese) of the original BMI are clustered separately (as 'favorable') from scale 1 (underweight) (as 'unfavorable'). In the next band, scales 1 and 2 are merged separately from scales 3 and 4. This process continues for all possible dichotomous cut points, as displayed in Figure 1. Therefore, from an n point ordinal scale, a total of $n-1$ different bands are articulated and each band has its own reformed version of 'favorable' and 'unfavorable' outcomes (Figure 1). All of the 'favorable' and 'unfavorable' values from all of the bands are pooled as a new outcome variable, so that the original k ordinal scales are

transformed into a final, single binary scale. (50) The traditional binary logistic model is then applied to fit the new binary outcome.

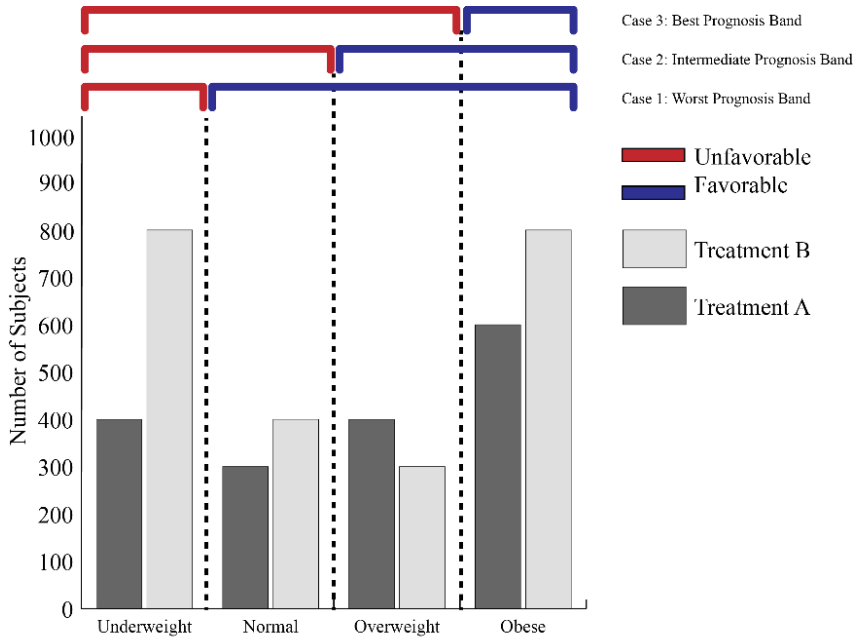


Figure 1: The sliding dichotomy model explained graphically

The R function *glm* was used to fit this model. Age and education of the respondents were used as covariates in the initial model to estimate fitted values (based on standard coarse dichotomization of the BMI). Cases were sorted based on fitted values and three BMI bands were defined: 'Worst', 'Intermediate' and 'Best' prognosis bands (Figure 1). For the worst prognosis band, the favorable outcome scales were normal, overweight, and obese; for the intermediate band, overweight and obese, and the only favorable outcome scale for the best prognosis band was obese. This method is highly praised in the area of clinical trial research for achieving high statistical power (51). The predictive model can be created using any number of covariates from the data set, which may vary from study to study. Hence, the choice of variables leads to subjectivity within the first stage of analysis.

Comparison of the models

Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and $Pseudo{-}R^2$ were quantified to assess the model fitness. These three criteria were calculated for each covariate category (Table 1) along with the final set of significant covariates. In total, five sets of AIC, BIC and $Pseudo{-}R^2$ were computed and compared.

AIC, a relative measure, can numerically express the amount of information in a model incorporating the number of covariates with maximized log-likelihood (52). It is defined by, $AIC = -2L_m + 2m$, where L_m is the maximized log-likelihood and m is the number of parameters in the model. (53) BIC is an extension of AIC, where sample size is taken into consideration. (54) It is defined by, $BIC = -2L_m + Klog(n)$, where L_m is maximized log-likelihood, K is the number of estimable parameters and n is the sample size. (55)

R^2 describes the proportion of variance in the dependent variable explained by the model. (56) It is quantified comparing the conditional variation of the response to the marginal variation, using the deviance of the observed, null and saturated model. One slightly controversial application of the deviance is to derive a $Pseudo{-}R^2$ measure, known as the log-likelihood or Hosmer-Lemeshow R^2. (57) Although $Pseudo{-}R^2$ is not the most reliable method, different versions of it have been developed. This study applied Mcfadden-R^2, which is defined by, $Mcfadden\ -\ R^2\ =\ 1\ -\ \frac{log\ (L_1)}{log\ (L_2)}$ where L_1 denotes the (maximized) likelihood value from the current fitted model and $\boldsymbol{L_0}$ denotes the corresponding value of the null model. (58)

Results

The four models did not show uniform results, with different variables identified as significantly influencing BMI in different models based on analysis of the same samples and covariates. Age ('Adult' group) was a significant covariate ($P - value < 0.05$) both in the POM and SDM; however, the CRM did not show any variable in the 'Individual Status' category to be significant (Table 2). Females were found to have a 45% greater chance of having higher BMI compared to males, according to the SDM. In the 'Family Demography' category, the CRM and SDM identified only the quantity of family members as a significant covariate; however, their odds were in opposite directions, which leads to a conflicting conclusion (Table 2). The POM identified both social class and education as significant factors, while the SDM only showed social class as a significant covariate. In the 'Exterior Health Status' category, the SDM showed only occupation as a significant covariate while the POM and CRM identified no significant covariates within this category (Table 2). The odds assumptions of the POM was assessed and Figure 2-5 shows that they were violated in all cases. Equal proportion was not evident in any of the variable levels. The PPOM identified age, number of family members, social class, education, physical exercise and smoking habit to be significant factors for BMI (Table 3).

Table 2: The four categories fitted with relevant models

Model	Individual Status [Odds ratio (95% C.I.)]							
	Gender: Female (ref: male)	**Age: Adult (ref: Old)**	**Age: Young (ref: Old)**	**Marital Status: Married (ref: Unmarried)**	**Marital Status: Widowed (ref: Unmarried)**	**Marital Status: Others (ref: Unmarried)**		
POM	1.223 (0.91, 1.65)	0.594* (0.41, 0.86)	0.943 (0.48, 1.84)	1.000 (0.46, 2.18)	1.099 (0.42, 2.85)	0.966 (0.14, 6.64)		
CRM	1.014 (0.85, 1.22)	0.988 (0.78, 1.24)	1.000 (0.66, 1.52)	1.009 (0.62, 1.64)	1.008 (0.55, 1.85)	0.988 (0.31, 3.12)		
SDM	1.449* (1.04, 2.02)	0.667* (0.45, 0.99)	1.437 (0.71, 2.91)	1.463 (0.62, 3.45)	0.983 (0.34, 2.87)	0.717 (0.06, 8.14)		
	Family Demography							
	No. of family Members	**Residence in Childhood: Urban (ref: Rural)**	**Residence in Childhood: Sub-urban (ref: Rural)**	**Residence in Childhood: (ref: Rural)**	**3/4th or more lifetime spent: Urban (ref: Rural)**	**3/4th or more lifetime spent: Sub-urban (ref: Rural)**	**3/4th or more lifetime spent: Others (ref: Rural)**	
POM	0.979 (0.91, 1.05)	2.598 (0.49, 13.89)	2.258 (0.59, 8.61)	No Observation	1.987 (0.35, 11.24)	2.143 (0.50, 9.19)	1.744 (0.09, 30.94)	
CRM	0.577* (0.34, 0.98)	1.175 (0.43, 3.19)	1.162 (0.52, 2.57)	No Observation	1.099 (0.40, 3.02)	1.079 (0.47, 2.46)	0.939 (0.18, 4.90)	
SDM	1.095* (1.02, 1.18)	3.018 (0.43, 21.20)	2.209 (0.43, 11.39)	No Observation	1.588 (0.22, 11.23)	1.147 (0.23, 6.10)	0.902 (0.05, 16.91)	
	Social Level							
	Social Class: Upper middle (ref: Rich)	**Social Class: Lower middle (ref: Rich)**	**Social Class: Poor (ref: Rich)**	**Social Class: Destitute (ref: Rich)**	**Education: Read only (ref: Illiterate)**	**Education: Write address (ref: Illiterate)**	**Education: Secondary or higher (ref: Illiterate)**	**Education: Tertiary/+ (ref: Illiterate)**
POM	0.667 (0.31, 1.44)	0.551 (0.26, 1.17)	0.432* (0.19, 0.94)	One observation	1.040 (0.62, 1.73)	0.716* (0.51, 0.99)	0.688 (0.45, 1.05)	0.680 (0.41, 1.14)
CRM	0.902 (0.52, 1.58)	0.848 (0.49, 1.47)	0.790 (0.45, 1.39)	One observation	1.015 (0.74, 1.39)	0.979 (0.79, 1.20)	0.964 (0.74, 1.26)	1.007 (0.73, 1.39)
SDM	11.714* (1.49, 91.91)	11.366 (1.47, 87.80)	11.194 (1.43, 87.38)	One observation	1.197 (0.68, 2.10)	1.182 (0.82, 1.69)	0.900 (0.56, 1.45)	0.790 (0.42, 1.48)
	Exterior Health Status							
	Occupation: Stress less (ref: Stressful)	**Occupation: Partial Stressful (ref: Stressful)**	**Physical Exercise: 31-60 min (ref: <30 min)**	**Physical Exercise: 61-90 min (ref: <30 min)**	**Physical Exercise: >90 min (ref: <30 min)**	**Smoking: stopped more than 6 months (ref: Never)**	**Smoking: Current (ref: Never)**	**Tobacco: No (ref: Yes)**
POM	1.099 (0.66, 1.83)	0.989 (0.58, 1.68)	0.793 (0.56, 1.12)	0.870 (0.56, 1.36)	0.602 (0.28, 1.28)	0.636 (0.34, 1.19)	0.706 (0.42, 1.18)	1.115 (0.82, 1.51)
CRM	1.025 (0.76, 1.39)	1.006 (0.73, 1.38)	0.966 (0.78, 1.19)	0.959 (0.73, 1.26)	0.899 (0.59, 1.38)	0.937 (0.66, 1.33)	0.915 (0.68, 1.23)	1.031 (0.85, 1.24)
SDM	0.514* (0.30, 0.87)	0.469* (0.27, 0.82)	0.846 (0.57, 1.25)	0.768 (0.47, 1.27)	0.869 (0.39, 1.89)	0.593 (0.30, 1.16)	0.791 (0.46, 1.37)	1.074 (0.76, 1.51)

* Significance at 5%, POM = Proportional odds model, CRM = Continuation Ratio Model, SDM = Sliding dichotomy model

Table 3: Partial Proportional Odds Model results for all covariate category

Covariate Category	Variable (reference cat.)	Panel I	Panel II	Panel III
		Odds (C.I.)		
Individual Status	Gender (male)	0.839 (0.56, 1.25)	1.140 (0.81, 1.61)	3.35* (1.27, 8.81)
	Age (Old)	0.630* (0.44, 0.91)	0.717 (0.49, 1.03)	0.629 (0.27, 1.47)
	Marital Status (Unmarried)	1.042 (0.62, 1.75)	1.110 (0.69, 1.77)	1.590 (0.72, 3.54)
Family Demography	No. of family Members	0.949 (0.87, 1.03)	0.906* (0.83, 0.99)	1.020 (0.87, 1.19)
	Residence in Childhood (Rural)	0.526 (0.16, 1.71)	1.099 (0.57, 2.12)	0.594 (0.26, 1.35)
	3/4th or more lifetime spent	0.666 (0.22, 2.02)	0.963 (0.49, 1.87)	0.858 (0.35, 2.04)
Social Level	Social Class (Rich)	0.920 (0.70, 1.20)	0.567* (0.45, 0.71)	0.560 (0.37, 0.85)
	Education (Illiterate)	1.413 (1.21, 1.65)	1.153* (1.02, 1.31)	1.113 (0.87, 1.42)
Exterior Health	Occupation (Stressful)	1.118 (0.81, 1.54)	1.128 (0.86, 1.48)	1.430 (0.79, 2.57)
	Physical Exercise (<30 min)	0.965 (0.76, 1.22)	0.761* (0.62, 0.94)	0.606* (0.38, 0.97)
	Smoking (Never)	0.939 (0.69. 1.29)	0.624* (0.45, 0.86)	0.341 (0.11, 1.09)
	Tobacco (Yes)	1.047 (0.68, 1.61)	1.270 (0.89, 1.80)	1.240 (0.62, 2.45)
Overall	Age (Old)	0.696* (0.49, 0.99)	0.837 (0.58, 1.19)	0.749 (0.36, 1.58)
	No. of family Members	0.972 (0.89, 1.06)	0.922 (0.84, 1.01)	1.020 (0.87, 1.19)
	Social Class (Rich)	0.727* (0.57, 0.93)	0.527* (0.43, 0.65)	0.546* (0.37, 0.79)

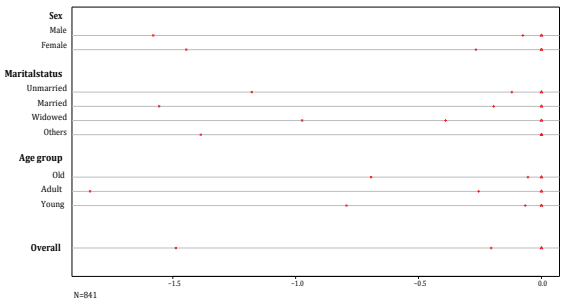
* Significance at 5%

From the initial results, portrayed in Table 2 and 3, three variables age, number of family members and social class were chosen for re-analysis, as they were significant in at least two models. These factors were again fitted using the four models and the results are provided in Table 4 and 3. Age was a significant covariate in the POM, SDM and PPOM (Table 4 and 3). The PPOM identified social class as an important factor as well; however, it was not significant in the other models. Hence, among the 12 original covariates only one, age, was found significant in the three models out of four. These results illustrate the variability among the models and the lack of consistency in the traditional methods in ordinal outcome analyses. In addition, the most popular model, the POM, violated the proportional odds assumption again, when the final three covariates were fitted (Figure 6).

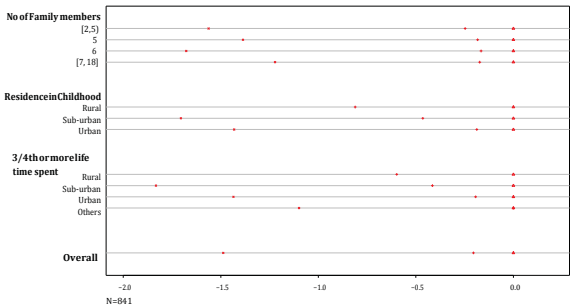
Table 4: Fitting the models with selected variables

Models [Odds ratio (C.I.)]	Age: Adult (ref: Old)	Age: Young (ref: Old)	No. of family Members	Social Class: Upper middle (ref: Rich)	Social Class: Lower middle (ref: Rich)	Social Class: Poor (ref: Rich)	Social Class: Destitute (ref: Rich)
POM	0.545* (0.38,0.79)	0.878 (0.48,1.61)	0.959 (0.89, 1.03)	0.719 (0.33, 1.58)	0.637 (0.30, 1.34)	0.548 (0.26, 1.16)	One observation
CRM	0.969 (0.77,1.22)	0.973 (0.67,1.42)	0.989 (0.95, 1.03)	0.901 (0.52, 1.57)	0.847 (0.50, 1.43)	0.797 (0.47, 1.35)	One observation
SDM	0.932 (0.62,1.39)	1.966* (1.03,3.76)	1.027 (0.96, 1.11)	1.072 (0.30, 3.78)	1.577 (0.47, 5.29)	1.920 (0.57, 6.45)	One observation

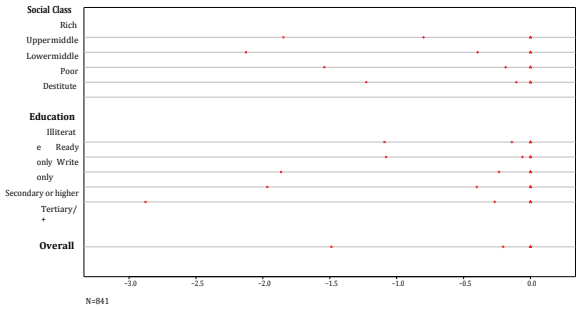
* Significance at 5%, POM = Proportional odds model, CRM = Continuation Ratio Model, SDM = Sliding dichotomy model



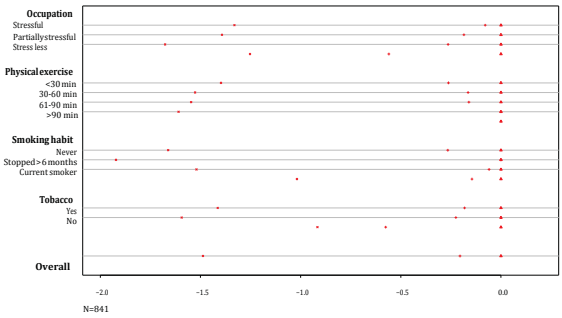
(a) Odds assumption in 'Individual Status'



(b) Odds assumption in 'Family Demography'



(c) Odds assumption in 'Social Level'



(d) Odds assumption in 'Exterior Health Status'

Figure 2: Test of odds assumption for Proportional Odds Model

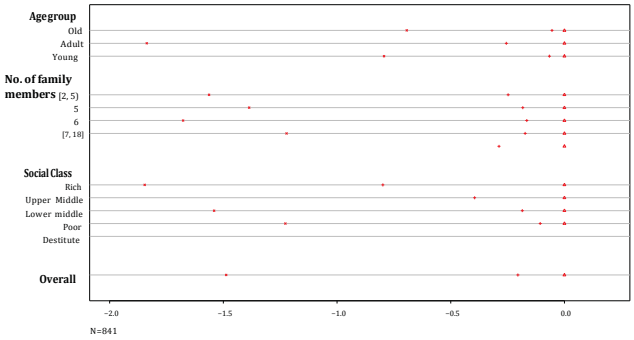
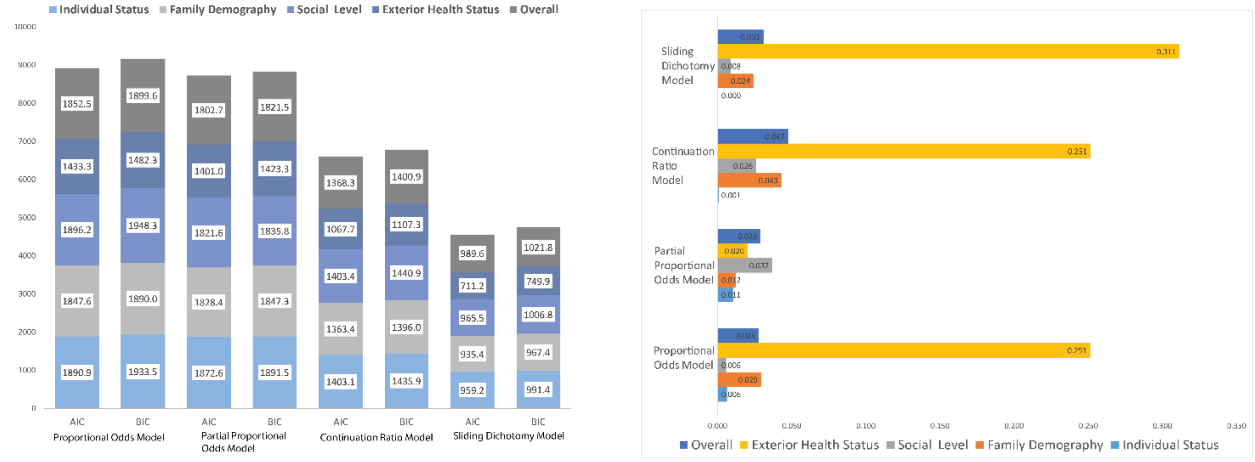


Figure 3: Test of odds assumption for Proportional Odds Model for selected variables

The goodness-of-fit indices of the models were assessed based on AIC, BIC and Mcfadden–R^2. Low values of AIC and BIC and high values of McFadden–R^2 indicate better fits of the models to the data. The SDM had obtained the lowest AIC and BIC among the four models followed by the CRM, PPOM and POM respectively (Figure 7). The *Pseudo*–R^2 accentuates these results, as the SDM and CRM achieved the highest R^2 value in five different models; except for 'Individual Status', where the PPOM had the highest R^2 Score (Figure 8). These results indicate that the sliding dichotomy model outperformed the other models for this particular data set.



(a) AIC and BIC for the models

(b) *Mcfadden*– R^2 for the models

Figure 4: Test of goodness-of-fit: a) AIC and BIC b) *Mcfadden* –R^2

Discussion

Results from one model were not consistent with other models. According to the POM, age, social class, and educational status significantly influenced BMI. However, only the number of family members was significant based on the CRM results. Gender, age, number of family members, social class, and occupation were all significant covariates based on the SDM. The PPOM showed age, number of family members, social class, education, physical exercise and smoking habit as significant factors for BMI. Each of these models fitted for another study like this one, would likely provide varying and potentially misleading results. Such variability in the results could undermine any study conclusions if the models are not cautiously fitted.

This paper used three goodness-of-fit indices to compare the results of the four models The SDM showed the highest fit across all three statistics. The final model comprised of the preliminary significant factors, showed only one variable, age, significantly influenced the BMI levels, which is consistent with the current literature. (59) However, this result was a derived from a combination of the all four models, and their goodness-of-fit indices. The variability in the outcomes of the models shows the necessity of evaluating models' performance based on such indices. (60; 61) We suggest application of multiple models and evaluation of multiple goodness of fit statistics, for analysis of a primary data where there is

In addition to the work in this paper, this **supplementary material** also contributes to objective 1 (section 1.6). This work investigates whether micro-level health condition improves gradually for every income class despite the existence of growing income inequality as macro-level economy flourishes. Bangladesh, exhibiting development in both economy and health, is chosen as a sample for assessing the hypothesis. Bangladesh Demographic and Health Surveys (BDHS) of 2007, 2011 and 2014 were applied in this paper. Only the data of female respondents were taken as BHDS focuses on women health.

Three statistical models: the proportional odds model, the continuation ratio model and the sliding dichotomy model were applied for robustness. The proportional odds model showed the highest $Pseudo - R^2$ and it was fitted to determine the relationship between wealth index and various health components. The direction of the relationship over the years were explored by trend of the odds ratios. The fitted models showed that most of the health components (e.g. family health consciousness, women's empowerment and reproductive health) significantly influenced wealth index, the outcome variable, at 5% level for all three data sets (Table 3.1 and Table 3.2). However, the odds showed a trend towards one (Odds = 1) depicting a reduction of influence on household economic status from 2007 to 2014. From 2007 to 2014, Bangladesh has doubled its per-capita GDP and showed remarkable achievements in Millennium Development Goals (MDGs) despite increasing inequality. Most health parameters showed less influence on wealth as the economy progressed. The pattern suggests that health accessibility increases as the country advances economically even though there exists high inequality. Therefore, overall development of a country is a blessing even for the most vulnerable part of the economic quantile because they can access health service despite their insolvency.

Table 3.1: Proportional odds model fitted with three categories of variables from 2007, 2011 & 2014

Variables	2007		2011		2014	
	Frequency (%)	Odds (95% C.I.)	Frequency (%)	Odds (95% C.I.)	Frequency (%)	Odds (95% C.I.)
Family Health Consciousness						
BMI Underweight	2349 (21.4)	1.00	3177 (17.8)	1.00	2543 (14.2)	1.00
BMI Normal	6557 (59.6)	1.47* (1.35, 1.60)	10162 (56.9)	1.71* (1.59, 1.83)	9742 (54.6)	1.63* (1.50, 1.76)
BMI Overweight	1687 (15.3)	5.59* (4.95, 6.32)	3519 (19.8)	4.88* (4.45, 5.34)	4614 (25.8)	4.06* (3.71, 4.45)
BMI Obese	403 (3.7)	6.13* (4.96, 7.58)	984 (5.5)	5.63* (4.91, 6.46)	964 (5.4)	7.25* (6.29, 8.36)
Number of Son Died: None	9411 (85.6)	1.00	15761 (88.3)	1.00	16046 (89.8)	1.00
Number of Son Died: One	1254 (11.4)	0.76* (0.68, 0.84)	1748 (9.8)	0.73* (0.67, 0.80)	1530 (85.6)	0.79* (0.72, 0.87)
Number of Son Died: >2	331 (3.0)	0.88 (0.72, 1.08)	333 (1.9)	0.89 (0.74, 1.09)	287 (1.6)	0.69* (0.56, 0.86)
Number of Daughter Died: None	9539 (86.8)	1.00	15977 (89.5)	1.00	16263 (91.1)	1.00
Number of Daughter Died: One	1168 (10.6)	0.85* (0.76, 0.95)	1151 (8.7)	0.83* (0.76, 0.91)	1400 (7.8)	0.72* (0.58, 0.89)
Number of Daughter Died: >2	289 (2.6)	0.78* (0.63, 0.97)	314 (1.8)	0.62* (0.51, 0.77)	200 (1.1)	0.58* (0.45, 0.75)
Aware of Com. Clinic: No	1072 (9.7)	1.00	14678 (82.3)	1.00	12662 (70.9)	1.00
Aware of Com. Clinic: Yes	9924 (90.3)	0.53* (0.47, 0.59)	3164 (17.7)	0.71* (0.66, 0.76)	5201 (29.1)	0.67* (0.63, 0.71)
Aware of AIDS: No	3308 (30.1)	1.00	5330 (29.9)	1.00	5270 (29.5)	1.00
Aware of AIDS: Yes	7687 (69.9)	4.18* (3.85, 4.53)	12512 (70.1)	3.94* (3.69, 4.19)	12593 (70.5)	3.78* (3.55, 4.03)
Women Empowerment						
Decision maker for Child health-care: Herself	3095 (33.3)	1.00	2423 (16.8)	1.00	2823 (19.3)	1.00
Decision maker for Child health-care: Joint	4459 (48.1)	0.76* (0.69, 0.82)	8894 (61.7)	0.75* (0.69, 0.82)	8667 (59.3)	0.73* (0.67, 0.79)
Decision maker for Child health-care: Husband	1723 (18.6)	0.63* (0.56, 0.69)	3109 (21.5)	0.61* (0.56, 0.68)	3124 (21.4)	0.56* (0.51, 0.61)
Working Status: Unemployed	7759 (70.6)	1.00	15468 (86.7)	1.00	12234 (68.5)	1.00
Working Status: Employed	3233 (29.4)	0.66* (0.62, 0.72)	2374 (13.3)	1.01 (0.92, 1.10)	5624 (27.5)	0.83* (0.78, 0.88)
Education: None	3525 (32.1)	1.00	4639 (26.0)	1.00	4206 (23.5)	1.00
Education: Primary	3268 (29.7)	1.25* (1.36, 1.14)	5332 (29.9)	1.25* (1.16, 1.35)	5226 (29.3)	1.18* (1.09, 1.28)
Education: Secondary	3345 (30.4)	2.11 (1.90, 2.33)	6406 (35.9)	2.18* (1.99, 2.37)	6722 (37.6)	1.98* (1.82, 2.16)
Education: Higher	855 (7.8)	5.81 (4.73, 7.14)	1465 (8.2)	5.47* (4.68, 6.38)	1709 (9.6)	4.52* (3.93, 5.19)
Partner's education: None	3602 (32.8)	1.00	5197 (29.1)	1.00	5062 (28.3)	1.00
Partner's education: Primary	2881 (26.2)	1.67* (1.52, 1.83)	4834 (27.1)	2.02* (1.87, 2.19)	4855 (27.2)	1.77* (1.64, 1.91)
Partner's education: Secondary	2900 (26.4)	3.63* (3.28, 4.02)	5175 (29.0)	4.66* (4.28, 5.08)	5266 (29.5)	4.10* (3.77, 4.47)
Partner's education: Higher	1598 (14.6)	9.07* (7.77, 10.59)	2627 (14.8)	12.04* (10.65, 13.61)	2677 (14.9)	10.51* (9.33, 11.85)
Reproductive Health						
Respondent's age at first birth		1.12* (1.11, 1.13)		1.11* (1.10, 1.12)		1.12* (1.11, 1.13)
Contraceptive Method: None	5368 (48.8)	1.00	7563 (42.3)	1.00	7391 (41.4)	1.00
Contraceptive Method: Folkloric	50 (0.5)	0.48* (0.29, 0.78)	64 (0.4)	0.64* (0.42, 0.98)	28 (0.2)	0.90 (0.46, 1.77)
Contraceptive Method: Traditional	827 (7.5)	1.38* (1.20, 1.58)	1499 (8.4)	1.13* (1.02, 1.25)	1310 (7.3)	1.13* (1.01, 1.26)
Contraceptive Method: Modern	4751 (43.2)	1.19* (1.10, 1.28)	8716 (48.9)	1.01 (0.96, 1.08)	9134 (51.1)	0.90* (0.85, 0.95)
Aware of MR: No	2167 (19.7)	1.00	12517 (70.2)	1.00	9433 (52.8)	1.00
Aware of MR: Yes	8828 (80.3)	1.85* (1.69, 2.03)	12517 (70.2)	2.04* (1.91, 2.17)	8430 (47.2)	2.40* (2.27, 2.54)
Ideal number of Boys: None	2568 (23.4)	1.00	3458 (19.4)	1.00	4923 (28.1)	1.00
Ideal number of Boys: One/two	8060 (73.4)	0.70* (0.64, 0.77)	14054 (79.0)	0.82* (0.77, 0.89)	12480 (71.2)	0.86* (0.81, 0.92)
Ideal number of Boys: >=3	123 (1.1)	0.48* (0.35, 0.65)	115 (0.7)	0.56* (0.39, 0.79)	118 (0.7)	0.48* (0.34, 0.67)
Ideal number of Boys: Faith decides	230 (2.1)	0.66* (0.51, 0.85)	158 (0.9)	0.58* (0.43, 0.78)	2 (0)	0.65* (0.08, 1.13)

* at 5% level of significance

Table 3.2: Continuation Ratio model and Sliding Dichotomy model fitted with three categories of variables from 2007, 2011 & 2014

	Continuation Ratio Model (Odds (C.I.))			Sliding Dichotomy Model (Odds (C.I.))		
Variables	2007	2011	2014	2007	2011	2014
Family Health Consciousness						
BMI Underweight	1	1	1	1	1	1
BMI Normal	1.01 (0.98, 1.04)	1.01 (0.98, 1.03)	0.99 (0.97, 1.02)	0.99 (0.89, 1.09)	1.02 (0.94, 1.11)	1.07 (0.98, 1.17)
BMI Overweight	1.11 (1.07, 1.15)*	1.07 (1.04, 1.09)*	1.05 (1.02, 1.08)*	1.02 (0.89, 1.16)	0.96 (0.87, 1.06)	1.14 (1.03, 1.26)
BMI Obese	1.13 (1.07, 1.19)*	1.09 (1.05, 1.14)*	1.09 (1.05, 1.14)*	1.16 (0.93, 1.44)	1.02 (0.88, 1.18)	0.97 (0.83, 1.12)
Number of Son Died: None	1	1	1	1	1	1
Number of Son Died: One	0.98 (0.94, 1.01)	0.99 (0.96, 1.02)	0.98 (0.95, 1.02)	0.97 (0.86, 1.09)	0.97 (0.88, 1.08)	1.03 (0.93, 1.15)
Number of Son Died: >2	0.99 (0.93, 1.06)	0.98 (0.92, 1.05)	0.99 (0.92, 1.06)	0.92 (0.74, 1.16)	0.99 (0.79, 1.24)	0.98 (0.78, 1.24)
Number of Daughter Died: None	1	1	1	1	1	1
Number of Daughter Died: One	0.99 (0.95, 1.02)	0.99 (0.96, 1.02)	0.98 (0.95, 1.02)	0.97 (0.86, 1.10)	1.06 (0.95, 1.18)	1.09 (0.97, 1.22)
Number of Daughter Died: >2	0.99 (0.93, 1.07)	0.98 (0.91, 1.05)	0.99 (0.91, 1.08)	1.23 (0.96, 1.58)	1.07 (0.84, 1.35)	0.87 (0.66, 1.15)
Aware of Com. Clinic: No	1	1	1	1	1	1
Aware of Com. Clinic: Yes	0.96 (0.93, 0.99)*	0.97 (0.95, 0.99)*	0.97 (0.95, 0.99)*	0.92 (0.81, 1.04)	0.91 (0.84, 0.99)	1.11 (1.04, 1.19)
Aware of AIDS: No	1	1	1	1	1	1
Aware of AIDS: Yes	1.03 (1.00, 1.06)*	1.02 (1.00, 1.04)*	1.01 (0.99, 1.03)	1.07 (0.98, 1.17)	1.03 (0.96, 1.10)	1.15 (1.08, 1.23)
Women Empowerment						
Decision maker for Child health-care: Herself	1	1	1	1	1	1
Decision maker for Child health-care: Joint	0.99 (0.97, 1.03)	0.87 (0.79, 0.95)*	0.98 (0.89, 1.07)	1.02 (0.93, 1.12)	0.99 (0.97, 1.02)	0.99 (0.97, 1.02)
Decision maker for Child health-care: Husband	0.97 (0.94, 1.01)	0.94 (0.84, 1.04)	0.99 (0.89, 1.10)	1.06 (0.94, 1.19)	0.98 (0.95, 1.01)	0.99 (0.96, 1.02)
Working Status: Unemployed	1	1	1	1	1	1
Working Status: Employed	0.99 (0.97, 1.02)	0.99 (0.89, 1.07)	0.93 (0.87, 0.99)*	0.78 (0.71, 0.84)	1.00 (0.97, 1.03)	0.98 (0.96, 0.99)
Education: None	1	1	1	1	1	1
Education: Primary	1.00 (0.97, 1.03)	1.05 (0.96, 1.14)	1.08 (0.99, 1.18)	0.97 (0.88, 1.07)	0.98 (0.95, 1.01)	0.99 (0.96, 1.01)
Education: Secondary	1.02 (0.99, 1.05)	1.02 (0.92, 1.12)	1.07 (0.98, 1.18)	0.99 (0.88, 1.11)	0.99 (0.97, 1.03)	0.99 (0.96, 1.02)
Education: Higher	1.08 (1.02, 1.14)*	1.20 (1.03, 1.41)*	1.08 (0.93, 1.25)	0.87 (0.72, 1.07)	1.07 (91.02, 1.11)*	1.05 (1.01, 1.09)
Partner's education: None	1		1	1	1	1
Partner's education: Primary	0.98 (0.95, 1.01)	1.07 (0.99, 1.17)	1.03 (0.95, 1.13)	1.10 (0.99, 1.22)	0.98 (0.96, 1.01)	0.98 (0.96, 1.01)
Partner's education: Secondary	1.02 (0.99, 1.06)	1.06 (0.96, 1.16)	1.06 (0.97, 1.16)	1.08 (0.96, 1.21)	0.98 (0.99, 1.05)	1.01 (0.98, 1.04)
Partner's education: Higher	1.10 (1.05, 1.15)*	1.18 (1.04, 1.34)	1.05 (0.93, 1.19)	1.10 (0.94, 1.29)	1.09 (1.06, 1.13)	1.09 (1.05, 1.13)*
Reproductive Health						
Respondent's age at first birth	1.01 (1.00, 1.01)*	1.01 (1.01, 1.01)*	0.99 (0.99, 1.00)	1.00 (0.99, 1.01)	1.01 (0.99, 1.02)	1.01 (1.00, 1.01)*
Contraceptive Method: None	1	1	1	1	1	1
Contraceptive Method: Folkloric	0.98 (0.82, 1.17)	0.96 (0.82, 1.11)	0.91 (0.43, 1.95)	1.16 (0.65, 2.07)	1.61 (0.95, 2.72)	0.97 (0.78, 1.21)
Contraceptive Method: Traditional	1.02 (0.98, 1.07)	1.00 (0.97, 1.03)	0.89 (0.79, 1.01)	0.85 (0.73, 0.99)*	0.95 (0.85, 1.07)	1.01 (0.98, 1.05)
Contraceptive Method: Modern	1.01 (0.99, 1.04)	0.99 (0.98, 1.02)	0.97 (0.91, 1.03)	0.95 (0.87, 1.03)	0.98 (0.92, 1.05)	0.99 (0.98, 1.02)
Aware of MR: No	1	1	1	1	1	1
Aware of MR: Yes	1.03 (0.99, 1.06)	1.02 (1.00, 1.04)*	1.27 (1.19, 1.35)*	0.93 (0.84, 1.03)	1.09 (1.01, 1.16)*	1.03 (1.01, 1.05)
Ideal number of Boys: None	1	1	1	1	1	1
Ideal number of Boys: One/two	0.98 (0.95, 1.00)	0.99 (0.97, 1.01)	1.06 (0.99, 1.14)	1.06 (0.96, 1.17)	1.13 (1.04, 1.23)*	0.99 (0.97, 1.01)
Ideal number of Boys: >=3	0.95 (0.85, 1.07)	1.00 (0.89, 1.13)	0.99 (0.69, 1.44)	0.91 (0.63, 1.32)	1.25 (0.85, 1.85)	0.99 (0.88, 1.11)
Ideal number of Boys: Faith decides	0.99 (0.91, 1.08)	0.99 (0.88, 1.11)	0.79 (0.05, 12.64)	0.93 (0.69, 1.24)	0.61 (0.44, 0.86)	0.83 (0.38, 1.81)

* at 5% level of significance

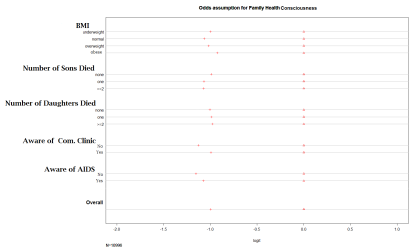


Figure 3.1: Odds assumption for Family Health Consciousness in 2007

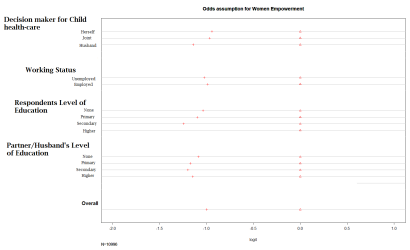


Figure 3.2: Odds assumption for Women Empowerment in 2007

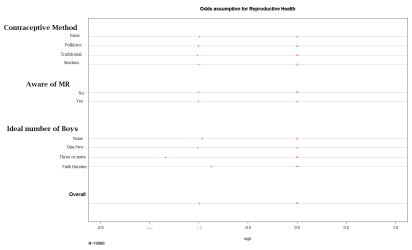


Figure 3.3: Odds assumption for Reproductive Health in 2007

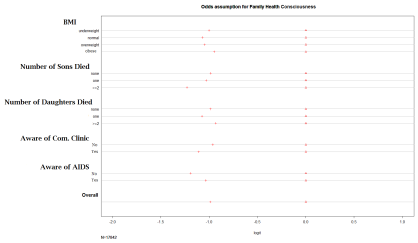


Figure 3.4: Odds assumption for Family Health Consciousness in 2011

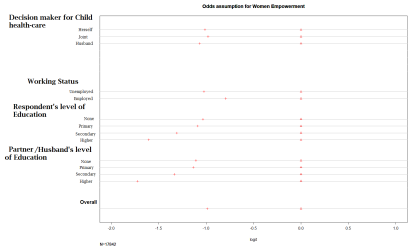


Figure 3.5: Odds assumption for Women Empowerment in 2011

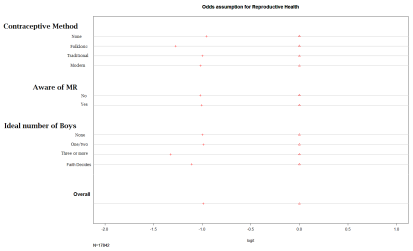


Figure 3.6: Odds assumption for Reproductive Health in 2011

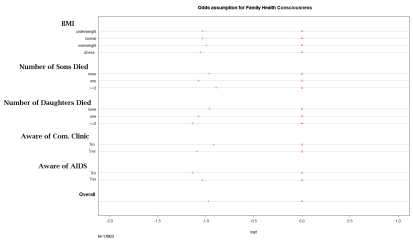


Figure 3.7: Odds assumption for Family Health Consciousness in 2014

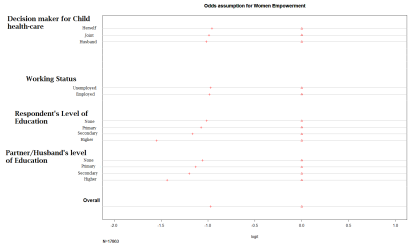


Figure 3.8: Odds assumption for Women Empowerment in 2014

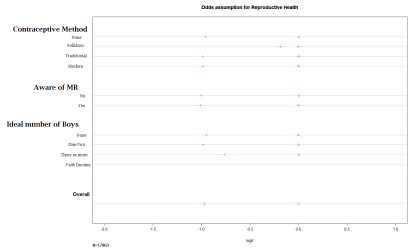


Figure 3.9: Odds assumption for Reproductive Health in 2014

The important take from the section is the application of POM, CRM and SDM and their goodness of fit (Table 3.3). The POM was the best-fitted model according to the $Pseudo - R^2$, whereas the SDM performed better in AIC and BIC scale. It also showed the most number of significant variables. According to the results, the POM seemed to find more significant variables compared to the CRM or SDM in all three different periods of time. The odds assumption of the POM were calculated for three covariate categories namely family health consciousness (FHC), women empowerment (WE) and reproductive health (RH), displayed in Figure 3.1 to Figure 3.9. Figure 3.1 to Figure 3.3 represented the odds assumptions of variables in 2007. Almost all the covariates seemed to maintain the odds assumption except the third scale of 'Ideal number of boys'. It is to be noted that sample size of 'Higher' educated class in the covariate 'Respondent's level of education' was very low (7.8%). Covariate categories of 2011 data sets were shown in Figure 3.4 to Figure 3.6, where the odds assumptions were moderately maintained for most of the variables. However, 'Education level' of both the respondent and their partner's did not fulfill it strictly. Similar pattern was observed in Figure 3.7 to Figure 3.9 which displayed the case of 2014. Considering the literature, the POM provided the most feasible outcomes. The results from the POM were consistent with previous studies on similar contexts. However, the goodness of fit statistics created a confusion (Table 3.3).

Table 3.3: Goodness of fit of the models

	2007			**2011**			**2014**		
Models	POM	CRM	SDM	POM	CRM	SDM	POM	CRM	SDM
Covariate category	$\boldsymbol{Pseudo - R^2}$								
Family Health Consciousness	0.0892839	0.0364856	0.0008662	0.0842164	0.0234319	0.0011381	0.0824582	0.0188594	0.0017967
Women Empowerment	0.1085427	0.0506828	0.0041048	0.1740152	0.1000024	0.0654576	0.1589352	0.0906695	0.0585122
Reproductive Health	0.1232335	0.1184091	0.1052463	0.1232949	0.1179843	0.1036988	0.1304737	0.1142604	0.1021853
	AIC								
Family Health Consciousness	31964.28	64906.29	14907.71	52437.76	104692.5	23229.43	52715.01	104383	24643.57
Women Empowerment	31293.01	64841.65	14903.04	47304.62	98050.5	22946.92	48331.55	98352.14	23256.49
Reproductive Health	30769.9	58051.09	13472.75	50199.32	93918.04	22034.55	49955.87	93804.83	22174.7
	BIC								
Family Health Consciousness	32081.16	65015.86	15002.68	52546.81	104793.7	23314.44	52824.08	104484.3	24729.27
Women Empowerment	31424.49	64965.83	15012.61	47435.92	98174.08	23055.05	48470.71	98483.56	23372.46
Reproductive Health	30870.64	58144.62	13551.89	50299.18	94010.22	22111.37	50055.78	93897.05	22251.56

This result in this chapter again shows the subjectivity involved in choosing models by the investigators. As suggested in the first paper (section 3.1), multiple models should be applied before concluding results from one model. Furthermore, goodness of fit statistics help to report and interpret results in accordance with the literature.

3.2 Effect of Sex and Age on Traumatic Brain Injury

The paper presented in this section was **accepted** in *Archives of Public Health* (reference ID: AOPH-D-17-00072) titled 'Effect of Sex and Age on Traumatic Brain Injury: A Geographical Comparative study' and scheduled for publication on 9 October 2017. RK Biswas was the primary contributor, where he conceptualized the study, collected the data, performed the statistical analysis, interpreted the findings and drafted the manuscript. The manuscript was critically reviewed and edited by E Kabir and R King. They supervised the whole process.

This paper addressed the second objective (section 1.6) of the thesis. We had two different data sets of head injury: a primary collected from Bangladesh and the multi-centre trial of CRASH data, collected from various parts of the world apart from Bangladesh. The objective was to assess the effect of sex and age on the recovery of TBI, and if these effects are different between Bangladesh and rest of the world. The results showed that age and sex significantly contributed to the TBI in CRASH data with old women being the worst victim of TBI. However, there was no such significant effects in Bangladesh. Figure 1 inside the following paper is the same Figure 2.2.5, mentioned in section 2.2.

Biswas *et al. Archives of Public Health* (2017) 75:43
DOI 10.1186/s13690-017-0211-y

Archives of Public Health

RESEARCH Open Access

Effect of sex and age on traumatic brain injury: a geographical comparative study

Raaj Kishore Biswas[1*], Enamul Kabir[2] and Rachel King[2]

Abstract

Background: Traumatic brain injury (TBI) is a much researched topic in medical health, which requires additional studies to understand various effects of demographic and geographic factors that can assist in developing the most effective treatments. Thousands of people of different ages are suffering from lifelong disabilities, either mild or severe, from TBI and the number is increasing. This study aims to increase our understanding of the effect of sex and age by applying five different statistical methods to evaluate the effect of these covariates on two independent TBI data sets representing patients from different geographical cohorts. A primary data was collected from Bangladesh and it was compared with CRASH (Corticosteroid Randomisation after Significant Head Injury) data, representing various countries around the world.

Methods: The outcome variable for TBI considered in this paper is Glasgow Outcome Scale, which is a four point scale. It was converted to a binary outcome scale for fitting of Fisher's exact test, a test of proportions and a binary linear model. For analyzing ordinal outcomes, the proportional odds model and the sliding dichotomy model were fitted. As the sample size of the Bangladeshi data set was small, parametric bootstrapping was applied for the consistency of results.

Results: Females were the worse sufferers of TBI compared to men, according to CRASH data set. The old (aged above 58 years) followed by adults (age 25 to 58) were the most vulnerable victims. Interaction effects concluded that old women tended to endure the worst outcomes of TBI. This conclusion came from the CRASH data set representing the world in general, whereas such effects were not present in the Bangladesh data set. Additional application of parametric bootstrapping for the smaller Bangladesh data set did not result into any significant outcome.

Conclusion: The effect of gender and age could be stronger in some countries than others which is driving the significance in CRASH and was not found in Bangladesh. It reflects the necessity of incorporating geographic patterns as well as demographic features of patients while developing treatments and designing clinical trials.

Keywords: Public health, Glasgow outcome scale, Health geography, Bangladesh, Ordinal outcome scale

Background

Identification of effective treatments for traumatic brain injury (TBI) has been the focus of much medical research in recent years [1]. Improved understanding of the role of sex and age will contribute to the development of more patient and geographic specific treatments. TBI is defined as an alteration in brain function, or other evidence of changed brain pathology, caused by an external force to the brain [2]. Alteration in brain function generally means any period of loss or a decreased level of consciousness (LOC). Although not all blows or jolts to the head result in TBI [3]. TBI is one of the most common forms of severe injury with a high death toll or life-long disabilities seen among patients. Among the injuries that occur due to TBI, the recorded deaths number more than 50,000 yearly in the USA [4]. Each year approximately 370,000 new cases of TBI are hospitalized in USA [5] and the figure is more than 100,000 for Europe [6]. Young people are the most common sufferers of TBI, resulting in long term disabilities which, in addition to the personal toll, affects both the work force and economy [7]. Expenditure on TBI related costs in the USA alone is estimated to be $17 billion per

*Correspondence: RaajKishore.Biswas@usq.edu.au
[1]Faculty of Health, Engineering and Sciences (HES), University of Southern Queensland, Darling Heights, Toowoomba QLD 4350, Australia
Full list of author information is available at the end of the article

year [8]. Severe and moderate forms of TBI, accidental or self-inflicted, are a major health and socioeconomic problem throughout the world [9]. Sex has been shown to be a key differentiating factor in many areas of medical research, and is often found to significantly interact with other predictor variables [10]. In the case of TBI patients it has been shown that the fatality rates are significantly higher for females than for males. Kraus et al. (2000) concluded that the mortality rate of women compared to men suffering from TBI was 1.28 times higher on average [11]. Moreover it was also found that even when death was not considered, women were 1.57 times more likely to suffer from post-traumatic symptoms than men. Klauber et al. (1981) also showed that fatality was higher for women compared to men in different age groups [12]. Even after one year of injury, severity of symptoms were more evident in women [13]. Several studies have also shown that women are prone to suffer more from TBI within one to three months of the traumatic incident [14–16]. On the other hand, the frequency of accidents leading to brain injury are more common in men than in women [17]. Males are more likely to be in recurrent accidents [11]; for example, motor cycle accidents are one of the most frequent causes of TBI in men [18]. Another common contributing factor leading to TBI incidents is the consumption of alcohol, which is more regularly consumed by men in general [19]. In summary, Farace et al. (2000) showed that in 17 out of 20 studies, which analyzed the effect of sex on TBI outcomes women suffered worse overall from TBI events [20].

It is generally accepted that the effect of diseases and injuries gradually worsens as age increases. TBI has shown to conform to this trend and different age groups are considered an important covariate in TBI studies [21]. Children have shown better recovery rates than older patients due to their greater degree of neuroplasticity [22]. Susman et al. (2002) found that mortality after a TBI event was approximately 24% in the elderly population while only 12.8% in other age groups [23]. Falls are the most common causes of brain injury in older patients and assaults or accidents in younger cohorts [24]. Even when older patients suffer comparatively minor head injuries and their overall injuries are seemingly less severe than non-elderly patients, they still have slower recovery rates and tend to experience more distress. Gómez et al. (2000) showed that the chance of an adverse outcome was 10 times higher for patients over 35 years of age compared to those aged between 15 to 25 years [21]. The large effect that age can play in long term outcomes for patients was shown in a study by Heiskanen et al. (1970), who found that less than 30% of patients aged 50 years or more went back to their former work, while more than 70% of patients under 20 years were able to go back to a normal life after their treatment [25]. Additionally, elderly patients had lower recovery rates than the young, while the young were more frequent sufferers of TBI [26].

Among older patient groups there is often a greater chance of co-morbidity occurring along with the primary disease or injury [27], as shown in studies of cardiovascular diseases [28], depression [29] and Alzheimer [30]. A positive association between depression and age following a TBI has been identified [31], indicating that older patients are more likely to suffer episodes of depression after head injury than younger patients. Guralnik and Jack (1996) identified a significant interaction effect between sex and age, with older women more likely to have higher prevalence of co-morbidity in contrast to older men [27]. Interestingly, sex difference had no impact on outcome scales if TBI was sustained by children, however for middle aged women TBI outcomes were significantly worse than for middle aged men. More elderly people suffered from TBI than middle aged people, however the difference between outcomes for elderly men and women (aged above 45) was less pronounced [19].

The aim of this paper is to investigate both the independent and interaction effects of age and sex on the TBI outcome scale, commonly known as the Glasgow Outcome Scale (GOS), (described in the following section). This study aims to increase our understanding of the effect of sex and age by applying five different statistical methods to evaluate the effect of these covariates on two independent TBI data sets representing patients from different geographical cohorts and find out the most vulnerable age-sex group for TBI.

Method

Data description

The first of the two data sets used in this study was the CRASH (Corticosteroid Randomisation after Significant Head Injury) data set which is comprised of data collected from TBI patients in a range of countries worldwide. The second data set, measuring the same variables as CRASH, was collected from TBI patients in Bangladesh, which was not one of the countries included in CRASH. These two data sets represent very different populations, with different levels of variation among a range of demographic and socio-economic variables not measured or included as covariates in this study. Of interest, was whether the different data sets suggested differences in the effects of age and sex on TBI for these different populations.

The CRASH data set was the result of a randomized controlled trial (ISRCTN74459797) [32]. This large trial was one of the most recent randomized trials monitoring the effect of corticosteroids on head injury and provided

a large data set; which was suitable for applying different statistical models on the ordinal outcome variable measuring changes in patients' post-treatment outcomes. Only the patients who were at least 16 years old and were observed whilst in hospital (in the absence of sedation) to have a GCS of 14 or less, and were within 8 hours of injury, were eligible for the trial entry. The CRASH collaboration includes data from various countries of Europe, Africa, South America, Asia and Oceania. The total number of patients in this data set was 10,800. Early results from the original CRASH study were published on 8 October 2004 (Lancet 2004;364:1321-28) and the 6-month follow-up results in May 2005 (Lancet 2005;365:1957-59) [33]. After removing cases with missing values, the final sample size was 7236. Results presented in this paper based on this data set do not distinguish between contributing countries. This Bangladesh data were collected from National Institute of Neuro Sciences and Hospital, Sher-e-Bangla Nagar, Dhaka in Bangladesh. The data is comprised of all brain injury patients treated in the hospital from May to September in 2015 with a total sample size 151. Patient information was collected from the hospital data base and cross checked with the resident physicians.

Glasgow Outcome Scale (GOS) was the outcome variable in both data sets. GOS is an ordinal variable commonly used to measure a recovering patients' neurological responses after some form of treatment, [34]. Although the TBI treatments applied in the CRASH and Bangladesh data sets were different, they were each applied consistently across sex and age groups within the data sets. There have been some recent adjustments to the GOS scale within the medical community; however, the general format from worst to best outcome scales are Death (D), Vegetative State (VS), Severe Disability (SD), Moderate Disability (MD) and Good Recovery (GR). In this study Death (D) and Vegetative State (VS) were merged into a single category (named Vegetative State (VS)) because the sample size of deaths in the data sets was small. The independent variables were sex and age. The grouping criteria of age has evolved over generations of research and varies due to differing research aims [35]. The following age groupings are commonly accepted and they have been utilized in this study: 'old' (greater than 59 years), 'adult' (in between 25 and 58 years), and 'young' (aging 0 to 24 years). A separate category with patients aged below 15 was not created as their proportion in either data set was very small.

Statistical methods

Frequencies for each level of the independent variables were calculated for each data set to provide a clear description of the data distributions. A cross tabulation of sex (two categories) and age (three categories) distribution with the GOS outcome variable (four categories) were also calculated. Five statistical models were applied as there is currently no single model that is considered the most robust approach when analyzing the ordinal outcomes of clinical trials [1, 9, 36, 37]. For the binary outcome analysis the GR & MD categories of the GOS scale were merged as favorable outcome and SD & VS were considered as unfavorable outcome. These binary levels of the outcome variable were created to analyze the effects of the covariates by applying Fisher's exact test, test of proportions and linear logistic regression model. To analyze the four point ordinal outcome scale of GOS, the proportional odds model and sliding dichotomy model were applied. All of the tests assessed the probability of a favorable outcome over a non-favorable or less favorable outcome which was consistently defined as the reference group. The multinomial regression model was not considered in this study as it does not provide one unique odds ratio for each category unlike other models. All statistical analyses were performed in R (version3.2.3).

All of the statistical methods were applied to both data sets and results were compared. As the primary data set from Bangladesh was small, the analysis was performed a second time implementing parametric bootstrapping with 1000 replications to attain more precision.

Binary outcome analysis

For assessing the significance of frequency distributions within a two-way contingency table, the common approach is to apply a chi-square goodness of fit test. However, this approach is only valid when expected frequency within cells is large [38]. Fisher's exact test, developed by R.A. Fisher [39], was applied to the collected data in Bangladesh due to insufficient expected values. The test is also valid for large samples as well allowing its application on comparatively bigger CRASH trial as well.

The test of proportions was applied to analyze the null hypothesis that the proportion of 'favorable outcome' results (probabilities of success) in several groups are similar [40]. It is an alternate to the Fisher's exact test and was applied here to consolidate the results from Fisher's test. Fisher's exact test and the test of proportions are appropriate methods only when the explanatory factor is also binary (e.g. sex).

The conventional binary linear logistic regression, or logit regression, first developed by D.R. Cox [41], is a popular model to analyze dichotomous forms of outcome variables. The logistic model is favored for its mathematical flexibility as well as clinically meaningful interpretation [42]. The linear logistic model is defined by Eq. 1,

$$\boldsymbol{logit}\pi_i = \log\frac{\pi_i}{1-\pi_i} = X^T\beta \tag{1}$$

where x_i is a vector measurement corresponding to covariates and dummy variables corresponding to factor

levels of the X covariate matrix and $\boldsymbol{\beta}$ is the parameter vector [43]. This model, referred in Eq. 1, is very widely used for analyzing data involving binary or binomial responses with several explanatory variables. It accommodates explanatory variables with more than two categories (e.g Age), providing a powerful technique analogous to multiple regression and ANOVA for continuous responses. The *glm* in R *MASS* Package was applied to fit this model.

Ordinal outcome analysis

The first method used to analyze the ordinal form of the GOS outcome variable was the proportional odds model. Naïve dichotomization of the full ordinal scale leads to loss of information and efficiency, when analyzing the outcomes. The proportional odds model is a popular choice for analyzing the full range of ordinal outcomes and avoiding the need for arbitrary dichotomization [44]. Where a random variable be $\boldsymbol{Y}$ with $\boldsymbol{J}$ categories and $\boldsymbol{\pi_1, \pi_2, \ldots, \pi_J}$ denote the respective probabilities, with $\boldsymbol{\pi_1 + \pi_2 + \cdots + \pi_J = 1}$. The cumulative logit model is defined by Eq. 2,

$$\log \frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J} = x_j^T \beta_j = \beta_{0j} + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1} \tag{2}$$

where the $\boldsymbol{x}$'s are the covariates and the $\boldsymbol{\beta}$'s are the unknown parameters with the intercept term $\boldsymbol{\beta_{0j}}$ (if exists) [43]. This model has a crucial odds assumption which claims that, the effects of the covariates $\boldsymbol{x_1, \ldots, x_{p-1}}$ are same for all categories of the logarithmic outcome scale, resulting in a constant $\boldsymbol{\beta}$ value. The method estimates a common odds ratio over all possible cut-offs of the outcome scale for a given change in category within the covariates [45]. Package *Polr* from R was applied to fit this model.

The sliding dichotomy model is a comparatively newer approach developed for clinical trials, particularly for TBI research [36]. This method is an improved version of the conventional logistic regression model. This model is assumed to provide the highest possible power and most robust results compared to the traditional methods in a number of scenarios. However, these scenarios are mostly limited to those cases when the probability of favorable outcomes is high [46]. Cases do exist where the fixed dichotomy and the proportional odds model performed better than the sliding dichotomy model [37]. Prior to analysis, outcome bands or successive dichotomous groups are created by segregating the fitted values (prognostic scores) from a binary logistic model [37]. Each band, displayed in Fig. 1, has its own reformed version of dichotomous 'favorable' and 'unfavorable' outcomes by combining a different subset of sequential outcomes from the original ordinal scale. The binary outcomes from all the bands are then compiled together and the traditional logistic regression model applied again to fit with available covariates, which were sex and age groups here. The *glm* in R *MASS* Package was applied to attain the fitted values as well as to analyze the complied favorable and unfavorable outcomes.

Results and discussion

Table 1 shows the frequency of female patients was proportionately higher in the Bangladeshi data (32.5%, $n = 102$) compared to the CRASH data (19%, $n = 7236$). Based on the age group categories chosen for this analysis, there were relatively more 'adults' in the CRASH data (59%, $n = 4286$) compared to the Bangladeshi data (40%, $n = 60$). In the Bangladeshi data, patients in the 'old' and 'adult' categories together represented the majority of the sample (75.5%, $n = 114$); however, 'old' category only comprised of 8.5% of the data in CRASH where the majority of patients were consisted of 'adult' and 'young'.

Sex

Table 2 displays the frequency distribution of patients by sex for each of the two forms of outcome variable, binary and ordinal.

Based on the binary outcome data (Table 2), the Bangladeshi data appeared to have a higher proportion of unfavorable outcomes for both males and females (24.5% & 24.5%, $n = 25$ & 12) compared to the CRASH data (16.3% & 20%, $n = 952$ & 286). The proportion of males and females did not vary much between the data sets for the GR and MD categories of the four point GOS ordinal scale; however, the SD and VS proportional differences were comparatively high.

Fisher's exact test and the test of proportions did not show any significant difference ($P > 0.05$) between sexes on the binary GOS measure (Table 3) for the primary data collected from Bangladesh. In contrast, the difference between sexes was found to be highly significant ($p < 0.001$) in the CRASH data for these same statistical methods. The odds indicate that females were 26% more likely to have unfavorable outcomes after treatment for TBI than men. In general terms, women were more prone to suffer unfavorable outcomes due to TBI than men demonstrated by global data. In addition, Table 3 presents the results of the binary logistic regression for dichotomous form of GOS and the proportional odds modeling and sliding dichotomy method for the four point ordinal GOS. The data from Bangladesh did not show any significant difference ($P > 0.05$) between males and females, neither in considering the binary GOS nor the four scale ordinal GOS. These results were consistent both with and without parametric bootstrapping; indicating that the smaller sample size of the Bangladeshi data (when compared to

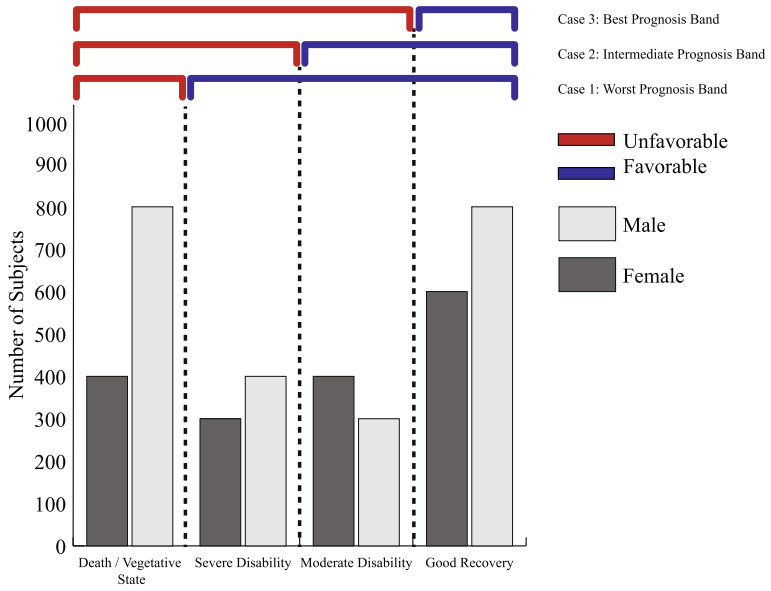


Fig. 1 Sliding Dichotomy model: binary bands are created from ordinal scales

CRASH) was not influencing the method performance or results. However, a significant difference was detected between males and females in the CRASH data for all the three methods ($P < 0.001$). The odds of unfavorable outcomes for women were 0.74 (0.64 $\sim$ 0.86), 0.91 (0.81 $\sim$ 1.03) and 0.97 (0.82 $\sim$ 1.14) in the logistic regression, the proportional odds model and the sliding dichotomy model respectively. These values agreed with previous tests and indicated that women have higher chance of having more suffrage from TBI, varying from 03 to 26%, compared to men.

Age

The data sets did not vary much in the proportion of different categories of age (Table 4). The 'young' category is comparatively higher in proportion in the CRASH data compared to the Bangladeshi data. The GR and MD groups had the higher proportion of samples for both Bangladeshi data and CRASH data.

The binary logistic model, the proportional odds model and the sliding dichotomy model were applied to fit age groups with GOS. Both the binary regression and proportional odds model agreed that the 'adult' and the 'young' groups were significantly different ($P < 0.001$) from the 'old' in CRASH data (Table 5). However, no significance was found in the sliding dichotomy model for the age groups in either data sets. According to the binary logistic model, applied in CRASH data, adults were 2.1 times and youths were 3.8 times more likely to have favorable outcomes in TBI compared to the olds. The Proportional odds model determined the likeliness of favorable outcomes in case of adults and youths were 1.6 and 2.5 times higher as TBI patients than olds respectively. These gave a summary stating olds were the worst victims of TBI. A contrasting conclusion was attained from the Bangladeshi data, both for normal and bootstrapping procedures. There were no mentionable differences between the age groups over the TBI outcomes in Bangladesh for the three models.

Table 1 Frequency distribution of sex and age in both data sets

Covariates	Levels	Bangladeshi data (151)	CRASH data (7236)
Gender	Female	102(67.5%)	5856 (80.9%)
	Male	49 (32.5%)	1380 (19.1%)
Age Groups	Old (>59)	54 (35.8%)	616 (8.5%)
	Adult (25∼58)	60 (39.7%)	4286 (59.2%)
	Young (0∼25)	37 (24.5%)	2334 (32.2%)

Sex and age

The demography of age groups and sex in each data set in cross-frequency distribution (Table 6) showed the proportions of adults varied between the data sets. The 'old' group and the 'young' group appeared to be similarly proportioned in both male and female. The percentage of female patients in both data sets were comparatively lower than males.

Sex, age and their interaction effect were fitted in the same model for the binary logistic regression, proportional odds model and the sliding dichotomy method. The results obtained from the Bangladeshi primary data set are displayed in Table 7 and CRASH data in Table 8. None

Table 2 Distribution of sex by GOS in binary and four point ordinal form

	Bangladeshi data (151)		CRASH data (7236)	
Outcome scales	Male (% among male)	Female(% among female)	Male (% among male)	Female(% among female)
Favorable (GR & MD)	77 (75.5%)	37 (75.5%)	4904 (83.7%)	1094 (79.3%)
Unfavorable (SD & VS)	25 (24.5%)	12 (24.5%)	952 (16.3%)	286 (20.7%)
GR	55 (53.9%)	23 (46.9%)	3511 (60%)	822 (59.6%)
MD	22 (21.6%)	14 (28.6%)	1393 (23.8%)	272 (19.7%)
SD	06 (5.9%)	07 (14.3%)	747 (12.8%)	209 (15.1%)
VS	19 (18.2%)	05 (10.2%)	205 (3.5%)	77 (5.6%)

of the tests displayed any significant covariates or interactions in the Bangladeshi data. In contrast, CRASH data showed the significance of sex (reference group 'male') and the two age groups: 'adult' and 'young' (reference group 'old') in the logistic regression model and the proportional odds model. The interaction effect between 'sex (female) and age (adult)' along with 'sex (female) and age(young)' were found to be significant ($p\ value < 0.05$). As the interaction effects were detected only in the CRASH data set, further analysis were conducted to assess the significance of interaction effects in that data set, displayed in Table 9. The interaction models were fitted for the binary regression model and the proportional odds model only, as the sliding dichotomy model failed to show any significant interactions.

Both the tests showed adult male, adult female, young male and young female groups face less severity in TBI compared to old males, however this was not true for old females. Additionally, the binary regression model showed that old females had a 47% lesser chance of favorable outcome than old men. Both adult and young females suffered more than adult and young men. However, adult women and young women faced 1.5 and 2.8 times better outcomes than old men respectively. The odds of favorable outcome were 1.7 and 3.0 for adult and young men respectively compared to old men showing the faster recovery by youths and adults in contrast with the older patients. Although the difference in odds between adult men and adult women were comparatively closer in the proportional odds model, the gap between young males and females were evident in both models but in opposite directions. CRASH, the multi country data which sampled a wider and more varied population, showed a significant interaction of age and sex on TBI. This suggests that the effect of gender and age could be stronger in some countries than others, which is driving the significance in CRASH that was not found in Bangladesh.

The worst victims of TBI, sequentially in descending order were old females, old men, adult women, adult men and then the young. This conclusion was derived from the CRASH data, which was a multi-country data set. These effects were not evident in the data from Bangladesh. The sensitivity of the human brain is higher than other organs and therefore it is likely that effective treatment

Table 3 Statistical models on GOS by sex

Tests		Bangladeshi data	Bootstrap of Bangladeshi data	CRASH data
Fisher's exact test	*P*-value	1		<0.001
	CI	0.427 ~ 2.441		0.639 ~ 0.863
	Odds	1.001		0.743
Test of proportions	*P*-value	0.99		<0.001
	CI	-0.174 ~ 0.173		-0.074 ~ -0.023
Binary logistic model	*P*-value	0.998	0.998	<0.001
	CI	0.453 ~ 2.211	0.453 ~ 2.211	0.641 ~ 0.861
	Odds	1.001	1.001	0.743
Proportional odds model	CI	0.477 ~ 1.682	0.477 ~ 1.682	0.813 ~ 1.025
	Odds	0.896	0.896	0.913
Sliding dichotomy model	*P*-value	0.841	0.841	0.688
	CI	0.441 ~ 2.735	0.441 ~ 2.736	0.816 ~ 1.144
	Odds	1.098	1.098	0.966

The reference level for sex was 'male'

Table 4 Distribution of GOS (as binary outcome) over age

	Bangladeshi data (151)			CRASH data (7236)		
Outcome scales	Old (% among Old)	Adult (% among adult)	Young (% among Young)	Old (% among Old)	Adult (% among adult)	Young (% among Young)
Favourable (GR & MD)	12 (22.2%)	17 (28.3%)	8 (21.6%)	197 (32%)	781 (18.2%)	260 (11.1%)
Unfavourable (SD & VS)	42 (77.8%)	43 (71.7%)	29 (78.4%)	419 (68%)	35.05 (81.8%)	2074(88.9%)
GR	26 (48.1%)	29 (48.3%)	23 (62.2%)	313 (50.8%)	2444 (54%)	1576 (67.5%)
MD	16 (29.6%)	14 (23.3%)	6 (16.2%)	106 (17.2%)	1061 (24.8%)	498 (21.3%)
SD	5 (9.3%)	6 (10%)	2 (5.4%)	138 (22.4%)	610 (14.2%)	208 (8.9%)
VS	7 (13%)	11 (18.3%)	6 (16.2%)	59 (9.6%)	171 (04%)	52 (2.2%)

Table 5 Statistical tests on age groups vs GOS

Tests		Bangladeshi data		Bootstrap of Bangladeshi data		CRASH data	
		Adult	Young	Adult	Young	Adult	Young
Binary logistic model	*P*-value	0.455	0.946	0.455	0.946	<0.001	<0.001
	CI	0.308 ∼ 1.695	0.376 ∼ 2.849	0.308 ∼ 1.695	0.376 ∼ 2.849	1.751 ∼ 2.542	3.031 ∼ 4.640
	Odds	0.723	1.034	0.723	1.036	2.110	3.751
Proportional odds model	CI	0.452 ∼ 1.761	0.650 ∼ 3.331	0.452 ∼ 1.761	0.649 ∼ 3.331	1.331 ∼ 1.849	0.855 ∼ 7.299
	Odds	0.892	1.471	0.892	1.471	1.569	2.498
Sliding dichotomy model	*P*-value	0.815	0.955	0.815	0.955	0.513	0.332
	CI	0.338 ∼ 2.349	0.334 ∼ 3.19	0.337 ∼ 2.349	0.334 ∼ 3.198	0.714 ∼ 1.18	0.673 ∼ 1.143
	Odds	0.891	1.0333	0.891	1.0333	0.919	0.877

The reference level for age group was 'old'

Table 6 Cross table of sex and age in both data sets

	Bangladeshi data (151)		CRASH data (7236)	
Sex	Male (% among the age group)	Female (% among the age group)	Male (% among the age group)	Female (% among the age group)
Age groups				
Old (>59)	37 (68.5%)	17 (31.5%)	411 (66.7%)	205 (33.3%)
Adult (25∼58)	39 (65%)	21 (35%)	3534 (82.5%)	752 (17.5%)
Young (15∼25)	26 (70.3%)	11 (29.7%)	1911 (81.9%)	423 (18.1%)

Table 7 Statistical tests on age groups and sex with interactions for Bangladesh data

Bangladeshi data						
Tests		Sex (Female)	Adult	Young	Sex*Adult	Sex*Young
Binary logistic model	*P*-value	0.876	0.681	0.627	0.774	0.289
	CI	0.229 ∼ 3.517	0.276 ∼ 2.315	0.233 ∼ 2.407	0.128 ∼ 4.613	0.300 ∼ 56.222
	Odds	0.897	0.800	0.749	0.769	4.109
Proportional odds model	CI	5.0218 ∼ 0.063	0.273 ∼ 1.972	0.350 ∼ 2.528	0.408 ∼ 6.831	0.649 ∼ 24.912
	Odds	0.563	0.734	0.941	1.669	4.020
Sliding dichotomy model	*P*-value	0.516	0.913	0.716	0.568	0.554
	CI	0.323 ∼ 9.477	0.334 ∼ 3.402	0.334 ∼ 4.931	0.061 ∼ 4.653	0.038 ∼ 5.795
	Odds	1.750	1.067	1.283	0.531	0.468

The reference level for sex was 'male' and age group was 'old'

Table 8 Statistical tests on age groups and sex with interactions for CRASH data

CRASH data						
Tests		Sex (Female)	Adult	Young	Sex*Adult	Sex*Young
Binary logistic model	*P*-value	<0.001	<0.001	<0.001	0.018	0.024
	CI	0.372 ∼ 0.752	1.367 ∼ 2.181	2.339 ∼ 3.935	1.089 ∼ 2.447	1.076 ∼ 2.819
	Odds	0.529	1.727	3.034	1.633	1.742
Proportional odds model	CI	0.415 ∼ 0.796	1.073 ∼ 1.598	1.671 ∼ 2.545	1.214 ∼ 2.496	1.252 ∼ 2.759
	Odds	0.574	1.309	2.062	1.741	1.858
Sliding dichotomy model	*P*-value	0.615	0.803	0.697	0.594	0.37
	CI	0.682 ∼ 1.905	0.712 ∼ 1.302	0.686 ∼ 1.286	0.489 ∼ 1.505	0.423 ∼ 1.378
	Odds	1.141	0.962	0.939	0.858	0.763

The reference level for sex was 'male' and age groups was 'old'

Table 9 Interaction effects for CRASH data

CRASH data						
Tests		Male*Adult	Male*Young	Female*Old	Female*Adult	Female*Young
Binary logistic model	*P*-value	<0.001	<0.001	<0.001	<0.001	<0.001
	CI	1.367 ∼ 2.181	2.339 ∼ 3.935	0.372 ∼ 0.752	1.126 ∼ 1.975	1.937 ∼ 4.031
	Odds	1.727	3.034	0.529	1.490	2.794
Proportional odds model	CI	1.073 ∼ 1.598	1.671 ∼ 2.545	0.415 ∼ 0.796	1.035 ∼ 1.656	1.668 ∼ 2.904
	Odds	1.309	2.062	0.574	1.309	2.201

The reference level considered here was Male*Old

of brain injury will require greater patient specificity. Inclusion of additional covariates measuring other patient demographic features, as well as information about the type of accident that has resulted in a patient's TBI, may improve understanding of why older females are suffering more than others. This paper highlights the necessity of incorporating geographic patterns as well as demographic features of patients while developing treatments and designing clinical trials.

Conclusions

This paper aimed to understand how demographic features, particularly sex and age, affect treatments of TBI and furthermore if location variation contributes to differences in these effects All results consistently indicated that there was no significant difference in GOS as a measure of TBI by either sex or age groups in the Bangladeshi sample. Additionally, no interactions between sex or age categories were found to be computationally significant. A clear distinction was found between males and females within the international CRASH data set, with females generally having worse outcomes. Significant differences were also found between some age groups within this data, with elderly patients more likely to suffer negative outcomes than patients within the adult and young age categories. Interaction effects identified for this data indicated that old women appeared to show the worst outcomes followed by old men.

From the analysis of these two data sets it appears that while sex and age were not strong covariates based on the Bangladeshi data, they were both significantly associated with GOS outcomes in the CRASH data. This suggests that head injury in Bangladesh and/or the impact of demographic factors on outcomes in Bangladesh may be different, or less important, than these factors in the rest of the world. Country wise analysis of the CRASH data is needed to determine if these results are common to all or most contributing countries in the CRASH data set or whether these results are influenced by one or a few countries only. This was not possible in the current study as country of origin information was not included in the available CRASH data. Furthermore, age in a continuous scale might provide additional information in future studies. The analysis of more health and demographic variables such as previous disease history particularly neurological or psychiatric problems, immunity level, mental health status, marital status, and workplace stress would help to clarify the recovery profile of patients.

Abbreviations

CRASH: Corticosteroid randomisation after significant head injury; GOS: Glasgow outcome scales;GR: Good recovery; MD: Moderate disability; SD: Severe disability; TBI: Traumatic brain injury; VS: Vegetative State

Acknowledgements

The authors would like to acknowledge CRASH (Corticosteroid Randomisation after Significant Head Injury) who made their data available for free, hence works can be done in this sector very swiftly. We would like to express our gratitude to the National Institute of Neurosciences & Hospital in Dhaka, Bangladesh who provided the primary data. And of course appreciation towards Faculty of Health, Engineering and Sciences (HES) of University of Southern Queensland for the technical supports.

Funding

Not applicable for this paper as there was no funding for this project.

Availability of data and materials

The CRASH data is available online. However, the data of Bangladesh cannot be shared without the consent of Ethical Review Committee of National Institute of Neuro Sciences and Hospital.

Authors' contributions

RKB conceptualized the study, compiled the data, performed the statistical analysis, interpreted the findings and drafted the manuscript. He also collected the data from Bangladesh. The manuscript was critically reviewed and edited by EK and RK. They supervised the whole process. All authors helped to finalize the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

To different data sets were used in this paper. The first one, CRASH data, is a secondary data set which is available online (http://www.crash.lshtm.ac.uk/). The second data set was collected from Bangladesh by one the authors. It was approved by 'Ethical Review Committee of National Institute of Neuro Sciences and Hospital' in Dhaka, Bangladesh. No indefinable information like names of patients were not collected. The approval letter have been attached with the document. Written consents from the subjects were not taken by the authors as it was dealt by the ethical committee of the hospital. The data was collected as part of a long project comparing the Sliding dichotomy model with other models in clinical trials. This paper is a part of that project. The contact of the Chairman of Ethical Review Committee:
Professor Md. Azharul Hoque
Email: azharul59@yahoo.com
Contact: +88 02 91373 ext: 209

Consent for publication
Approval has been taken from 'Ethical Review Committee of National Institute of Neuro Sciences and Hospital'.

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

Publisher's Note

Author details
[1]Faculty of Health, Engineering and Sciences (HES), University of Southern Queensland, Darling Heights, Toowoomba QLD 4350, Australia. [2]School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Darling Heights, Toowoomba QLD 4350, Australia.

Received: 21 April 2017 Accepted: 27 June 2017
Published online: 09 October 2017

References

1. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the impact project. Clin Trials. 2010;7(1):44–57.
2. Menon DK, Schwab K, Wright DW, Maas AI, et al. Position statement: definition of traumatic brain injury. Arch Phys Med Rehabil. 2010;91(11): 1637–40.
3. Faul M, Xu L, Wald MM, Coronado VG. Traumatic Brain Injury in the United States. Atlanta: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010.
4. Tiesman HM, Konda S, Bell JL. The epidemiology of fatal occupational traumatic brain injury in the us. Am J Prev Med. 2011;41(1):61–7.
5. Reeves RH, Beltzman D, Killu K. Implications of traumatic brain injury for survivors of sexual abuse: A preliminary report of findings. Rehabil Psychol. 2000;45(2):205.
6. Maas AI, Menon DK, Lingsma HF, Pineda JA, Sandel ME, Manley GT. Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. J Neurotrauma. 2012;29(1):32–46.
7. Rutland-Brown W, Langlois JA, Thomas KE, Xi YL, et al. Incidence of traumatic brain injury in the united states, 2003. J Head Trauma Rehabil. 2006;21(6):544.
8. National Center for Injury Prevention and Control (US). Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem. Atlanta: Centers for Disease Control and Prevention; 2003.
9. Maas AI, Marmarou A, Murray GD, Teasdale SGM, Steyerberg EW. Prognosis and clinical trial design in traumatic brain injury: the impact study. J Neurotrauma. 2007;24(2):232–8.
10. Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, Kapucu OL, Kluge A, Knudsen GM, Koulibaly PM, et al. European multicentre database of healthy controls for [123i] fp-cit spect (enc-dat): age-related effects, gender differences and evaluation of different methods of analysis. Eur J Nucl Med Mol Imaging. 2013;40(2):213–27.
11. Kraus JF, Peek-Asa C, McArthur D. The independent effect of gender on outcomes following traumatic brain injury: a preliminary investigation. Neurosurg Focus. 2000;8(1):1–7.
12. Klauber MR, Marshall LF, Barrett-Connor E, Bowers SA. Prospective study of patients hospitalized with head injury in san diego county, 1978. Neurosurgery. 1981;9(3):236–41.
13. Rutherford W, Merrett J, McDonald J. Symptoms at one year following concussion from minor head injuries. Injury. 1979;10(3):225–30.
14. Bazarian JJ, Atabaki S. Predicting postconcussion syndrome after minor traumatic brain injury. Acad Emerg Med. 2001;8(8):788–95.
15. Fenton G, McClelland R, Montgomery A, MacFlynn G, Rutherford W. The postconcussional syndrome: social antecedents and psychological sequelae. Br J Psychiatr. 1993;162(4):493–7.
16. Ponsford J, WILLMOTT C, Rothwell A, Cameron P, Kelly AM, Nelms R, Curran C, Ng K. Factors influencing outcome following mild traumatic brain injury in adults. J Int Neuropsychol Soc. 2000;6(05):568–79.
17. Rimel RW, Giordani B, Barth JT, Boll TJ, Jane JA. Disability caused by minor head injury. Neurosurgery. 1981;9(3):221–8.
18. Marion DW, Penrod LE, Kelsey SF, Obrist WD, Kochanek PM, Palmer AM, Wisniewski SR, DeKosky ST. Treatment of traumatic brain injury with moderate hypothermia. N Engl J Med. 1997;336(8):540–6.
19. Pentland B, Jones PA, Roy CW, Miller JD. Head injury in the elderly. Age Ageing. 1986;15(4):193–202.
20. Farace E, Alves WM. Do women fare worse: a metaanalysis of gender differences in traumatic brain injury outcome. J Neurosurg. 2000;93(4): 539–45.
21. Gómez PA, Lobato R, Boto G, De la Lama A, González P, De la Cruz J. Age and outcome after severe head injury. Acta Neurochir. 2000;142(4): 373–81.
22. Carroll L, Cassidy JD, Peloso P, Borg J, Von Holst H, Holm L, Paniak C, Pépin M. Prognosis for mild traumatic brain injury: results of the who collaborating centre task force on mild traumatic brain injury. J Rehabil Med. 2004;36(0):84–105.
23. Susman M, DiRusso SM, Sullivan T, Risucci D, Nealon P, Cuff S, Haider A, Benzil D. Traumatic brain injury in the elderly: increased mortality and worse functional outcome at discharge despite lower injury severity. J Trauma Acute Care Surg. 2002;53(2):219–24.
24. Mosenthal AC, Livingston DH, Lavery RF, Knudson MM, Lee S, Morabito D, Manley GT, Nathens A, Jurkovich G, Hoyt DB, et al. The effect of age on functional outcome in mild traumatic brain injury: 6-month report of a prospective multicenter trial. J Trauma Acute Care Surg. 2004;56(5): 1042–8.
25. Heiskanen O, Sipponen P. Prognosis of severe brain injury. Acta Neurologica Scandinavica. 1970;46(3):343–8.
26. Ghajar J. Traumatic brain injury. The Lancet. 2000;356(9233):923–9.
27. Guralnik JM. Assessing the impact of comorbidity in the older population. Ann Epidemiol. 1996;6(5):376–80.
28. Lakatta EG. Age-associated cardiovascular changes in health: impact on cardiovascular disease in older persons. Heart Fail Rev. 2002;7(1): 29–49.
29. Seeman TE, Guralnik JM, Kaplan GA, Knudsen L, Cohen R. The health consequences of multiple morbidity in the elderly the alameda county study. J Aging Health. 1989;1(1):50–66.
30. van Reekum R, Simard M, Cohen T. The prediction and prevention of alzheimer's disease–towards a research agenda. J Psychiatry Neurosci. 1999;24(5):413.
31. Glenn MB, O'Neil-Pirozzi T, Goldstein R, Burke D, Jacob L. Depression amongst outpatients with traumatic brain injury. Brain Inj. 2001;15(9): 811–8.
32. Collaborators CT, et al. Final results of mrc crash, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury outcomes at 6 months. The Lancet. 2005;365(9475):1957–9.
33. CRASH. CRASH Corticosteroid Radomisation after Significant Head Injury. 2005. http://www.crash.lshtm.ac.uk. Accessed 5 Jan 2015.
34. Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. The Lancet. 1975;305(7905):480–4.
35. Eisenstadt SN. From Generation to Generation: Age Groups and Social Structure. Piscataway: Transaction Publishers; 1956.
36. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, Marmarou A, Mendelow AD, Steyerberg EW, et al. Design and analysis of phase iii trials with ordered outcome scales: the concept of the sliding dichotomy. J Neurotrauma. 2005;22(5):511–7.
37. Ilodigwe D, Stat M, Murray GD, Kassell NF, Torner J, Kerr RS, Molyneux AJ, Macdonald RL. Sliding dichotomy compared with fixed dichotomization of ordinal outcome scales in subarachnoid hemorrhage trials: Clinical article. J Neurosurg. 2013;118(1):3–12.
38. Fisher RA. On the interpretation of χ 2 from contingency tables, and the calculation of p. J R Stat Soc. 1922;85(1):87–94.
39. Upton GJG. Fisher's exact test. J R Stat Soc Ser A (Stat Soc). 1992;155: 395–402.
40. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22(158):209–12.
41. Cox DR. Regression models and life-tables. J R Stat Soc Ser B (Methodol). 1972;34:187–220.

42. Hosmer Jr DW, Lemeshow S. Applied Logistic Regression. Milton: John Wiley & Sons; 2004.
43. Dobson AJ, Barnett A. An Introduction to Generalized Linear Models. Boca Raton: CRC press; 2008.
44. Liu I, Mukherjee B, Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL. Proportional Odds Model. Encyclopedia of Statistical Sciences: Wiley Online Library; 2004. http://dx.doi.org/10.1002/0471667196.ess7271.
45. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, Maas A, Steyerberg EW, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. Crit Care. 2011;15(3):127.
46. Price M, Hertzberg V, Wright DW. Does the sliding dichotomy result in higher powered clinical trials for stroke and traumatic brain injury research? Clin Trials. 2013;10(6):924–34.

3.3 Type I error analysis

We are currently compiling a paper on the third objective of the this study (section 1.6) titled 'Sliding Dichotomy compared with Fixed Dichotomization of Ordinal Outcome Scales: Does it control Type I Error?'. This paper explores the weakness of the sliding dichotomy model. The type I error rate of the binary logistics model, proportional odds model and sliding dichotomy model were compared by devising eight different clinical trial scenarios (explained below) and sample sizes ranging from 150 to 2500. Data set from CRASH, a multi-centre TBI trial, was used to analyze the scenarios. The results of the paper are explained in this section. However, the following results are obtained from 1000 simulations, which are not adequate for an academic paper. We are currently running 10,000 simulations, which will be reported in the paper. However, we do not think the following interpretation will vary due to the simulation numbers.

Quantification and interpretation of type I error should be done carefully and delicately. Both $p-value$ and level of significance (α) are related to the type I error rate. A statistical test with size α is expected (at least asymptotically) to have type I error rate equal to α. However, $p-value$ provides the level of false positives found in a specific sample when the test is applied. That is why if the $p-value$ is below the significance level (α), we consider null hypothesis to be rejected and conclude that there is significant treatment effect. When null hypothesis is true, that is no significant treatment effect, $p-value$ allows the computation of type I error. In this case, if the $p-value$ is below then asymptotic level (α) then we suspect the existence of type I error or false positive based on our prior knowledge on true treatment effect from the population. We applied this property to quantify the error rate in this study. We simulated different sample sizes multiple times to acquire the $p-values$ from a known population and compiled the number of samples that has $p-values$ less the 5% threshold (α). From there, the cumulative proportion of the type I error was calculated.

The sliding dichotomy approach was developed as an applied method, without assessment of theoretical considerations. It is a natural expectation of medical researchers applying the SDM that this method will maintain type I error rate in-line with the traditional statistical methods, if not more. We wanted to investigate that expectation, and determine whether the SDM maintains type I error rate when applied to a data set from CRASH under several simulated scenarios. Different scenarios were simulated to recreate various practical environments of clinical trials, to evaluate the performance of each method in controlling the cumulative proportion of the type I error.

To design the study, twelve samples of size 150, 180, 300, 450, 600, 900, 1200, 1500, 1800, 2100, 2400 and 2700 were drawn from the 7,236 full cases in the CRASH data. Patients in each sample were randomly assigned to one of the artificially generated 'Treatment A' or 'Treatment B' in two different ways. Patients were first randomly assigned to one of the two treatments in equal proportions and then reassigned with 30% in Treatment A and 70% in Treatment B. The grouping of patients to form prognosis bands for the sliding dichotomy method was also performed in two different ways. In the first instance, equal numbers of patients were allocated to each prognosis band, based on the sorted fitted values (see section 2.4) (Figure 2.1). In the second instance the proportion of patients allocated to the 'Worst', 'Intermediate' and 'Best' prognosis bands were 50%, 30% and 20% of the total sample size respectively. In both scenarios, patient GOS was then classified as either favorable or unfavorable within each band (Figure 2.1).

Each sample was analyzed using two different sets of covariates in each model. In the first approach, three covariates from CRASH (age, sex and Glasgow Coma Scale motor response) plus the created treatment variable were included in each analysis. In the second approach, only the treatment variable was used as a covariate.

These variations in treatment levels, GOS allocation and in covariates assessed, resulted in eight ($2 \times 2 \times 2 = 8$) different scenarios for analysis across each of the initial 12 sample sizes: (1) four covariates, equal treatment sizes and equal band sizes; (2) four covariates, equal treatment sizes and unequal band sizes; (3) four covariates, unequal treatment sizes and equal band sizes; (4) four covariates, unequal treatment sizes and unequal band sizes; (5) one covariate, equal treatment sizes and equal band sizes; (6) one covariate, equal treatment sizes and unequal band sizes; (7) one covariate, unequal treatment sizes and equal band sizes; (8) one covariate, unequal treatment sizes and unequal band sizes.

Each sample was fitted with the binary logistic model, the proportional odds model and the sliding dichotomy model. The process was simulated 1000 times for each sample and scenario combination, and the resulting *p-values* were proportioned based on greater or lower than 0.05 to calculate the threshold of 5% for the type I error rate. Statistical power was also quantified in each case and it was averaged for every model in each scenario. We simulated treatments in a way so that the null (H_0) hypothesis is not rejected, that there is no significant effect of the treatment. The results from the whole data set (n = 7236) showed the simulation was correct (Table 3.4).

Table 3.4: Assessment of the null hypothesis for the whole CRASH data set

Model scenario	**Statistical Model**	**P-value**
Model 1: Four covariates, equal treatment sizes and equal band sizes	Binary logistic	0.5352
	Proportional odds	0.2244
	Sliding dichotomy	0.9565
Model 2: Four covariates, equal treatment sizes and unequal band sizes	Binary logistic	0.5352
	Proportional odds	0.2244
	Sliding dichotomy	0.7470
Model 3: Four covariates, unequal treatment sizes and equal band sizes	Binary logistic	0.7962
	Proportional odds	0.4475
	Sliding dichotomy	0.1887
Model 4: Four covariates, unequal treatment sizes and unequal band sizes	Binary logistic	0.7962
	Proportional odds	0.4475
	Sliding dichotomy	0.4736
Model 5: One covariate, equal treatment sizes and equal band sizes	Binary logistic	0.842
	Proportional odds	0.5541
	Sliding dichotomy	0.8930
Model 6: One covariate, equal treatment sizes and unequal band sizes	Binary logistic	0.8420
	Proportional odds	0.5541
	Sliding dichotomy	0.6830
Model 7: One covariate, unequal treatment sizes and equal band sizes	Binary logistic	0.9210
	Proportional odds	0.6395
	Sliding dichotomy	0.1850
Model 8: One covariate, unequal treatment sizes and unequal band sizes	Binary logistic	0.9210
	Proportional odds	0.6395
	Sliding dichotomy	0.4780

It is a general expectation that the results generated from samples will be close to the population if the sample size is high (Hernandez et al., 2006; Maas and Hox, 2005). With higher sample sizes, the distance between fitted values and observed values should be low (less error). Results from Table 3.4 shows the rejection of null hypothesis in our *population* with P-values more than 0.15; it is a natural expectation that higher sample sizes will provide homogeneous results, where higher bias might appear in smaller sample sizes. That is, the cumulative proportion of false results ($P - value < 0.05$) or type I error will decrease with higher sample sizes or at least a particular pattern will appear in repeated simulations.

The comparison of the type I error rates (in %) among the three models and eight different scenarios was done graphically by plotting the type I error rates against sample sizes in

ascending order, from 150 to 2700. A plot comparing the performance of the three statistical methods was produced for each of the eight scenarios (Figure 3.10 to Figure 3.17) to visualize the change of error rate by sample size for each experimental scenario. In each Figure, the red, green and blue lines represent the binary logistic, proportional odds and sliding dichotomy model respectively.

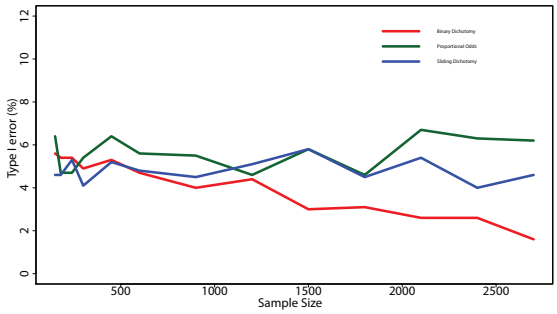


Figure 3.10: Type I error in scenario 1: Four covariates, equal treatment sizes and equal band sizes

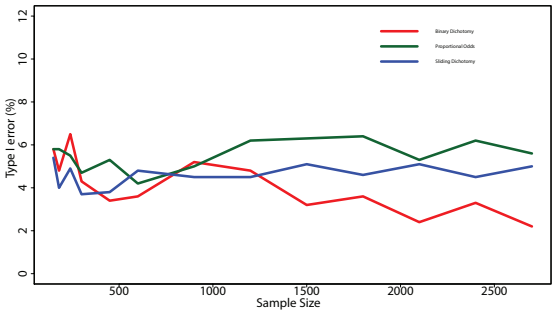


Figure 3.11: Type I error in scenario 2: Four covariates, equal treatment sizes and unequal band sizes

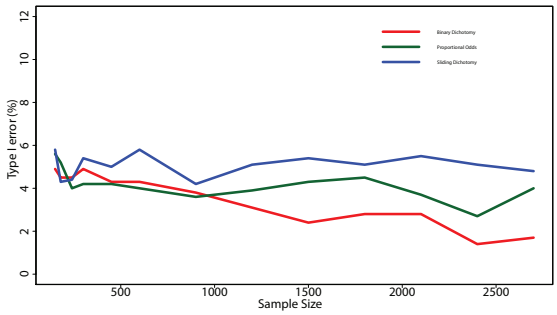


Figure 3.12: Type I error in scenario 3: Four covariates, unequal treatment sizes and equal band sizes

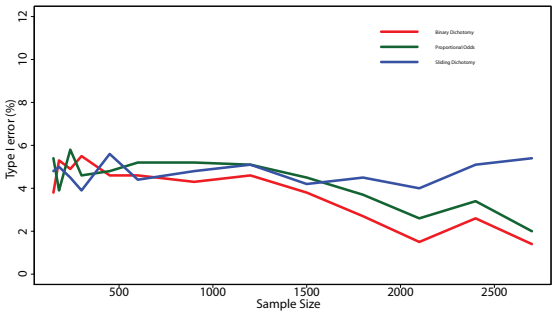


Figure 3.13: Type I error in scenario 4: Four covariates, unequal treatment sizes and unequal band sizes

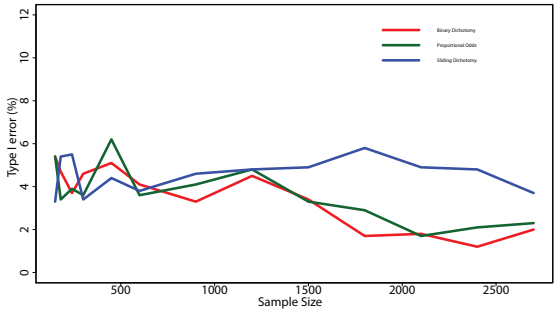


Figure 3.14: Type I error in scenario 5: One covariate, equal treatment sizes and equal band sizes

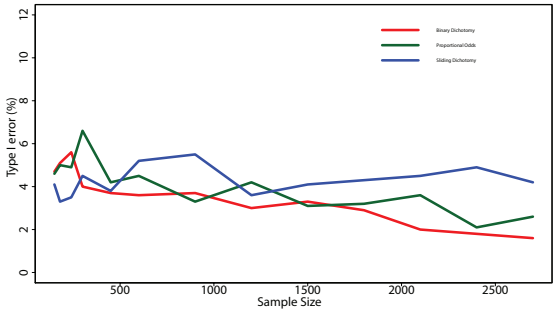


Figure 3.15: Type I error in scenario 6: One covariate, equal treatment sizes and unequal band sizes

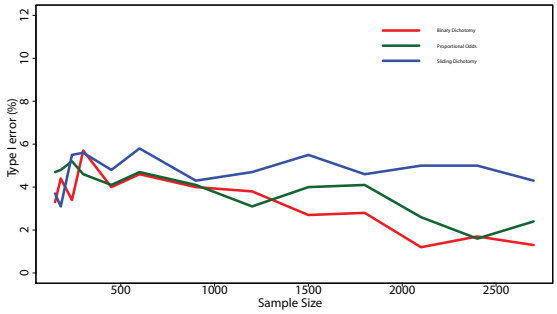


Figure 3.16: Type I error in scenario 7: One covariate, unequal treatment sizes and equal band sizes

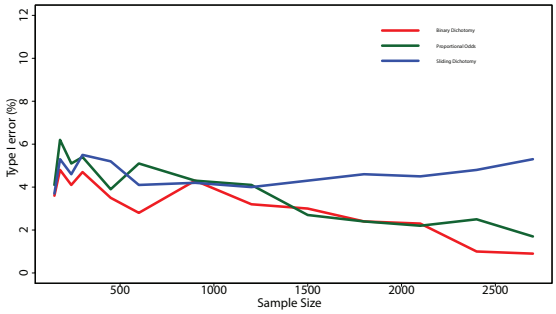


Figure 3.17: Type I error in scenario 8: One covariate, unequal treatment sizes and unequal band sizes

In the first two scenarios (Figures 3.10 and 3.11), the binary logistic model showed the expected gradual (although not always consistent) decline of cumulative proportion of the type I error as sample size increased, maintaining a rate convincingly below the 5% threshold for larger sample sizes. The two scenarios differ only in band sizes, which was expected to have limited effect on the coarse dichotomization required of the binary logistic model. Unfortunately, both the proportional odds model and the sliding dichotomy model did not demonstrate the expected tendency for the error rate to decline as sample size increased. Moreover, the proportional odds model seemed to have higher cumulative error rates over the sliding dichotomy model. Unequal band size (Figure 3.11) may have produced similar results for the sliding dichotomy model, however the proportional odds model showed a high type I error rate thought the sample size variation.

The unusual behavior of the proportional odds model was not seen in the subsequent scenarios. From Figure 3.12 to Figure 3.17, both the binary logistic model and the proportional odds models displayed expected behavior of decreasing type I error rate with increased sample size, and greater fluctuations and uncertainty in behavior for smaller sample sizes. However, for these same scenarios, the sliding dichotomy model failed to display either a consistent pattern of declining cumulative proportion of type I error with increasing sample size, or the ability to reliably achieve a type I error rate close to other two model for any of the sample sizes above 1500.

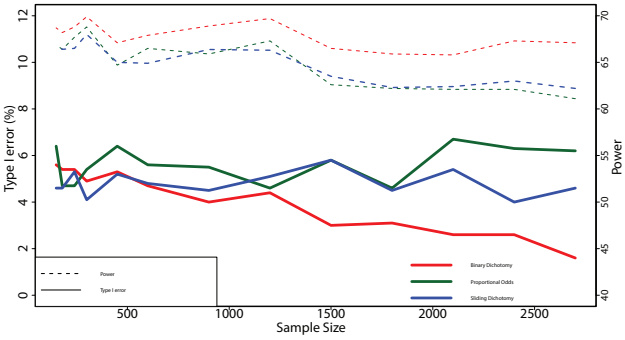
The first two scenarios (Figures 3.10 and 3.11) represent models with four covariates and equal treatment sizes. The GOS band size difference included in the models shown in Figure 3.11 do not affect the proportional odds model, which means the covariates or the treatment sizes are responsible for the proportional odds model failing to control the type I error rate in these two scenarios. The model may be over-fitted with four covariates, particularly for the smaller sample sizes, or the equal treatment sizes may have forced a violation of the proportional odds assumption. As each sample size was simulated 1000 times, it was not feasible to test the assumption in every case. Furthermore, the aim of this analysis was not to evaluate the proportional odds model or its assumptions in detail.

We assumed, based on the clinical effectiveness of the sliding dichotomy model, that it would control type I error rate in-line with the model it is based on, the binary logistic model. Unfortunately, the sliding dichotomy model failed to do so, showing fluctuating yet generally poor behavior across all eight scenarios. This indicates the risk of producing a higher proportion of false-positive results by the sliding dichotomy model, when it is applied to clinical data analyses that the method was specifically developed to analyze.

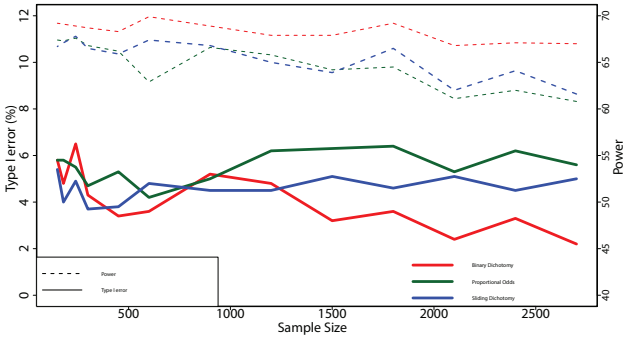
3.4 Power Analysis

The papers exhibited earlier fulfilled the first three objectives. The fourth objective, identifying the relationship between power, sample size and type I error rate in ordinal outcomes of phase III trials of TBI, is addressed in this section. In the previous papers, the CRM did not perform up to the mark compared to other models. It is not highly regarded as a robust model in clinical trials. Moreover, the CRM is best suited for cases when the individual categories of the response variable are of intrinsic interest, and are not merely an arbitrary grouping of an underlying continuous variable (Ananth and Kleinbaum, 1997). Additionally, the outcome of the POM is unchanged even if the ordering of the categories is reversed which is not the same

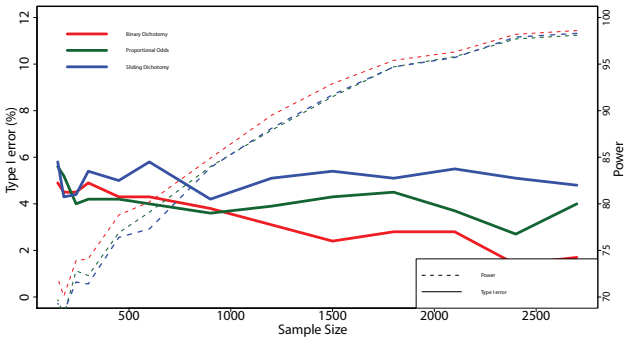
case for the CRM (ARMSTRONG and Sloan, 1989; Dos Santos and Berridge, 2000). However, this reversibility was praised in certain cases, particularly in longitudinal studies. Then again, it makes this model data dependent (Lindsey et al., 1997; Dos Santos and Berridge, 2000). Going forward, we only considered the binary logistic model, the POM, and the SDM.



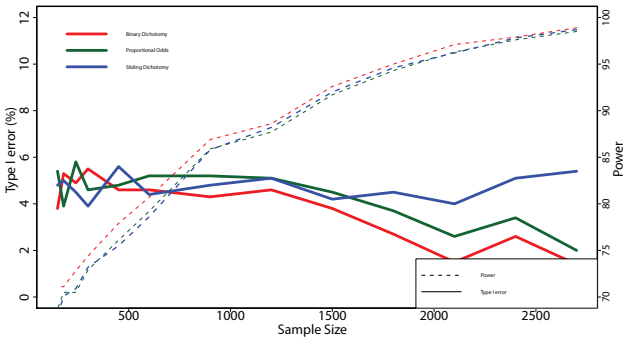
(a) Type I error in scenario 1: Four covariates, equal treatment sizes and equal band sizes



(b) Type I error in scenario 2: Four covariates, equal treatment sizes and unequal band sizes



(c) Type I error in scenario 3: Four covariates, unequal treatment sizes and equal band sizes



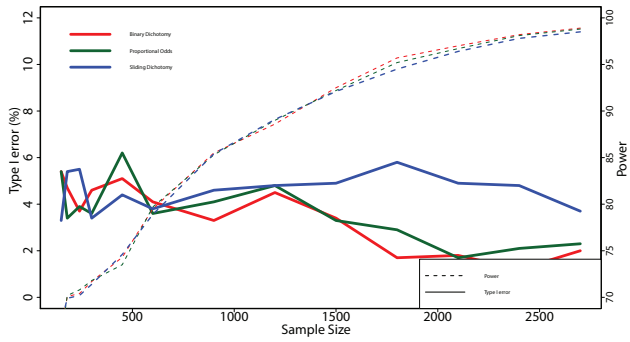
(d) Type I error in scenario 4: Four covariates, unequal treatment sizes and unequal band sizes

Figure 3.18: The comparison among power, sample size and type I error rate for CRASH data

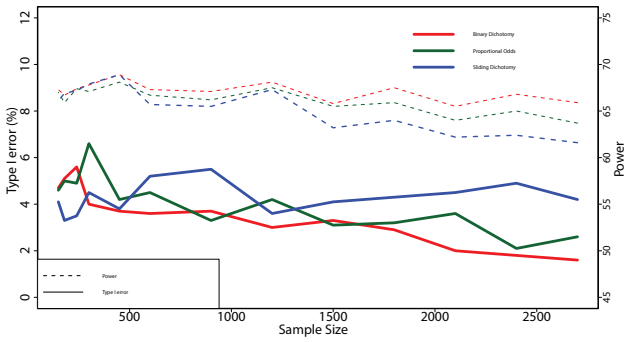
As mentioned in Chapter 3.3, variations in the treatment level, GOS allocation, and number of covariates resulted in eight different scenarios. Each scenario was assessed by a sample size range of 150 to 2700. Statistical power and type I error rate was quantified and compared graphically (Figure 3.4 and 3.4). The dotted lines and the smooth lines represent statistical power and type I error rate respectively.

In the first two scenarios (Figure 3.4(a) and (b)), where four covariates (age, sex, GCS motor response and treatment) were applied alongside equal treatment sizes (50% 'treatment A' and 50% 'treatment B'), binary dichotomy showed the highest power followed by the POM and SDM. The gap between the SDM and POM was not every high. In the following scenarios, similar statistical powers were demonstrated by the three models. As discussed before the cumulative proportion of type I error of the SDM were mostly unsatisfactory (apart from the

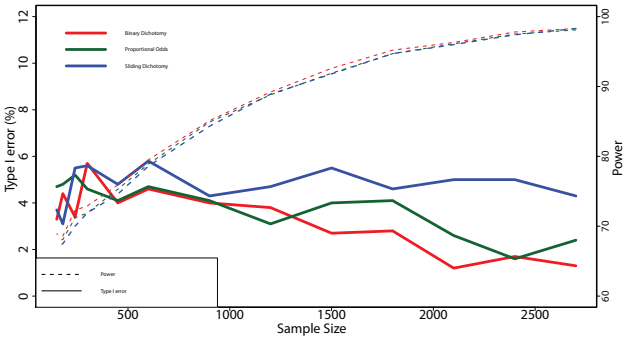
first two scenarios) with the blue line hovering along 5% line. Compared to that the power displayed by the SDM was not satisfactory.



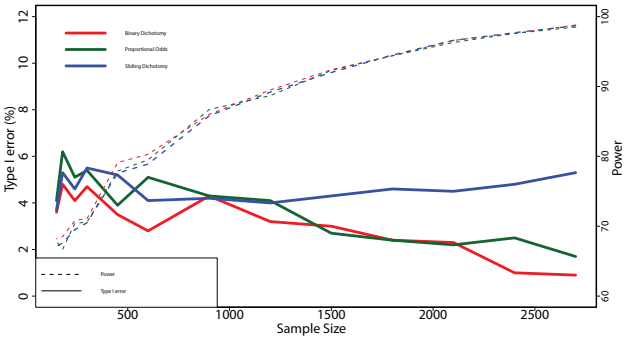
(a) Type I error in scenario 5: One covariate, equal treatment sizes and equal band sizes



(b) Type I error in scenario 6: One covariate, equal treatment sizes and unequal band sizes



(c) Type I error in scenario 7: One covariate, unequal treatment sizes and equal band sizes



(d) Type I error in scenario 8: One covariate, unequal treatment sizes and unequal band sizes

Figure 3.19: The comparison among power, sample size and type I error rate for CRASH data

It is important to note that the binary logistic model might have shown the highest power; however, their power lies in detecting how many patients shifted from 'unfavorable' to 'favorable' state without counting the underlying transitions. Furthermore, the null hypothesis was rejected in all eight cases, which is no treatment effect was found. It should be considered while interpreting the statistical powers. Particularly, as the SDM performs better when probability of a favorable outcome is high, which is not the case here (Price et al., 2013).

The weakness of the SDM is the lack of theoretical design and the uncontrolled type I error rate. The other model has their weaknesses as well, discussed in previous literatures. Comparing the power, sample size and type I error rate shows the success of the SDM, high power, lies with its uncontrolled type I error, which supports the primary statistical concept (section 1.1).

Table 3.5: Summarizing the three models for ordinal analysis of TBI

Model	Type I error rate	Statistical power	Inadequacy
Binary logistic	Strong	Weak*	Information loss
Proportionals odds	Strong	Moderate	Restrictive assumption
Sliding dichotomy	Weak	Strong	False positive results

* underlying shifts are not considered

3.5 New model outline

An optimum model should be able to fit the data with the power of the SDM, error control of the binary logistic model and without any strict assumptions like the proportional odds assumption. The final objective is to theorize possible strategies to develop the SDM to articulate a *robust* model with greater control over the type I error rate. Three problems are required to be addressed to rectify the weakness of the SDM:

1. The subjectivity in covariate selection for bands formulation
2. The dichotomization technique that ignores the negative effects of the treatment effect
3. The control over type I error rate

The first and best possible step should be to formulate a theoretical framework for the SDM. A mathematical expression like the binary logistic or the POM will allow theoretical development addressing the issues stated above. However, at this point we attempted to look for temporary solutions of these issues.

To tackle the first issue, the POM can be applied prior to the SDM and predetermine the 'important' covariates. These covariates can be used to formulate the bands, which will reduce the subjectivity to some extent. Here the odds assumption should not be an issue, as the odds ratios will not be interpreted.

Instead of binary logistic regression, multivariate generalized linear regression can be applied in the final model fitting of the SDM. This could improve the process, where the outcomes will not necessary have to be binary. According to Fahrmeir and Tutz (2001), multivariate GLM is special cases of multinomial response models defined by equation 3.1.

$$\mathbf{Pr}(\boldsymbol{Y_i = r}) = \frac{\boldsymbol{e}^{(\beta_{r0}+z_i\beta_r)}}{\mathbf{1} + \sum_{s=1}^{q} \boldsymbol{e}^{\beta_{s0}+z_i\beta_s}} \tag{3.1}$$

where $\boldsymbol{z_i}$ is a $(q \times p)$ design matrix composed of covariates $\boldsymbol{x}$ and $\boldsymbol{\beta}$ is a $(p \times 1)$-vector of unknown parameters. This is based on a distributional assumptions and a structural assumption. However, his model can become computationally infeasible if the cluster size gets large (Carey et al., 1993). Applying it in the bands might improve the SDM.

A potential way to develop the POM is to articulate the outcome variable to satisfy the proportional odds assumption. Simulations can be applied in a practical data set to create equal proportion of ranked scales in the outcome variable. It will allow a better scope of fulfilling the assumption without altering the primary results.

Chapter 4

Discussion and Conclusions

The results in previous sections display the limitations of the contemporary models applied for analyzing ordinal outcomes. The same data set with the same variables reached different conclusions when different models were used. These results show the sensitiveness of ordinal outcomes as well as the failure of current models in fitting them. Moreover, the goodness of fit statistics did not support one specific model. Our results were not limited to only clinical trials data. We used other public health data sets to reassess the models, and found these limitations.

The first section of the results (section 3.1) shows the applications of contemporary ordinal models in a primary data on regional health. The results agreed with previous research that the SDM is the best-fitted model according to the goodness of fitness statistics (Murray et al., 2005). The additional part in section 3.1 showed the application of the models in another health survey, which had a higher sample size. However, the goodness of fit statistics did not agree on a unanimously best-fitted model. The POM and SDM both showed adequacy. Furthermore, the POM satisfied the proportional odds assumption as the outcome variable was scaled equally in five parts. This section showed the subjectivity involved in selecting an ordinal model in public health data sets. This is an extension of works of Price et al. (2013), Hardin et al. (2007) and O'Connell and Liu (2011), who showed the problems of each of the models but did not apply all the models together to compare goodness of fits, particularly with the SDM. Moreover, we discussed the importance of considering multiple models for analyzing ordinal outcome variables.

The second section of the results (section 3.2) was a demonstration of using five different models and a consistent interpretation from the outcomes. Two different data sets were compared and a novel conclusion was drawn regarding the geographical difference in TBI experience for patients. Moreover, the models that showed significant covariate effects were extended (in-

tegration effects) to explore the effect of age and sex on TBI. The results were consistent with the recent literature of TBI (Hurn et al., 2005; Bazarian et al., 2010; Coronado et al., 2011; Chan et al., 2017; Sandel et al., 2017). We contributed to this area of research by applying different ordinal methods that ensured the validity of results and its interpretation.

The results section 3.3 is an important addition to this area of research, where we displayed the drawback of the SDM, and why it is necessary to apply multiple models instead of relying on only one. The SDM failed to control the cumulative proportion type I error for eight different simulated scenarios compared to the binary logistic model or the POM. It is particularly important for clinical trial outcomes to maintain false positives in phase III trial, as the results from this stage are considered applicable (as medicines or treatments) for treating diseases/injuries. We extended the literature by scrutinizing the SDM, which was generally praised in previous studies (Young et al., 2005; Murray et al., 2005).

The fourth and fifth sections of results are works in progress and future implications of this project. We showed how statistical power compares across ordinal models and the importance of considering the type I error rate as well as power while formulating a new model. Several plans are suggested in section 3.5 to improve the SDM. The most important one among them would be to devise a theoretical model for the SDM and then attempt to control the error rate.

Discussion regarding each result are detailed in each paper or at the end of each section in Chapter 3. Various factors are associated with the traditional models' fitness, as they varied over data sets. The weaknesses of the models can be summarized as following.

- Binary logistic model, designed for two scale outcome variables, suffers from much information loss when applied to any multiple ranked order outcome analysis (Lovric, 2011)
- Proportional odds model regularly fails to meet the proportional odds assumption, particularly in RCTs, where the researchers have no control over the outcome scale proportions (Meter et al., 2011; Jäckle et al., 2010)
- Continuation ratio model and partial proportional odds model lacks fitness compared to the POM and their application is restricted to particular scenarios; and applied mostly when parallel assumption of POM is violated (Bender and Grouven, 1998; Das and Rahman, 2011)
- Sliding dichotomy model works better only in data sets where true effects of treatment exists (Young et al., 2005; Bath et al., 2012). Furthermore, selection of the covariates for

initial band formulation induces subjectivity. False conclusions are likely due to the lack of type I error rate control.

To address these limitations, formulation of a robust model is required that can consistently fit ordinal outcomes under most scenarios, especially without any strict assumptions that are easily violated (O'Connell, 2006). We suggest some steps that would improve the current analysis process,

- Test of assumptions before applying the POM, or any statistical model
- Apply multiple models before reporting the results from a single model
- Report goodness of fit statistics along with the results, which will allow future studies with same data sets to compare their models' fitness
- Evaluate the type I error rate when a model is formulated or assess the literature before the application

Time constraint has limited this study to explore and test a few more options of developing a new model. A theoretical framework for the SDM would contribute to develop a robust model. In Chapter 3.3, 1000 simulations were conducted; perhaps the results would have been more comprehensive with 100,000 simulations. The statistical power comparison needs to be assessed with a data set where true treatment effect exists (null hypothesis rejected), unlike Chapter 3.4. Although the focus of this study was on public health, a data outside this area could have generalized the conclusions.

Despite these limitations, this study came to a reproducible conclusion. The available models are inadequate to fit ordinal outcome variables and lack proper assessment on the fitted covariates. Future works can include developing either the POM or SDM. Both model should be revised to improve their application. An error control mechanism in the SDM could make it more robust. Moreover, the subjectivity involved in selecting the bands should also be minimized. On the other hand, a revised POM could be developed that would allow the strict parallel assumption to be breached, but still provide robust fitness.

The objective of this study was to evaluate the performances of the available statistical methods and determine their performance reliability by comparing the inference drawn across models when applied consistently to the same data sets. Furthermore, we explicitly assessed the incapability of the SDM in controlling the type I error rate. The results showed a lack of fitness for all the models, sometimes even at the most basic level of analysis based on the

binary logistic model. Moreover, the problem stands for most types of public health studies including the clinical trials. A patient's status in traumatic brain injury is sensitive and the slightest of improvement (or decline) should be specifially analyzed. Unfortunately, none of the present statistical models is capable of reliably doing that in their current form.

Two statistical methods when applied to the same data sets should not demonstrate two different conclusions due to lack of a robust statistical model (Huber, 2011). Most ordinal methods are data dependent and often require data manipulation to reach a comprehensive result. Consequently, meta-analysis is becoming more popular and opportunistic results are interpreted in the papers, where only the models with favorable conclusions are reported.

Ordinal outcome analyses still require researchers to apply models with known limitations and a history of unsatisfying results. This study analyzed clinical trial rank outcomes for TBI as well as some public health data sets to comprehensively assess the contemporary models. None of these models showed encouraging results in their present form. Thus, we fulfilled the final objective of this research by providing a comprehensive exploration of current literature gap, analyzing the performance of the models in various scenarios, discussing each model's specifi weaknesses, and indicating pathways to new development of ordinal outcome models.

More empirical studies with theoretical development are required to formulate a more robust, generalized model. Until this occurs, we suggest application of multiple models (and the appropriate tests of assumptions for each), comparison of results and the use of goodness of fit statistics to help determine the best-fitted model for final reporting and interpretation of results from ordinal outcome variable.

Appendix

List of Abbreviations

1. BDHS = Bangladesh Demographic and Health Surveys
2. CRASH = Corticosteroid Randomisation after Significant Head Injury
3. CRM = Continuation Ratio Model
4. D = Death
5. DHS = Demographic and Health Surveys
6. GCS = Glasgow Comma Scale
7. GOS = Glasgow Outcome Scale
8. GR = Good Recovery
9. MD = Moderate Disability
10. OLS = Ordinary Least Square
11. POM = Proportional Odds Model
12. PPOM = Partial Proportional Odds Model
13. SD = Severe Disability
14. SDM = Sliding Dichotomy Model
15. TBI = Traumatic Brain Injury
16. VS = Vegetative State

References

Ananth, C. V., Kleinbaum, D. G., 1997. Regression models for ordinal responses: a review of methods and applications. International journal of epidemiology 26 (6), 1323–1333.

Anderson, J., Philips, P., 1981. Regression, discrimination and measurement models for ordered categorical variables. Applied statistics, 22–31.

Andrich, D., 1979. A model for contingency tables having an ordered response classification. Biometrics, 403–415.

ARMSTRONG, B., Sloan, M., 1989. Ordinal regression models for epidemiologic data. American Journal of Epidemiology 129 (1), 191–204.

Bagiella, E., Novack, T. A., Ansel, B., Diaz-Arrastia, R., Dikmen, S., Hart, T., Temkin, N., 2010. Measuring outcome in traumatic brain injury treatment trials: recommendations from the traumatic brain injury clinical trials network. The Journal of head trauma rehabilitation 25 (5), 375.

Bath, P. M., Lees, K. R., Schellinger, P. D., Altman, H., Bland, M., Hogg, C., Howard, G., Saver, J. L., 2012. Statistical analysis of the primary outcome in acute stroke trials. Stroke 43 (4), 1171–1178.

Bazarian, J. J., Blyth, B., Mookerjee, S., He, H., McDermott, M. P., 2010. Sex differences in outcome after mild traumatic brain injury. Journal of neurotrauma 27 (3), 527–539.

Bender, R., Grouven, U., 1998. Using binary logistic regression models for ordinal data with non-proportional odds. Journal of clinical epidemiology 51 (10), 809–816.

Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Johnson, V., 2017. Redefine statistical significance.

Berge, E., Barer, D., 2002. Could stroke trials be missing important treatment effects? Cerebrovascular Diseases 13 (1), 73–75.

Blomberg, S., 2014. Power analysis using r.

Bolland, K., Sooriyarachchi, M., Whitehead, J., 1998. Sample size review in a head injury trial with ordered categorical responses. Statistics in medicine 17 (24), 2835–2847.

Campbell, M. K., Donner, A., Webster, K. M., 1991. Are ordinal models useful for classification? Statistics in medicine 10 (3), 383–394.

Carey, V., Zeger, S. L., Diggle, P., 1993. Modelling multivariate binary data with alternating logistic regressions. Biometrika 80 (3), 517–526.

Chan, V., Mollayeva, T., Ottenbacher, K. J., Colantonio, A., 2017. Clinical profile and comorbidity of traumatic brain injury among younger and older men and women: a brief research notes. BMC Research Notes 10 (1), 371.

Chu, W., Ghahramani, Z., 2005. Gaussian processes for ordinal regression. Journal of machine learning research 6 (Jul), 1019–1041.

Collaborators, C. T., et al., 2005. Final results of mrc crash, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury outcomes at 6 months. The Lancet 365 (9475), 1957–1959.

Coronado, V. G., Xu, L., Basavaraju, S. V., McGuire, L. C., Wald, M. M., Faul, M. D., Guzman, B. R., Hemphill, J. D., et al., 2011. Surveillance for traumatic brain injury-related deaths: United states, 1997-2007.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological), 187–220.

CRASH, 2005a. Crash corticosteroid radomisation after significant head injuryAccessed: 2015-01-05.

CRASH, 2005b. International study of recovery after head injury: Six month follow-up questionnaireAccessed: 2015-01-05.

Das, S., Rahman, R. M., 2011. Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in bangladesh. Nutrition journal 10 (1), 124.

Diez-Roux, A. V., 2000. Multilevel analysis in public health research. Annual review of public health 21 (1), 171–192.

Dobson, A. J., Barnett, A., 2008. An introduction to generalized linear models. CRC press.

Dos Santos, D. M., Berridge, D. M., 2000. A continuation ratio random effects model for repeated ordinal responses. Statistics in medicine 19 (24), 3377–3388.

Easterbrook, P. J., Gopalan, R., Berlin, J., Matthews, D. R., 1991. Publication bias in clinical research. The Lancet 337 (8746), 867–872.

European Medicines Agency, 2003. Guidelines: Medicinal products for human use safety, environment and information 3B, accessed: 2015-12-31.

Fagerland, M. W., Hosmer, D. W., 2016. Tests for goodness of fit in ordinal logistic regression models. Journal of Statistical Computation and Simulation, 1–21.

Fahrmeir, L., Tutz, G., 2001. Models for multicategorical responses: Multivariate extensions of generalized linear models. In: Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, pp. 69–137.

Farewell, V. T., 1982. A note on regression analysis of ordinal data with variability of classification. Biometrika 69 (3), 533–538.

Faul, M., Xu, L., Wald, M. M., Coronado, V., 2010. Traumatic brain injury in the united states. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Fisher, R. A., 1922. On the interpretation of χ 2 from contingency tables, and the calculation of p. Journal of the Royal Statistical Society 85 (1), 87–94.

Food and Drug Administration (FDA), 2014. Guidances (drugs). http://www.fda.gov/Drugs/accessed: 2015-12-31.

Freeman, E. A., 1987. The catastrophe of coma: a way back. David Bateman Limited.

Friedman, L. M., Furberg, C., DeMets, D. L., et al., 2010. Fundamentals of clinical trials. Vol. 4. Springer.

Goldstein, M., 1990. Traumatic brain injury: a silent epidemic. Annals of neurology 27 (3), 327–327.

Goos, P., Meintrup, D., 2016. Statistics with JMP: Hypothesis Tests, ANOVA and Regression. John Wiley & Sons.

Greenland, S., 1985. An application of logistic models to the analysis of ordinal responses. Biometrical Journal 27 (2), 189–197.

Greenwood, C., Farewell, V., 1988. A comparison of regression models for ordinal data in an analysis of transplanted-kidney function. Canadian Journal of Statistics 16 (4), 325–335.

Hardin, J. W., Hilbe, J. M., Hilbe, J., 2007. Generalized linear models and extensions. Stata Press.

Harrell, F., 2015. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

Hemri, S., Haiden, T., Pappenberger, F., 2016. Discrete post-processing of total cloud cover ensemble forecasts. Monthly Weather Review (2016).

Hernandez, P. A., Graham, C. H., Master, L. L., Albert, D. L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29 (5), 773–785.

Homer, J. B., Hirsch, G. B., 2006. System dynamics modeling for public health: background and opportunities. American journal of public health 96 (3), 452–458.

Hosmer Jr, D. W., Lemeshow, S., 2004. Applied logistic regression. John Wiley & Sons.

Huber, P. J., 2011. Robust statistics. In: International Encyclopedia of Statistical Science. Springer, pp. 1248–1251.

Huff, T. E., 2003. The rise of early modern science: Islam, China and the West. Cambridge University Press.

Hurn, P. D., Vannucci, S. J., Hagberg, H., 2005. Adult or perinatal brain injury. Stroke 36 (2), 193–195.

Iasonos, A., Zohar, S., O'Quigley, J., 2011. Incorporating lower grade toxicity information into dose finding designs. Clinical Trials 8 (4), 370–379.

Ilodigwe, D., Stat, M., Murray, G. D., Kassell, N. F., Torner, J., Kerr, R. S., Molyneux, A. J., Macdonald, R. L., 2013. Sliding dichotomy compared with fixed dichotomization of ordinal outcome scales in subarachnoid hemorrhage trials: Clinical article. Journal of neurosurgery 118 (1), 3–12.

Ivan, L., Bruce, D., 1982. Coma–physiopathology, diagnosis and management. Charles C. Thomas, Spring-field, IL.

Jäckle, A., Roberts, C., Lynn, P., 2010. Assessing the effect of data collection mode on measurement. International Statistical Review 78 (1), 3–20.

Jakobsen, J. C., Gluud, C., Winkel, P., Lange, T., Wetterslev, J., 2014. The thresholds for statistical and clinical significance–a five-step procedure for evaluation of intervention effects in randomised clinical trials. BMC medical research methodology 14 (1), 1.

Jennett, B., Plum, F., 1972. Persistent vegetative state after brain damage: a syndrome in search of a name. The Lancet 299 (7753), 734–737.

Jennett, B., Snoek, J., Bond, M., Brooks, N., 1981. Disability after severe head injury: observations on the use of the glasgow outcome scale. Journal of Neurology, Neurosurgery & Psychiatry 44 (4), 285–293.

Jüni, P., Altman, D. G., Egger, M., 2001. Assessing the quality of randomised controlled trials. Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition, 87–108.

Khan, F., Baguley, I. J., Cameron, I. D., 2003. 4: Rehabilitation after traumatic brain injury. Medical Journal of Australia 178 (6), 290–297.

Lall, R., Campbell, M., Walters, S., Morgan, K., Co-operative, M. C., 2002. A review of ordinal regression models applied on health-related quality of life assessments. Statistical methods in medical research 11 (1), 49–67.

Lieberman, M. D., Cunningham, W. A., 2009. Type i and type ii error concerns in fmri research: re-balancing the scale. Social cognitive and affective neuroscience, nsp052.

Lindsey, J., Jones, B., Ebbutt, A., 1997. Simple models for repeated ordinal responses with an application to a seasonal rhinitis clinical trial. Statistics in Medicine 16 (24), 2873–2882.

Lipsitz, S. R., Buoncristiani, J. F., 1994. A robust goodness-of-fit test statistic with application to ordinal regression models. Statistics in medicine 13 (2), 143–152.

Liu, I., Mukherjee, B., 2008. Proportional odds model. Encyclopedia of Statistical Sciences.

Liu, X., 2010. Ordinal regression analysis: Fitting the continuation ratio model to educational data using stata.

López, M. F., Dupuy, J.-F., Gonzalez, C. V., 2012. Effectiveness of adaptive designs for phase ii cancer trials. Contemporary clinical trials 33 (1), 223–227.

Lovric, M., 2011. International Encyclopedia of Statistical Science. Springer.

Maas, A. I., Marmarou, A., Murray, G. D., Teasdale, S. G. M., Steyerberg, E. W., 2007. Prognosis and clinical trial design in traumatic brain injury: the impact study. Journal of neurotrauma 24 (2), 232–238.

Maas, A. I., Menon, D. K., Lingsma, H. F., Pineda, J. A., Sandel, M. E., Manley, G. T., 2012. Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. Journal of neurotrauma 29 (1), 32–46.

Maas, C. J., Hox, J. J., 2005. Sufficient sample sizes for multilevel modeling. Methodology 1 (3), 86–92.

Mann, H., Wald, A., 1942. On the choice of the number of class intervals in the application of the chi square test. The Annals of Mathematical Statistics 13 (3), 306–317.

McCullagh, P., 1980. Regression models for ordinal data. Journal of the royal statistical society. Series B (Methodological), 109–142.

McCullagh, P., 2005. Proportional-odds model. Encyclopedia of Biostatistics.

McHugh, G. S., Butcher, I., Steyerberg, E. W., Marmarou, A., Lu, J., Lingsma, H. F., Weir, J., Maas, A. I., Murray, G. D., 2010. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the impact project. Clinical Trials 7 (1), 44–57.

Mendelow, A. D., Gregson, B. A., Fernandes, H. M., Murray, G. D., Teasdale, G. M., Hope, D. T., Karimi, A., Shaw, M. D. M., Barer, D. H., investigators, S., et al., 2005. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the international surgical trial in intracerebral haemorrhage (stich): a randomised trial. The Lancet 365 (9457), 387–397.

Menon, D. K., Schwab, K., Wright, D. W., Maas, A. I., et al., 2010. Position statement: definition of traumatic brain injury. Archives of physical medicine and rehabilitation 91 (11), 1637–1640.

Meter, E. M. V., Garrett-Mayer, E., Bandyopadhyay, D., 2011. Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading. Statistics in medicine 30 (17), 2070–2080.

Mocullagh, P., 1977. A logistic model for paired comparisons with ordered categorical data. Biometrika 64 (3), 449–453.

Mondol, M. A., Rahman, N., Akhter, S., Ahmed, B., Rahman, A., Momen, M. A., Rahman, M., Talukder, D. C., 2013. Sociodemographical and clinical presentation of traumatic brain injury patients. Journal of Dhaka Medical College 22 (1), 45–50.

Mooradian, J., Ivan, J. N., Ravishanker, N., Hu, S., 2013. Analysis of driver and passenger crash injury severity using partial proportional odds models. Accident Analysis & Prevention 58, 53–58.

Moore, D. S., McCabe, G. P., Craig, B. A., 2012. Introduction to the practice of statistics. WH Freeman New York.

Morgan, A., Ziglio, E., 2007. Revitalising the evidence base for public health: an assets model. Promotion & Education 14 (2_suppl), 17–22.

Murray, G. D., Barer, D., Choi, S., Fernandes, H., Gregson, B., Lees, K. R., Maas, A. I., Marmarou, A., Mendelow, A. D., Steyerberg, E. W., et al., 2005. Design and analysis of phase iii trials with ordered outcome scales: the concept of the sliding dichotomy. Journal of neurotrauma 22 (5), 511–517.

Myburgh, J. A., Cooper, D. J., Finfer, S. R., Venkatesh, B., Jones, D., Higgins, A., Bishop, N., Higlett, T., for the Australian, A. T. B. I. S. A. I., et al., 2008. Epidemiology and 12-month outcomes from traumatic brain injury in australia and new zealand. Journal of Trauma and Acute Care Surgery 64 (4), 854–862.

National Center for Injury Prevention and Control (US), 2003. Report to Congress on mild traumatic brain injury in the United States: Steps to prevent a serious public health problem. Centers for Disease Control and Prevention.

O'Connell, A. A., 2006. Logistic regression models for ordinal response variables. No. 146. Sage.

O'Connell, A. A., Liu, X., 2011. Model diagnostics for proportional and partial proportional odds models. Journal of Modern Applied Statistical Methods 10 (1), 15.

Peterson, B., Harrell Jr, F. E., 1990. Partial proportional odds models for ordinal response variables. Applied Statistics, 205–217.

Price, M., Hertzberg, V., Wright, D. W., 2013. Does the sliding dichotomy result in higher powered clinical trials for stroke and traumatic brain injury research? Clinical Trials 10 (6), 924–934.

Reeves, R. H., Beltzman, D., Killu, K., 2000. Implications of traumatic brain injury for survivors of sexual abuse: A preliminary report of findings. Rehabilitation Psychology 45 (2), 205.

Rogers, M. A., 2009. What are the phases of intervention research. American Speech-Language-Hearing Association, Rockville.

Roozenbeek, B., Lingsma, H. F., Perel, P., Edwards, P., Roberts, I., Murray, G. D., Maas, A., Steyerberg, E. W., et al., 2011. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. Crit Care 15 (3), R127.

Roozenbeek, B., Maas, A. I., Lingsma, H. F., Butcher, I., Lu, J., Marmarou, A., McHugh, G. S., Weir, J., Murray, G. D., Steyerberg, E. W., et al., 2009a. Baseline characteristics and statistical power in randomized controlled trials: Selection, prognostic targeting, or covariate adjustment?*. Critical care medicine 37 (10), 2683–2690.

Roozenbeek, B., Maas, A. I., Marmarou, A., Butcher, I., Lingsma, H. F., Lu, J., McHugh, G. S., Murray, G. D., Steyerberg, E. W., 2009b. The influence of enrollment criteria on recruitment and outcome distribution in traumatic brain injury studies: results from the impact study. Journal of neurotrauma 26 (7), 1069–1075.

Rutland-Brown, W., Langlois, J. A., Thomas, K. E., Xi, Y. L., et al., 2006. Incidence of traumatic brain injury in the united states, 2003. Journal of Head Trauma Rehabilitation 21 (6), 544.

Sandel, N. K., Schatz, P., Goldberg, K. B., Lazar, M., 2017. Sex-based differences in cognitive deficits and symptom reporting among acutely concussed adolescent lacrosse and soccer players. The American journal of sports medicine 45 (4), 937–944.

Schneier, A. J., Shields, B. J., Hostetler, S. G., Xiang, H., Smith, G. A., 2006. Incidence of pediatric traumatic brain injury and associated hospital resource utilization in the united states. Pediatrics 118 (2), 483–492.

Scott, S. C., Goldberg, M. S., Mayo, N. E., 1997. Statistical assessment of ordinal outcomes in comparative studies. Journal of clinical epidemiology 50 (1), 45–55.

Senn, S., Julious, S., 2009. Measurement in clinical trials: a neglected issue for statisticians? Statistics in medicine 28 (26), 3189–3209.

Shen, Y., Liao, K. P., Cai, T., 2015. Sparse kernel machine regression for ordinal outcomes. Biometrics 71 (1), 63–70.

Singh, R., 2011. A review: Clinical trial and data management. trials 2, 3.

Stern, J. M., Simes, R. J., 1997. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. Bmj 315 (7109), 640–645.

Sterne, J. A., Egger, M., Smith, G. D., 2001. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ: British Medical Journal 323 (7304), 101.

Teasdale, G. M., Pettigrw, L. E., Wilson, J. L., Murry, G., Jennet, B., 1998. Analyzing outcome of treatment of severe head injury: a review and update on advancing the use of the glasgow outcome scale. Journal of neurotrauma 15 (8), 587–597.

Thurman, D. J., Alverson, C., Dunn, K. A., Guerrero, J., Sniezek, J. E., 1999. Traumatic brain injury in the united states: a public health perspective. The Journal of head trauma rehabilitation 14 (6), 602–615.

Tiesman, H. M., Konda, S., Bell, J. L., 2011. The epidemiology of fatal occupational traumatic brain injury in the us. American journal of preventive medicine 41 (1), 61–67.

Upton, G. J., 1992. Fisher's exact test. Journal of the Royal Statistical Society. Series A (Statistics in Society), 395–402.

Washington, S. P., Karlaftis, M. G., Mannering, F. L., 2010. Statistical and econometric methods for transportation data analysis. CRC press.

Weir, J., Steyerberg, E. W., Butcher, I., Lu, J., Lingsma, H. F., McHugh, G. S., Roozenbeek, B., Maas, A. I., Murray, G. D., 2012. Does the extended glasgow outcome scale add value to the conventional glasgow outcome scale? Journal of neurotrauma 29 (1), 53–58.

Wells, R., Dywan, J., Dumas, J., 2005. Life satisfaction and distress in family caregivers as related to specific behavioural changes after traumatic brain injury. Brain Injury 19 (13), 1105–1115.

Wright, J., 2011. Glasgow outcome scale. In: Encyclopedia of Clinical Neuropsychology. Springer, pp. 1150–1152.

Young, F. B., Lees, K. R., Weir, C. J., et al., 2005. Improving trial power through use of prognosis-adjusted end points. Stroke 36 (3), 597–601.