Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Adair Minihan Date

An Exploration of Regression Analysis Methods to Identify Significant Predictors of Visual Outcomes in Unilaterally and Bilaterally Injured Ocular Trauma Patients

By

Adair Minihan

Master of Public Health

Department of Biostatistics and Bioinformatics

C. Christina Mehta, PhD, MSPH

Committee Chair

Randi N. Smith, MD, MPH

Committee Member

An Exploration of Regression Analysis Methods to Identify Significant Predictors of Visual Outcomes in Unilaterally and Bilaterally Injured Ocular Trauma Patients

By

Adair Minihan

Bachelor of Science

The University of Texas at Austin

2017

Thesis Committee Chair: C. Christina Mehta, PhD, MSPH

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University in partial fulfillment of the requirements

for the degree of Master of Public Health

in Biostatistics

2019

Abstract

An Exploration of Regression Analysis Methods to Identify Significant Predictors of Visual Outcomes in Unilaterally and Bilaterally Injured Ocular Trauma Patients

By Adair Minihan

Background: Acknowledging the paired nature of ocular data is imperative to a sound statistical analysis, as it differs in unilaterally versus bilaterally injured patients and also in cross-sectional versus longitudinal analysis. The dataset in question contains unilaterally and bilaterally injured patients and is longitudinal. Previous literature does not address regression methods accounting for both unilaterally and bilaterally injured patients. It also does not attempt to identify significant clinical and demographic predictors for ocular trauma outcomes.

Objective: Identify regression methods for analysis of unilaterally and bilaterally injured ocular trauma patients in cross-sectional and longitudinal contexts.

Methods: The main predictor of interest was injury type and the outcome of interest was the logMAR score. LogMAR scores were divided into three categories: satisfactory (logMAR <= 0.3), moderately impaired $(0.3 <$ logMAR \le 1.0), and severely impaired visual outcomes $(log_{MAP} > 1)$. Cross-sectional analyses consisted of selecting the injured eye for unilaterally injured patients and the most injured eye at immediate follow up for bilaterally injured patients. Three logistic regressions with a baseline logit link were created, one for each follow up time. In longitudinal analysis, both eyes were included for bilaterally injured patients. Two mixed models were created with a logit link and random effect for subject id, one comparing satisfactory to moderately impaired outcomes, the other comparing satisfactory to severely impaired outcomes.

Results: Injury type was significant in both the cross-sectional logistic regression for the immediate follow up and the longitudinal mixed model comparing satisfactory outcomes to severely impaired outcomes. For patients with blunt injuries (vs. penetrating), the odds of having severely impaired outcomes was 0.236 (95% CI 0.099, 0.563) times the odds of having satisfactory outcomes at the immediate follow up. For patients with blunt injuries (vs. penetrating), the odds of having severely impaired outcomes was 0.15 (95% CI 0.055, 0.408) times the odds of having satisfactory outcomes in the longitudinal analysis.

Conclusions: The mixed models better captured the complexity of the data. They account for the longitudinal and paired nature, as both eyes were included for bilaterally injured patients.

Acknowledgements

I would like to thank the Rollins School of Public Health and the Department of Biostatistics and Bioinformatics for providing such an excellent education and learning environment where I have had many opportunities to apply my new skills. My mentor, faculty advisor, and thesis committee chair, C. Christina Mehta, PhD, MSPH, has been nothing short of incredible in guiding me throughout my two years as an MPH student at Rollins. I would also like to thank Randi Smith, MD, MPH for her collaboration, support, and data in making this thesis possible. I am also grateful to my peers both within the biostatistics department and other departments at Rollins for their support and comradery.

I am grateful to the University of Texas at Austin School of Public Health and the University of Colorado Summer Institute in Biostatistics for inspiring me to pursue higher education in public health and biostatistics. Most importantly, I would like to thank my family and close friends for their unwavering enthusiasm in all of my endeavors. I would not be here without them.

Table of Contents

INTRODUCTION

The proper analysis of health data is essential to producing statistically sound results. In public health and biomedical research, the publication of incorrect findings could have serious implications for health practices and outcomes in the general population. This is especially relevant to ophthalmological studies, as ocular trauma is the leading cause of monocular impairment, and skewed results could prevent researchers from identifying new ocular hazards in the environment.¹ Many ophthalmological studies measure variables of interest from both eyes for each patient. To an inexperienced researcher, analyzing ophthalmological data may involve treating each eye as an independent observation. This, however, would be incorrect, as the researcher would be ignoring correlation between the right and left eyes. By ignoring the intraclass correlation, the type one error rate would be inflated, thus producing invalid results.⁵ After appropriately acknowledging the paired nature of ocular data, hypothesis testing such as t tests, paired t tests, and chi square tests have been well-established methods for analyzing the variables of interest in these studies. Modeling, however, has been an increasingly popular method of analysis for paired ocular data as it allows researchers to control for multiple covariates. In this paper, different regression methods of analysis will be applied to a longitudinal ocular trauma dataset in order to identify significant variables.

BACKGROUND

Correct statistical methodology is essential for valid statistical inference in ophthalmological studies. Because ocular data is paired, the eyes cannot be treated as independent observations. Methods for hypothesis testing are well established and are dependent on three types of ophthalmological study design. More recent studies have relied on regression methods rather than hypothesis testing in order to account for covariates. Different methods of regression analysis yield different strengths and weaknesses.

Paired Ocular Data and Hypothesis Testing

As previously stated, it is important that ocular data are not treated as independent observations. In a simulation study, researchers ran 200 simulations where 22 patients were allocated to two treatment groups, A and B. The researchers then analyzed the data using four different methods: the comparison of the IOP (the visual outcome of interest) in the right eyes in group A with the right eyes in group B, the comparison of the IOP in the left eyes in group A with the left eyes in group B, the comparison of the average IOP of both eyes between group A and B, and the comparison of the IOP in all eyes of group A with all eyes of group B. This study demonstrated that the fourth comparison, where the eyes were treated as independent observations, inflated the type 1 error rate by a factor of nearly four. The study also concluded that averaging the observation of interest for each pair yielded the target type 1 error rate.⁵

As explored in the simulation study, paired ocular data can be analyzed using different hypothesis testing based off of study design and preference of the researcher. Designs or comparisons where only one eye is selected for each patient can be analyzed using tests such as chi-square and t tests. Designs or comparisons in which each patient has a treated eye and a controlled eye can be analyzed using paired t tests. For designs or comparison in which each patient has either two treated eyes or two control eyes, the outcome of interest is averaged for each pair. Chi-square and t tests can then be applied to these averages.⁶ Although all these three designs are commonly used, it was argued in "People and eyes: statistical approaches in ophthalmology" that the use of only one of the patient's eyes in the analysis leads to a loss in information. A powerful analysis involves data where one eye has been treated and the other is a control.4 These designs can also be used in regression analysis, where researchers can control for important covariates and potential confounders.

Regression Analysis Methods

Similar to the results found in the simulation study, the "Statistical Methods in Ophthalmology: An Adjustment for the Intraclass Correlation between Eyes" publication established that models where eye measurements were assumed to be independent were also not statistically valid.⁷ In the study "Tutorial on Biostatistics: Statistical Analysis for Correlated Binary Eye Data," three paired ocular datasets were analyzed in order to determine the optimal methods for analysis. When covariates needed to be controlled for, generalized linear mixed models (GLMM) and marginal models using generalized estimating equations (GEE) yielded similar results for all three datasets while accounting for the paired structure.¹¹ The study "Regression methods when

3

the eye is the unit of analysis" also found that GLMM and GEE methods yielded similar estimates and controlled for paired data in longitudinal analyses. Logistic regression also can be used for paired data when analyzed cross-sectionally and the outcome variable of interest is categorical or binary.^{3,8}

Global odds ratio regression models were also argued to be effective for paired ocular data. In comparison to GEE models, global odds ratio regression models allowed researchers to account for variables that affect only a specific eye and variable that effect the specific patient. This method, however, assumed that the treatment of one eye will not affect the control eye in the same patient.10 ANOVA methods can also be used when each individual has the same number of repeated observations, or when the data is balanced. When the data is unbalanced, GLMM and GEE models were considered optimal.²

Based off of this review, many methods can be used to analyze paired ocular data once the pairs have been accounted for. Because the dataset in question was paired and longitudinal, the main method of interest was GLMM. The data was also broken down cross-sectionally and analyzed using logistic regressions. This data was unique because it was ocular trauma data where some patients only injured one eye (unilateral injury) while some patients injured both eyes (bilateral injury). Regression methods had to account for pairing in both types of injuries. This data was also unique in the sense that it is trauma data. Unlike studies where specific doses of medication are administered to one or both eyes, the type and severity of injury could not be predetermined. Therefore, there was less clarity in the nature of the pairing for the injured and uninjured eyes. Because of this, the visual outcomes for uninjured eyes were not necessary in predicting visual outcomes of injured eyes. By identifying effective regression methods of analysis, significant predictors of visual outcomes in ocular trauma cases can be identified. This

is novel because there are very few studies that address demographic and clinical predictors of visual outcomes of unilateral and bilateral ocular injuries in urban settings.

METHODS

Ocular Trauma Data

The ocular trauma data was gathered at the Grady Memorial Hospital emergency department. A total of 291 patients with eye injuries were recorded, with 290 having visual acuity data. Demographic information was recorded for each patient, including variables such as race, age at time of injury, and gender. Variables relating to the type and severity of the injury were recorded as well, such as whether the injury was blunt or penetrating, the injury severity score, and the Glasgow coma score. The primary predictor of interest was injury type. It was also noted whether the patient injured one eye or both. Each patient had their vision assessed immediately after the injury. Follow-up vision assessments were performed one month after the injury and one year after the injury, where the visual acuity was measured by logMAR.

Data Cleaning and Categorization of Variables

The visual acuity data was then categorized into three classifications based on the LogMar values. Patients with satisfactory vision had LogMar values of less than or equal to 0.3. Patients with moderate visual impairment had LogMar values greater than 0.3 but less than or equal to 1.0. Patients with severe visual impairment had LogMar values greater than 1.0. Patients who had conditions such as losing an eye due to the ocular trauma were also classified as having severe visual impairment. Because there were only 291 patients and some patients were lost to follow up for subsequent visual acuity measurements, there was some sparsity in the data.

Cross-sectional Methods

The data was first analyzed cross sectionally. For each of three follow up times, multinomial logistic models with baseline logit link were created to predict visual acuity outcomes based on a variety of variables. Unilaterally injured eyes were first pooled independently of whether it was the right eye or left eye that was injured. For patients with bilateral injuries, the eye that was the most severely impaired initially was selected for each patient, as both eyes could not be included in cross-sectional analysis while still controlling for the paired nature. Because the logMAR value was categorical, averaging the two eyes was not appropriate. The unilaterally injured eyes and more impaired eye of the bilateral patients were included in the same analysis. For each follow up time, one model was created using only the injured eyes. The uninjured eye for unilateral patients was not included in the models, as it led to model instability and was deemed not necessary for the analysis. An indicator variable was included in the models to account for any differences between unilaterally and bilaterally injured patients. The outcomes of this analysis were later compared to the results of the longitudinal analysis.

Longitudinal Methods

The data was then analyzed longitudinally. Linear mixed models were used to predict the visual acuity outcomes. Because linear mixed models can account for paired data, both eyes for bilaterally injured patients were included in the analysis. The uninjured eyes of the unilaterally injured patients were excluded from the analysis, as they again created model instability, similar to the logistic models created in the cross-sectional analysis. A random effect by subject was included in the model. A generalized mixed model using a cumulative logit link was initially applied in order to account for the paired and longitudinal nature of the data. This model was not stable, as many patients had been lost to follow up and lacked longitudinal data. To attempt to find more model stability, two new mixed models were created using a logit link. In the first model, the outcome variable was limited to only satisfactory visual outcomes and moderately impaired outcomes. In the second model, the outcome variable was limited to satisfactory visual outcomes and severely impaired outcomes. By comparing only two visual outcome levels at once, the models were more stable.

RESULTS

Descriptive Analysis

The data had a total of 290 patients with visual acuity data, where 175 patients (94.73%) were unilaterally injured and 15 patients (5.17%) were bilaterally injured. 220 patients (75.86%) were male and 70 (24.14%) were female, as shown in Table 1. The patients had a median age of 41 and a median hospital length of stay of 2. The injury type for this group of patients was mostly blunt with 232 individuals (80.28%) while the rest had penetrating injuries (19.72%). Of the 290 patients, 243 (83.79%) participated in the immediate follow up, 165 (56.9%) participated in the one month follow up, and 23 (7.93%) participated in the one year follow up. 163 (56.21%) patients had only one follow up visit, while 14 (4.83%) patients participated in all three follow ups.

Cross-Sectional Analysis

In the cross-sectional analysis, a total of three baseline logistic regression were used, one for each respective follow up time. In the immediate follow up logistic regression, the type of injury (blunt vs. penetrating) and age was a significant covariate. For patients with blunt injuries (vs.

penetrating), the odds of having a logMAR measurement of greater than 1 was 0.236 (95% CI 0.099, 0.563) times the odds of having a logMAR measurement of less than 0.3 (p value = 0.0011), as shown in Table 2. For every one year increase in age, the odds of having a logMAR measurement of greater than 1 in comparison to a logMAR value of less than or equal to 0.3 increased by 2% (95% CI 0.02 %, 3.8%) (p value $= 0.0288$). For the logistic regressions corresponding to the one month and one year follow up, none of the covariates were statistically significant, as shown in Tables 3 and 4. Because of loss to follow up, the data for the one year follow up was sparse, and therefore could not account for covariates other than injury type.

Longitudinal Analysis

In the mixed model comparing satisfactory visual outcomes to severely impaired visual outcomes, injury type was again significant in addition to follow up time and whether the patient suffered from unilateral or bilateral injuries. For patients with blunt injuries (vs. penetrating), the odds of having a logMAR measurement of greater than 1 was 0.15 (95% CI 0.055, 0.408) times the odds of having a logMAR measurement of less than 0.3 (p value = 0.0002), as shown in Table 5. For patients during their immediate visual assessment (in comparison to the one year follow up), the odds of having a logMAR measurement of greater than 1 was 2.505 (95% CI 0.711, 8.821) times the odds of having a logMAR measurement less than 0.3 (p value = 0.0042). For patients that were unilaterally injured (in comparison to those bilaterally injured), the odds of having a logMAR measurement of greater than 1 was 9.575 (95% CI 1.301, 70.456) times the odds of having a logMAR measurement of less than 0.3 (p value = 0.0266). In the mixed model comparing satisfactory visual outcomes to moderately impaired visual outcomes, follow up time

was again a significant covariate, but injury type and whether the patient was unilaterally or bilaterally injured were not, as shown in Table 6.

DISCUSSION

Conclusions

The cross-sectional and longitudinal analysis of this data have comparable results. The logistic regression for the immediate follow up yields significance for injury type when comparing satisfactory visual outcomes and severe visual impairment, but the one month and one year follow up logistic regressions do not. Initially, this may suggest that injury type influences visual outcomes immediately after the injury, but over time blunt and penetrating injuries have no significant difference. After performing the longitudinal analysis, however, injury type was still significant when comparing satisfactory visual outcomes and severe visual impairment in the mixed model, suggesting that the one month and one year follow up logistic regression models in the cross-sectional analysis did not have a large enough sample size to detect a difference. This may also have been observed because the cross-sectional analysis only included one eye for each bilaterally injured patient. The mixed models accounted for both eyes of each bilaterally injured patient and also for the paired nature of these eyes. Because of this, the mixed models seem to better address the complexity of this dataset.

Limitations and Future Research

The sample size and loss to follow up was a major limitation during the analysis of this dataset. Because only 14 patients had visual acuity measurements for all three follow up times, there was some sparsity in the longitudinal data, decreasing the power for both the mixed models and the

logistic regression models for the one month and one year follow up. In the future, studies may try to increase the sample size or offer incentives for follow ups in order to increase the power of the analyses. Future studies can also simulate the missing observations or new datasets entirely with less loss to follow up in order to better explore different methodologies.

REFERENCE

- 1. Cooling, R. J. (1996). The burden of serious ocular injury. *British Journal of Ophthalmology, 80*, 585.
- 2. Fan, Q., Teo, Y.-Y., & Saw, S.-M. (2011). Application of Advanced Statistics in Ophthalmology. *Investigative Ophthalmology & Visual Science, 52*, 6059-6065.
- 3. Glynn, R. J., & Rosner, B. (2012). Regression methods when the eye is the unit of analysis. *Ophthalmic Epidemiology, 19*(3).
- 4. Murdoch, I. E., Morris, S. S., & Cousens, S. N. (1998). People and eyes: statistical approaches in ophthalmology. *British Journal of Ophthalmology, 82*, 971-973.
- 5. Newcombe, R. G., & Duff, G. R. (1987). Eyes or patients? Traps for the unwary in the statistical analysis of ophthalmological studies. *British Journal of Ophthalmology, 71*, 645-646.
- 6. Ray, W. A., & O'Day, D. M. (1985). Statistical Analysis of Multi-Eye Data in Ophthalmic Research. *Investigative Ophthalmology & Visual Science, 26*, 1186-1188.
- 7. Rosner, B. (1982). Statistical Methods in Ophthalmology: An Adjustment for the Intraclass Correlation between Eyes. *Biometrics, 38*(1), 105-114.
- 8. Rosner, B. (1984). Multivariate Methods in Ophthalmology with Application to Other Paired-Data Situations. *Biometrics, 40*(4), 1025-1035.
- 9. Rupesh V. Agrawal, Stephen Teoh, Xiaoling Ou, Sue Wei Ho; Validation of Ocular Trauma Score and Prognostic Factors for Vision Outcome After Surgical Repair of Open Globe Injuries - 10 year Study at a Tertiary Referral Eye Care Centre in Singapore. *Invest. Ophthalmol. Vis. Sci.* 2011;52(14):5627.
- 10. Williamson, J., & Kim, K. (1996). A Global Odds Ratio Regression Model for Bivariate Ordered Categorical Data from Ophthalmologic Studies. *Statistics in Medicine, 15*, 1507- 1518.
- 11. Ying, G.-s., Maguire, M. G., Glynn, R., & Rosner, B. (2017). Tutorial on Biostatistics: Statistical Analysis for Correlated Binary Eye Data. *Ophthalmic Epidemiology, 25*(1).

Appendix:

Table 1: Descriptive statistics

APPENDIX

Table 2. Baseline logistic regression for immediate follow up

Table 3. Baseline logistic regression for one month follow up

Table 4. Baseline logistic regression for one year follow up

Variable	OR	95% CI LB	95% CI UB	p-value
Age	1.029	1.004	1.05	0.0221
ISS	0.99	0.934	1.048	0.7221
GCS	0.904	0.742	1.103	0.32
Injury Type Blunt (vs Penetrating)	0.15	0.055	0.408	0.0002
Race Black/AA (vs Other)	0.812	0.367	1.795	0.6056
Gender Female (vs Male)	1.106	0.441	2.77	0.8301
Visit 1 vs 3	2.505	0.711	8.821	0.0042
Visit $2 \text{ vs } 3$	1.007	0.287	3.536	
BothEyes No (vs Yes)	9.575	1.301	70.456	0.0266

Table 5. Mixed model comparing satisfactory visual outcomes and moderately impaired visual outcomes

Table 6. Mixed model comparing satisfactory visual outcomes and severely impaired visual outcomes