

MA678 Midterm Project: Modeling Length of Rehabilitation Stay for Spinal Cord Injury Patients

Rose Determan

December 12, 2021

Abstract

Spinal cord injuries (SCI) are costly and lifelong medical conditions. The first weeks and months after an injury are crucial to recovery, but many variables impact the timeline of patient recovery. Both health care professionals and patients can benefit from an accurate prediction of length of stay. Using fixed effect linear and multi-level linear models, I modeled the number of days patients spent hospitalized in inpatient rehabilitation facilities following a SCI. Although still imperfect, the mixed effect model is more effective and accurate.

Introduction

Spinal cord injury (SCI) is a serious condition where the spinal cord is damaged and has a decreased ability to send and receive nerve signals. Symptoms can include paralysis, pain or pressure in the neck or back, weakness and inability to move any part of the body, and difficulty breathing (Mayo Clinic, 2021). Jain et al., 2015 estimates that in the United States, there are approximately 54 cases of SCI per 1 million. Medical professionals, divide the spinal cord into neurological segments. For analysis, the cervical and thoracic segments were selected. Each named segment roughly corresponds to muscle groups and functions, and damage at a level closer to the head likely indicates more serious impairment. For example, if there is an injury at the C4 neurological level, the individual will have weak deltoids (in the shoulder region) and reduced strength and sensation everywhere below this region (Young, 2021).

Since 1973 the National Spinal Cord Injury Model Systems have been collecting data in a database. The database is well managed and extensively documented. These data come from twenty-nine facilities across the United States. The database includes information on 32,159 individuals. Collected data include, injury year, age at injury, sex, use of mechanical ventilation, functional independence scores, American Spinal Cord Injury Association (ASIA) motor index scores, and ASIA sensory scores, among many others. This report focuses on modeling the number of days a patient spends hospitalized in inpatient rehabilitation. This is the number of days after a patient has been treated medically/surgically and before they are discharged from care. A model of predicted days spent in rehab is relevant to hospitals, so they can plan treatment for patients and estimate the number of patients they can care for in a given time. This information is also relevant to patients and their families, so they can also plan the patient's recovery timeline.

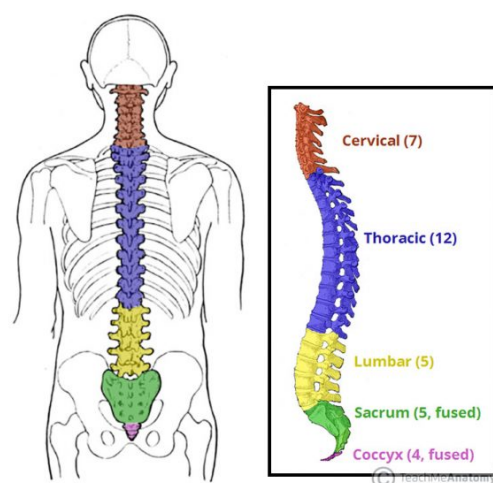


Figure 1: Source: Jones, Olivia. The Vertebral Column, 23 September 2020. <https://teachmeanatomy.info/back/bones/vertebral-column/>

Variable	Abbreviation	Definition	Notes
Total days inpatient rehab	AHDSyRb_log	Total length of stay in inpatient acute/subacute rehabilitation unit until discharge, only days for which charges are incurred	log transformed
Days from injury to rehab admission	AI2RhADa_log	Number of days from the date of injury to the first admission to the inpatient rehab unit	log transformed
BMI	ABMI_st	Calculated from provided height and weight. \$weight (lb) / [height (in)]^2 \times 703\$	centered at 0 and scaled by st.dev
Age at injury	AInjAge	Age of patient (years) on the date the SCI occurred	categorical
Functional independence motor score	AFScorRb	Total score of measures including self care, sphincter control, mobility, locomotion, and eating	valid range = 13-91
Use of mechanical ventilation	AUMVAdm_rcc	Any use of mechanical ventilation used to sustain respiration at admission to rehab?	recoded to be binary (1 = yes. 0 = no)
Vertebral injury	AVertInj	Was there a spinal fracture and/or dislocation in addition to the SCI?	binary (1 = yes. 0 = no)
Spinal surgery	ASpinSrg	Was there surgery performed? Includes limited list of relevant surgeries.	binary (1 = yes. 0 = no)
Associated injury	AAsscInj	Was there an additional injury? Includes a limited list of injuries, for example, chest injury, fractures, and amputations	binary (1 = yes. 0 = no)
Neurologic level of injury	ANurLvlR_rcc	The highest point on the spine where normal sensory and motor function can be identified at admission to rehab	Categorical. Selected C01-C08 and T01-T12

Methods

First, I selected relevant variables which are listed in Table 1. Although the data set was well organized and documented, I removed missing and unknown values. In addition I centered the BMI variable at 0 and scaled it by the standard deviation. I also log transformed the variables that measured days (AHDSyRb and AI2RhADa). Any additional modifications to the data set are in the `clean_data.R` file, and brief notes are listed in the table. I selected cervical and thoracic injuries since there are 4582 and 2510 cervical and thoracic injuries in the complete data set and only 607 lumbar injuries and nine sacral injuries listed. In addition, individuals 0-14y were excluded from the analysis due to a high rate of missing data. After removing the missing data and the excluded variables, the dataset includes 5168 individuals. Next, I fit a fixed-effect model and random-effect model to predict log(days hospitalized in inpatient rehab). The random effect model includes a random intercept and random slope for the log(days from injury to inpatient rehab admission) based on neurological level of injury.

$$\begin{aligned}
 AHDSyRb_log = & AI2RhADa_log + ABMI_st + AInjAge + AFScorRb + \\
 & AUMVAdm_rcc + AVertInj + ASpinSrg + AAsscInj + ANurLvlR_rcc \\
 & \text{(fixed effect)}
 \end{aligned}$$

$$\begin{aligned}
 AHDSyRb_log = & AI2RhADa_log + ABMI_st + AInjAge + AFScorRb + \\
 & AUMVAdm_rcc + AVertInj + ASpinSrg + AAsscInj + (1 + AI2RhADa_log | ANurLvlR_rcc) \\
 & \text{(random effect)}
 \end{aligned}$$

I anticipate that the level of injury will impact the baseline, or intercept of the model. The intercept shows the expected average log(days) hospitalized in inpatient rehab, if all predictors are zero. I also anticipate that neurological level of injury would also have a varied impact on the slope of log(days) from injury to rehab admit on the log(days) hospitalized in inpatient rehab. The models were evaluated using posterior predictive checks and leave-one-out validation (loo).

Results

Exploratory Data Analysis

The mean log(days) hospitalized in inpatient rehab varies based on the neurological level of SCI. I expected high level injuries (injuries that occur closer to the head) to have longer hospital stays on average, and this was largely true. The average stay was 52 ± 1 days (95% confidence interval of the mean(CI)). Nearly 23% of individuals in the dataset had a C04 level injury and an average rehab stay of 62 ± 2 days (95%

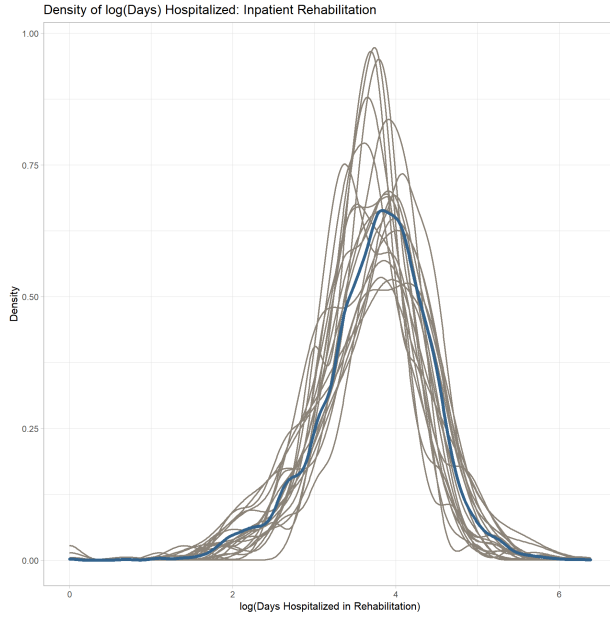


Figure 2: Density of log(days) hospitalized in inpatient acute and subacute rehabilitation unit. The blue line is the density of all selected data which excludes individuals who were never admitted either due to recovery or death. Each gray line represents a distribution of log days for a given neurologic level of injury.

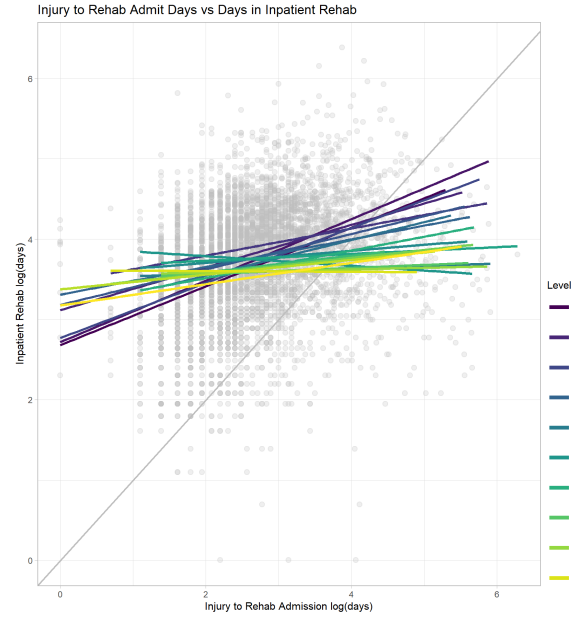


Figure 3: log(Days) from injury to rehab admission compared to log(days) spent in inpatient rehab. The lines are colored by the neurologic level of injury. Darker purple corresponds to higher on the spine, and yellow corresponds to lower on the spine. As the injury level decreases, it appears that the slope also decreases.

CI). In addition, this was also the longest stay of all injury levels. Examining the plot of comparisons, patterns and trends in the data are revealed (Figure 3). A longer time from injury to rehab admission corresponds to a longer stay in rehab, on average. Individuals who have a more complex health situation or a more traumatic injury will probably take longer to reach a point where they can be admitted to inpatient rehab where they can begin the recovery process, and they will probably spend a longer time recovering. A higher ASIA Independence Score corresponds to a shorter stay in rehab. This makes sense because if an individual starts out being able to feed themselves or go to the bathroom on their own, have less work to do in rehab when compared to someone who is less independent. Based on the boxplots, also in Figure 3, individuals who have additional injuries (AVertInj, AAsscInj), and surgeries (ASpinSrg) have a longer stay in rehab when compared to those without secondary injuries. Individuals who have a need for ventilation (AUMVAdm_rcc) also have longer stays compared to those without ventilator assistance. Again these relationships make logical sense when thinking about the impact of additional trauma on the length of recovery.

The various levels of neurological injury have different estimated slopes when predicting log(days) hospitalized in inpatient rehab from the variable log(days) from injury to rehab admission. This justifies the use of a model that includes random slopes for log(days) from injury to rehab admission. Figure 4 suggests that cervical spine level injury categories have a steeper slope when compared to the slope of the thoracic level injuries, which have slopes closer to zero.

Models

As mentioned, I fit two models to examine the differences between the fixed effect model and the random effect model. For both I used Bayesian methods for the regression. Please refer to the appendix for detailed outputs of both models and model validation. A leave one out cross validation that compares the fixed effect model and the random effect model shows that the random effect model is a better fit. The p_{loo} value returned

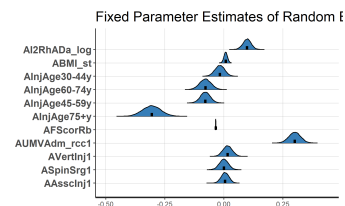


Figure 5: The parameter estimate and density for the nine parameters that are fixed in the random effect model. The estimate for those 75 and older is negative, so individuals in this group spend less time in rehabilitation than someone who is 15-29 years old. It also seems that spinal surgery and associated injuries do not play a strong role in the length of time in rehab.

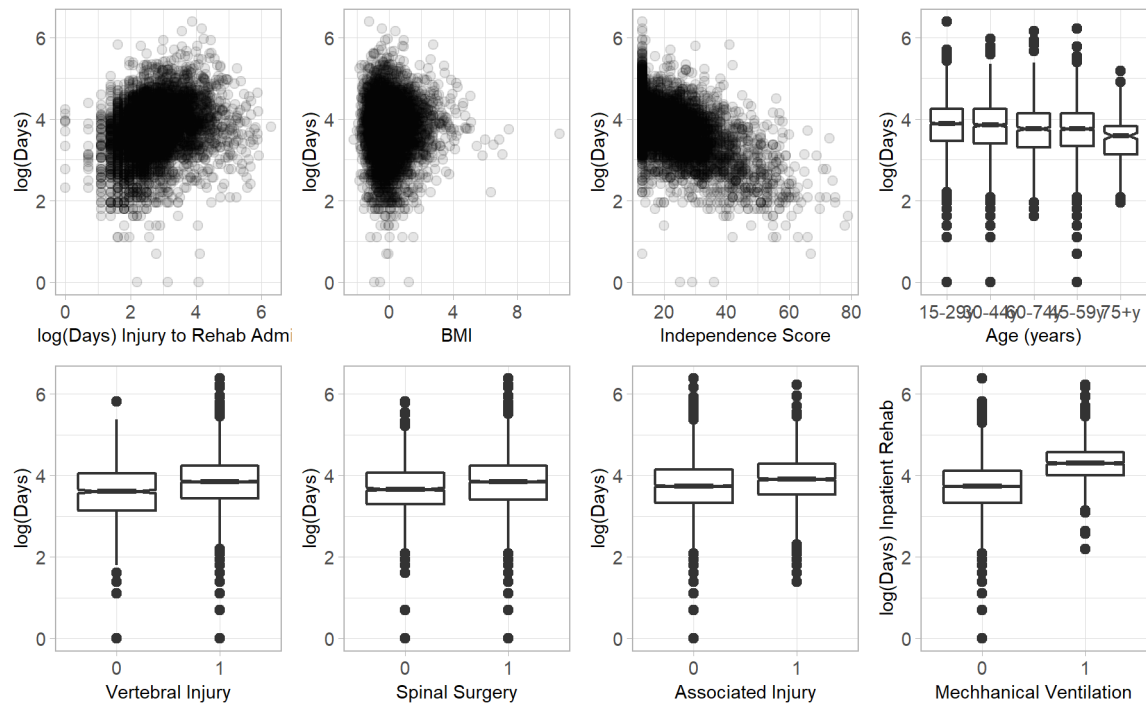


Figure 4: Selected variables compared to log(days) hospitalized in inpatient rehab. These variables were selected for inclusion in the models. Continuous variables are plotted in a scatterplot. BMI has been centered at zero and scaled to have a standard deviation of 1. Categorical or binary variables are displayed as boxplots.

from the validation indicates the number of “effective parameters.” In the random effect model, approximately 83% (43/52) of the parameters are effective while in the fixed effect model, the p_{100} value indicates weak predictive power of the model.

Looking

at

the

cater-

pillar

plots,

we

can

see

the

ran-

dom

slope

and

inter-

cept

es-

ti-

mates.

In

gen-

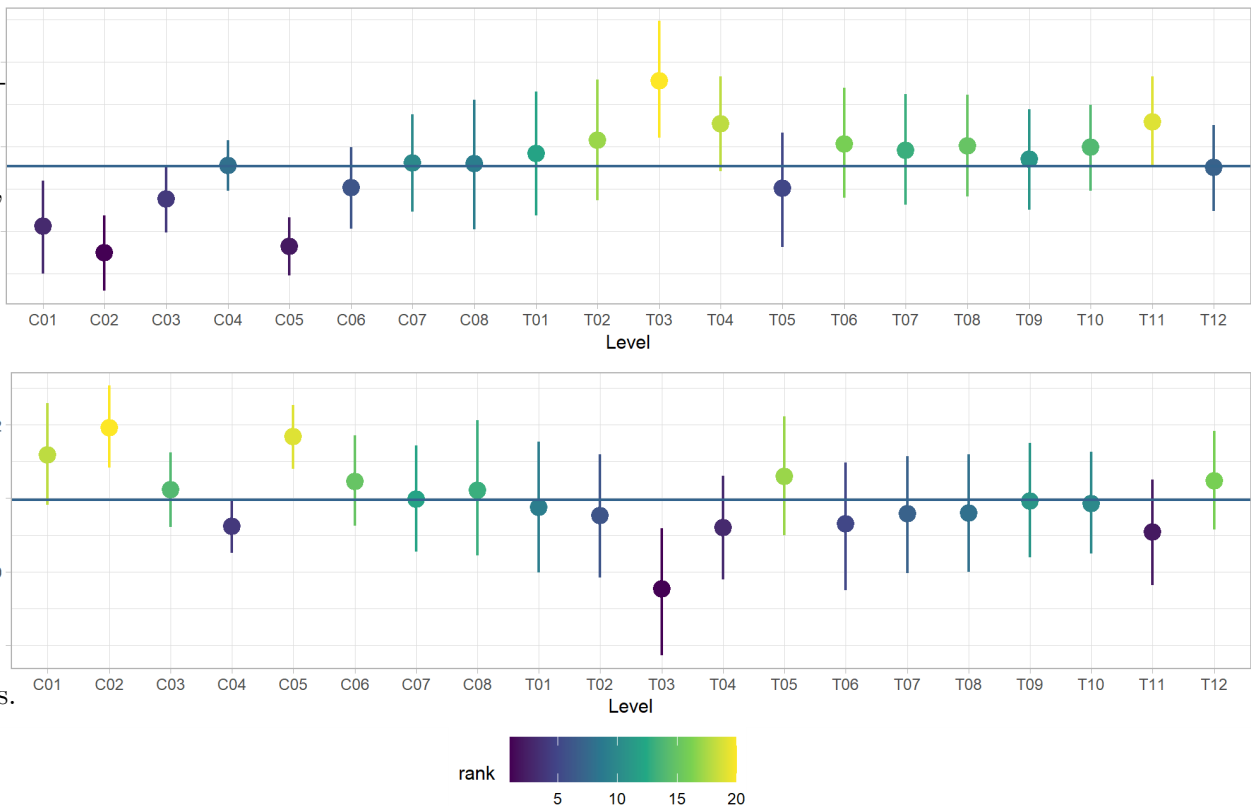


Figure 6: The caterpillar plots show the point estimate of the random intercepts at top and slopes at bottom and the 95% credible interval. The x-axis lists the level of neurologic injury from highest level to lowest when read left to right. The color represents the “rank” of the intercept or slope. When compared to the other slope is the lowest rank 1 or highest rank 20. The dark horizontal line

Injury Level	Mean log(days)	Std.Dev. log(days)
C04	3.8	0.52
C05	3.74	0.52
T10	3.97	0.52
T12	3.95	0.52

eral,
in-
juries
closer
to
each

other, for example (T06, T07, and T08) have similar coefficient estimates. If a neurologic level has an above average intercept, it also often has a below average intercept. The most prominent of this pattern is for injuries at the T03 level, since it has the largest intercept estimate and the smallest slope estimate.

I selected the median value for numeric inputs and the most common input for categorical variables, and used posterior predication with 15,000 draws to estimate the predicted log(days) spent hospitalized in inpatient rehabilitation. I also selected the two most common cervical levels injuries and the two most common thoracic levels of injuries (C04, C05, T10, T12). With these simple examples lower levels of injuries have a longer expected stay in rehab.

Discussion

In the random effect model, there were several elements that were surprising and should be investigated further. It was surprising too that patients with lower levels of injury (less of their body is impacted) were predicted to have longer stays. Maybe this is because they have more body functions, so there are more areas where they can improve and progress, while individuals with high levels of injuries are more limited. It should also be noted that these estimations also have large standard deviations which indicate a larger degree of uncertainty.

Additionally, I was not expecting that individuals in the 75+ age group would have a shorter duration of rehab hospitalization. Perhaps this is due to a lack of data for this population. SCIs are most common in young people, and in the dataset used for the models less than 5% of patients were in the 75+ group.

Works Cited

Chen, Yuying. National Spinal Cord Injury Model Systems Database. Version 2016ARPublic. Birmingham, AL: National Spinal Cord Injury Statistical Center [distributor]. <https://doi.org/10.17605/OSF.IO/NP24C>. Accessed 28 October 2021.

Jain, N. B., Ayers, G. D., Peterson, E. N., Harris, M. B., Morse, L., O'Connor, K. C., & Garshick, E. (2015). Traumatic Spinal Cord Injury in the United States, 1993–2012. *JAMA*, 313(22), 2236–2243. <https://doi.org/10.1001/jama.2015.6250>

Spinal cord injury—Symptoms and causes. (October 2, 2021). Mayo Clinic. Retrieved November 19, 2021, from <https://www.mayoclinic.org/diseases-conditions/spinal-cord-injury/symptoms-causes/syc-20377890>

Young, Wise (2021). Spinal Cord Injury Levels & Classification. Travis Roy Foundation. Retrieved November 29, 2021 from <https://www.travisroyfoundation.org/sci/resources/spinal-cord-injury-levels-classification/>