Revisiting Laser Tattoo Removal Complications on the Formerly Incarcerated and Reformed

Richard Paul Yim

16 June 2023

Contents

1 Introduction

This project revisits an older analysis that I completed while I was an undergraduate studying pure math and statistics. The original project was a joint collaboration with a nonprofit, Homeboy Industries, which is the largest gang rehabilitation program in the United States, and headquartered in Los Angeles, providing job training, employment and most important to former gang members, tattoo removal.

A lot of gang members have very visible and dramatic gang-affiliated tattoos that would otherwise prevent them from having any real job prospects after exiting the state/federal prison system(s). Therefore, it would be very important for these individuals to have visible tattoos removed if they hope to acquire serious and long-lasting employment and reintegration into society.

This project is organized as follows: the introduction section continues with some findings of the original paper and its flaws as well as the motivation for a refreshed analysis; Section 2 discusses the data and relevant wrangling with the context of the revisit; Section 3 introduces the main tool for modeling the data, namely the generalized estimating equation GLM; and Section 4 concludes the paper with implications and analysis of concordance with original findings as well as potential followup analyses.

1.1 Original Paper

The original paper published in 2022 primarly studied the following research questions: 1. What demographic factors place increased risk for complications? 2. What laser setting are associated with increased risk for complication? The original paper analyzed data at the tattoo level with randomization techniques and basic statistical inference tests. (In my defense, my undergraduate curriculum like many in today in statistics has heavier emphasis on data mining and statistical learning tools, lacking the rigor of more classical methods like mixed effects and matching techniques used in the social sciences.)

What's glaringly incorrect with much of the analysis in the paper is the very application of randomization tests. The data is historical in nature, and is essentially observational study data. It employed many nonparametric methods (e.g., rank sum and Kruskal-Wallis), but in a way that was completely unsuited for the data since the data is not randomized study data. Still, there is no implication of causal inference being made, and the project still managed to make statistically sensible conclusions on the association of certain covariates with complication occurrence as would be expected at a practical level. In other words, we can still conclude sample difference in complication occurrence across different demographic factors.

Besides various sample mean/median type tests, I also applied boosted decision trees to produce a rank of important covariates in determining treatment outcomes. In particular, the Gini impurity from the ablation of certain covariates were recorded, and used to develop a rank distribution for each covariate, from which the mode rank was used to determine the relative rank of importance for increased complication risk.

Finally, since I didn't know how to handle dynamic, varying treatment data, I created surrogate variables of the sample statistics over treatment periods for each tattoo and studied complications outcomes with respect to those statistics. For example, in the laser removal process, the laser tool has four parameter settings of laser fluence (intensity), spot size (spread of the laser spot), wavelength, and frequency. If the treatment sequence of a tattoo was of length m, I would compute

the sample variance and mean of the parameter values used over the whole treatment sequence for a particular tattoo.

1.2 Conclusions and Motivation

The original project made the following key conclusions:

1. Occurence of tattoo complication is associated with decreased likelihood of completion of tattoo removal.
2. Tattoo removal complications are more likely for tattoos with color as opposed to black/blue tattoos.
3. Greater mean laser fluence is associated with increased risk for complication as is greater laser wavelength (heavier radiation)

I presented Point 1. as the number reason why laser removal clinicians need to be extra wary of causing complications since the completion rate by z-test was lower for those with complications; Point 2. emphasized that there are certain tattoo characteristics that are more sensitive to complication occurrence; and Point 3. indicated that tattoo removal clinicians have agency during the removal process to prevent complications by carefully and selectively choose more conservative treatment practices.

In particular, I aim to reproduce the characterization made by Point 3. in this revised analysis using subject-specific generalized linear models, accounting for the implicit correlation structure in the data.

2 Data

This section will discuss the exact endpoint used in the study as well as covariates involved in this revised analysis.

2.1 Original Data

The original data consists entirely of three different datasets: 1. Patient demographic data: information such as race, ethnicity, gender, as well as other social-type factors such as reasons for tattoo removal are included. 2. Patient tattoo data: information such as number of tattoos, skin tone or Fitzpatrick score, a rating of skin pigmentation from 1 (lightest) to 6 (darkest). 3. Tattoo treatment data: information on the laser tool settings used such as spot size, fluence, over variable number of treatment appointments as well as whether the patient experienced complication in the next appointment due to that appointments procedure.

The particular details of this dataset are studied comphrehensively in the original paper. The main focus of this paper is to build a better associative model that more accurately and precisely complication risks as a function of covariates that were determined to be the most important and increased risk for complications.

In particular, the endpoint described as complications broadly encompasses the occurrence of any of the following:

- Keloids: bumpy scarring on the site of treatment. Considerably noticeable nonuniform skin surface.
- Hypopigmentation: skin lightening at the site of treatment.
- Hyperpigmentation: skin darkening at the site of treatment.
- Scarring: Similar to keloids, but with the addition of discoloration (can be simultaneously dark and light in different parts of scar).

These skin complications from laser treatment, one can surmise, is the result of the actual laser treatment. Of course individual patient attributes can determine reactivity to certain treatments, but the original paper general found some association with laser treatment settings and treatment outcomes.

2.2 Working Data and Data Exploration

With the endpoint of skin complications established we discuss the working dataset for our analysis. The working dataset used for our analysis will consist of the following covariates:

1. Colored tattoo: binary (coded 0 for blue/black, 1 for colored like red or yellow or other).
2. Male or not: binary (coded 0 for not, 1 for yes).
3. Ethnicity: Latino/Hispanic or not (coded 0 for not Hispanic, 1 for Hispanic).
4. Age: continuous varible of patient age.
5. Fitzpatrick score: scale on 1 through 6 from lightest possible to darkest possible—human graded score from pictures of patients.
6. Laser wavelength: wavelength of laser applied at treatment j in nanometers (nm).
7. Laser frequency: frequency of laser measured in Hertz (Hz).
8. Spot size: laser size measure in milimeters of radius of spot size.
9. Fluence: laser intensity and energy measured in joules/cm^2.
10. Treatment Date: time since first treatment for one tattoo on fixed patient (i.e., each tattoo is its own subject).

The key thing to note is that correlation structure occurs not only in time due to repeated measurements and treatments, but also at the patient level. Each tattoo will be the subject, not the patient, since the tattoo for the given patient is what is under specific risk for complication. With this point in mind one might ask, "what about the tattoo location?"

We do not consider this data point for two main reasons, the sample is heavily biased towards tattoos that are visible, in particular those on the arms. Additionally, even after adjusting for tattoo location, which in theory affects blood flow and healing rates for each patient, the actual application of the laser arguably is what decides complication occurrence not necessarily the latent reactivity of the patient to laser treatment.

Finally, we do some very basic data exploration by look at the correlation between the covariates to be analyzed in our model. In particular we look at the correlation of the covariates ignoring subject and between/within subject variation, we see that in Figure 1 all of the covariates have low Pearson correlation where the asterisks indicate the level of significance if we were to apply simple linear regression covariate-wise. In particular, notice that fluence and spot size have high correlation—this makes sense since lower fluence could be due to greater spot size, hence the negative but somewhat larger correlation than other covariate pairs.

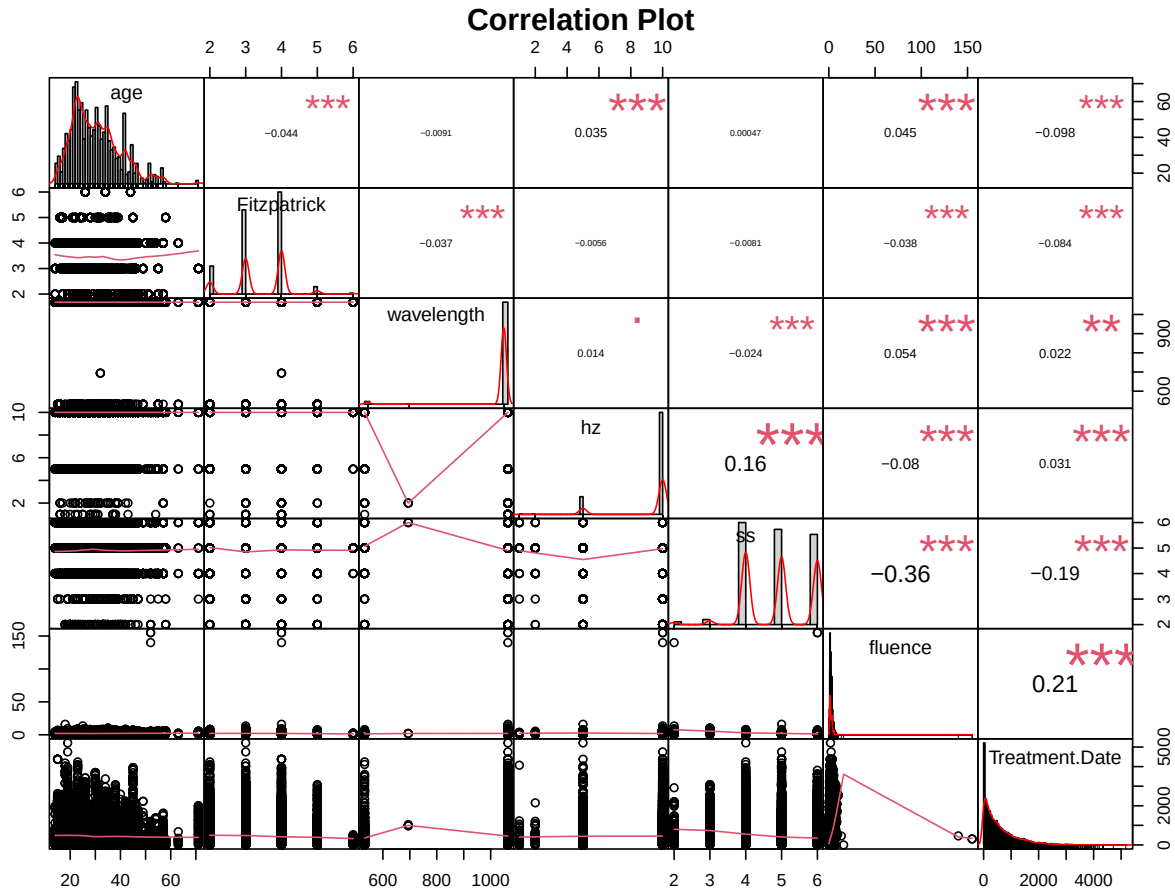

Figure 1: Correlation figure of the covariates. The upper triangle gives Pearson correlation coefficients where asterisks indicate levels of significance of a hypothetical simple linear coefficient modeled pairwise against those covariates. The lower traingle corresponds to scatterplots with the red line indicating a smoothing spline. The main diagonal corresponds to histograms of the covariates.

The working dataset has 18,518 observations corresponding to 1911 distinct tattoos across 443 different patients. Furthermore, we report that there are 18,417 nonoccurences of complication and 101 occurrences of complications. So there is quite a bit of imbalance in the response. But we claim that accounting for the correlation structure shared for this longitudinal dataset across seven years starting in 2013, using a generalized estimating equation generalized linear model (GEE GLM) will still provide robust coefficient estimates in our model—to be discussed in the next section.

Table 1: Summary statistics of continuous covariates from working dataset to be considered in our model.

age	Fitzpatrick	wavelength	hz
Min. :14.00	Min. :2.000	Min. : 532	Min. : 1.000
1st Qu.:23.00	1st Qu.:3.000	1st Qu.:1064	1st Qu.:10.000
Median :29.00	Median :3.000	Median :1064	Median :10.000
Mean :30.46	Mean :3.411	Mean :1050	Mean : 9.218
3rd Qu.:36.00	3rd Qu.:4.000	3rd Qu.:1064	3rd Qu.:10.000
Max. :71.00	Max. :6.000	Max. :1064	Max. :10.000

ss	fluence	Treatment.Date
Min. :2.000	Min. : 0.050	Min. : 0.0
1st Qu.:4.000	1st Qu.: 1.250	1st Qu.: 124.0
Median :5.000	Median : 2.000	Median : 386.0
Mean :4.897	Mean : 2.283	Mean : 577.2
3rd Qu.:6.000	3rd Qu.: 3.000	3rd Qu.: 835.0
Max. :6.000	Max. :155.000	Max. :5210.0

3 Modeling

The original paper's model incorrectly used a misspecified model that did not account for any correlation in the dataset. In particular, it simply used the following binomial regression model with logit link:

$$\begin{aligned}\text{logit}(\mathbb{E}(Y|X)) = \beta X = \beta_0 &+ \beta_1 \texttt{meanFlu}_i + \beta_2 \texttt{meanWave}_i + \beta_3 \texttt{SDss}_i + \beta_4 \texttt{MeanDays}_i \\ &+ \beta_5 \texttt{MeanFreq}_i + \beta_6 \texttt{MeanDeltaSS}_i + \beta_7 \texttt{MeanDeltaFreq}_i \\ &+ \beta_8 \texttt{MeanDeltaFlu}_i + \beta_9 \texttt{MeanDeltaWave}_i + \beta_{10} \texttt{MeanSS}_i,\end{aligned}$$

where we essentially did feature engineering to chracterize periods of treatment sequences into a single scalar value. We created statistics like the autocorrelation of different treatment parameters and mean difference of different laser parameters, where the response would be a whether the corresponding tattoo under the treatment period characterization experienced a complication. This model is clearly a bit stiff, since it only accounts for laser treatment parameters ignoring tattoo-subject demographic features entirely.

3.1 GEE GLM

In this revised analysis we will fit a new model that will allows us to account for the dynamic treatment period with varying days between appointment with varying number of appoints, and

we also add more features for a fuller and more rich model account for tattoo-characteristic features as described in the previous section.

In particular, we will consider the following model:

$$\begin{aligned}\text{logit}(\mathbb{E}(Y_{ij}|U_i)) = \mathbf{x}'_{ij}\beta = \beta_0 &+ \beta_1 \texttt{color}_i + \beta_2 \texttt{male}_i + \beta_3 \texttt{fitzpatrick}_i + \beta_4 \texttt{spotsize}_{ij} \\ &+ \beta_5 \texttt{fluence}_{ij} + \beta_6 \texttt{wavelength}_{ij} + \beta_7 \texttt{frequency}_{ij} \\ &+ \beta_8 \texttt{ethnicity}_i + \beta_9 \texttt{age}_i + \beta_{10} \texttt{daysSinceFirst}_{ij},\end{aligned}$$

where the covariates with double indices vary within subject, and the single indexed covariates vary between subejects, where subject in this setting refers to a particular tattoo. Clearly, we ignore the patient-specific correlation since that would needlessly increase our model complexity, and in some ways we incorporate patient specific correlation using patient demographic features such as whether they are male or not, or their ethnicity and age at first treatment with Homeboy Industries.

For the formulation of the working variance-covariance correlation model for V_i, since we have a variable number of time points for each subject the dimension will vary with size m_i where m_i indicates the number of treatment appointments given to tattoo i. in particular, we will have a variance structure of the following form:

$$\text{var}(Y_i) = V_i = \phi A_i^{1/2} C_i A_i^{1/2},$$

where

$$A_i = A_i(\beta) = \text{diag}\{v(\mu_{i1}), \ldots, v(\mu_{i(m_i-1)}), v(\mu_{i(m_i)})\},$$

where this matrix is of rank m_i, and v is the associated variance function of the binomial regression model with canonical logit link. Furthermore, we also have that

$$C_i = C(\alpha),$$

is just a correlation matrix for a fixed tattoo, i, where its entries are correlations of the response of tattoo i from times 1 through m_i, where m_i is variable. Table 3 shows the estimates of the covariates in our model along with the specified levels of significance.

Another important thing to note is that we specify an exchangeable correlation structure since the spacing between laser tatto appointments are generally long average at around month at a time for each tattoo (i.e., a patient can come into the clinic and get all tattoos done, and come back in the following month).

3.2 Hypothesis Testing

From our coefficient estimates we can tell that color, Fitzpatrick score, laser wavelength, and patient age are not statistically significant. We can test this more rigorously with a likelihood ratio type test, or an approximate F-test. In particular, we will test the following:

$$H_0 L : L\beta = \mathbf{0} \text{ v.s. } L\beta \neq \mathbf{0},$$

with (LaTeX compilation error):

	Estimate	Std.err	Wald	Pr(>|W|)
(Intercept)	-9.3842	2.0173	21.6398	0.0000*
ColorsYes	0.4994	0.2614	3.6482	0.0561
GenderM	-1.0388	0.2202	22.2509	0.0000*
Fitzpatrick	0.1879	0.1554	1.4625	0.2265
ss	-0.3108	0.1046	8.8267	0.0030*
fluence	0.0172	0.0087	3.9300	0.0474*
wavelength	0.0019	0.0012	2.6042	0.1066
hz	0.2404	0.1001	5.7670	0.0163*
EthnicityNot Hispanic or Latino	-0.8642	0.3660	5.5740	0.0182*
age	-0.0056	0.0118	0.2288	0.6324
Treatment.Date	0.0014	0.0001	209.1241	0.0000*

```
#L=
#\begin{pmatrix}
#0 1 0 0 0 0 0 0 0 0 0\\
#0 0 0 1 0 0 0 0 0 0 0\\
#0 0 0 0 0 0 1 0 0 0 0\\
#0 0 0 0 0 0 0 0 0 1 0
#\end{pmatrix}.
```

From this null hypothesis we can do an approximate F-test using a χ^2 approximation, where we have 4 degrees of freedom since we have four between subjects covariates such as whether patient is male, their ethnicity, age and Fitzpatrick score. Performing this F-test we find that we have a p-value of 0.0044942, so actually reject the null, and find that once jointly, these coefficients are not actually much different from zero, and we can leave them in our model.

4 Conclusions

We conclude the project in this section.

4.1 Interpretation

Since we are doing a logistic regression model accounting for the correlation structure, we acheive very robust standard error estimates allowing us to make by extension the usually odd ratio interpretation for logistic regression models (fixing all other covariates...): 1. If a patient is male, they are only 35% at risk for complications compared to nonmales. 2. If a patient is given greater spotsize laser treatment, they are only 73% at risk for complication per millimeter increase in spot size. 3. For a unit increase in Joules/cm.^{2} in fluence a patient has increased risk for complication at 1.7%. 4. A patient has 27% increased likelihood for complication if they are given laser treatment at greater laser frequency per unit increase in Hertz. 5. A patient that is not Hispanic or Latino has roughly 58% less risk for complication.

All of these conclusions are consistent with the analysis made in the original project.

For practical considerations, recommendations can be made when education patients that are hispanic/latino, since they are at increased risk for complication by this analysis. Treatment date was found to statistically significant, but for pracitcal reasons, the estimated coefficient is very small at 0.0014, such that the corresponding likelihood is only marginally greater than 1 in the thousandths.

The recommendations of more conservative laser practices remain the same. Clinicians should focus on more conserative laser removal with larger spot sizes and lower laser fluency. Additionally, clinicians should aim for lower laser frequencies since higher frequencies emulate higher radiation like ultraviolet rays that can hyperpigmentation, an instatnatiation of complication by the definition in this study.

Other patient demographic features indicate that patients that are not male are at higher risk for complication. This can suggest that clinicians should be generally more careful with laser settings and more conservative when treating females. When I was consulting for Homeboy Industries and the physician at USC Keck Medicine, they were a bit adamant about the Fitzpatrick score being related to complication occurrence, but in this robust GEE GLM we find that this is again not the case. And this is a good thing because complication occurrence can be more heavily attributed to laser treatment practices and laser parameter settings: the main point being that conservative practices are always the best.

4.2 Followup Studies

There is strong associations made from this study. A natural followup could be to perform a causal inference analysis using either optimal paired matching where we can match a tattoo to a similar tattoo using some combination of covariate balancing to adjust for hidden biases. This dataset is sufficiently rich and comprehensive to the point where we can potentially create very robust matches between tattoos with complications and those without and measure along some laser parameter setting the occurrence of complication and perform rank sum tests and corresponding sensitivity analyses.

The conclusion remains the same: conservative practice is safe practice. But we do not know if it means effective tattoo removal—this is a more nuanced question that can potentially take advantage of the very general philosophy and practice of modern “precision medicine.”

5 Appendix

5.1 Code Chunks

```
# ================================================================================
# data wrangling
# ================================================================================
# load openxlsx
if(0){
library(openxlsx)

# load CSVs, slice correct columns for iterating through
tattoo <- read.xlsx(xlsxFile = "individual.xlsx", sheet = 1,
                    skipEmptyRows = FALSE)
observations <- tattoo[,-c(1:22, dim(tattoo)[2]-1, dim(tattoo)[2])]
# check column names
colnames(observations)
# optional max print
options(max.print=100000)
# check data frame dimensions
dim(observations)

# instantiate empty list
observationslist <- list()
# begin looping
for(i in 1:nrow(tattoo)){
  for(j in 1:40){
    rangecol <- (1:25)+(j-1)*25
    candidate <- observations[i,rangecol]
    # check if observations end
    ifelse(is.na(candidate[1,1]),
           break,
           observationslist[[length(observationslist)+1]] <- c(tattoo[i,1:22],
                                                               observations[i,rangecol]))
  }
}
# create data frame for tattoo level
observations_list <- as.data.frame(do.call("rbind", observationslist))
length(unlist(observations_list$X..of.Total.Treatments))
tattooLevel <- as.data.frame(apply(observations_list, 2, unlist))

# create data frame for patient level
patientLevel <- read.xlsx(xlsxFile = "individual.xlsx",
                          sheet = 2, skipEmptyRows = FALSE)
head(patientLevel)
wrangledData <- merge(patientLevel, tattooLevel, by ="HB.ID")
```

```
# ================================================================================
# generate working dataset
# ================================================================================
# from wrangled data
colnames(wrangledData)
workingData <- wrangledData[,c("File.Order.y","HB.ID", "Colors",
                               "Gender","Ethnicity","Age.at.first.treatment",
                               "Fitzpatrick.Score.(Skin.tone,.if.applicable)",
                               "Setting.1:.Wavelength", "Setting.1:.Hz.",
                               "Setting.1:.SS", "Setting.1:.F",
                               "Side.Effects", "Treatment.Date")]

# make sure everything is a leved factor
workingData$Side.Effects <- workingData$Side.Effects=="Yes"
workingData$File.Order.y <- factor(workingData$File.Order.y)
workingData$Colors <- factor(workingData$Colors)
workingData$Gender <- factor(workingData$Gender)
colnames(workingData)[6:11] <- c("age","Fitzpatrick",
                                 "wavelength", "hz", "ss", "fluence")
workingData$fluence <- as.numeric(workingData$fluence)
workingData$ss <- as.numeric(workingData$ss)
workingData$wavelength <- as.numeric(workingData$wavelength)
workingData$hz<- as.numeric(workingData$hz)
workingData$Fitzpatrick <- as.numeric(workingData$Fitzpatrick)
workingData$Treatment.Date <- as.numeric(workingData$Treatment.Date)
workingData$Ethnicity <- factor(wrangledData$Ethnicity,
                                levels = c("Not Hispanic or Latino",
                                           "Hispanic or Latino"))

# missingness structure
apply(workingData, 2, function(x){mean(is.na(x))})
# omit missing datapoints
workingDataNoNA <- na.omit(workingData)

if(1){
  # first instance of complication data (survival process)
  for(i in unique(workingDataNoNA$File.Order.y)){
    compSequence <- workingDataNoNA[workingDataNoNA$File.Order.y==i,]$Side.Effects
    dateSequence <- as.numeric(
      workingDataNoNA[workingDataNoNA$File.Order.y==i,]$Treatment.Date)
    if("TRUE" %in% compSequence){
      newSequence <- c(rep(FALSE, length(compSequence)-1),TRUE)
      workingDataNoNA[workingDataNoNA$File.Order.y==i,
                      ]$Side.Effects <- newSequence
    } else {
      workingDataNoNA[workingDataNoNA$File.Order.y==i,
                      ]$Side.Effects <- compSequence=="TRUE"
```

```
    }
    dateSequence <- dateSequence - min(dateSequence)
    workingDataNoNA[workingDataNoNA$File.Order.y==i,]$Treatment.Date <- dateSequence
  }
}
colnames(workingDataNoNA)[1] <- "id"

write.csv(workingDataNoNA,"workingData.csv")
}

# ============================================================================
# working data
# ============================================================================
workingDataClean <- read.csv(file ="workingData.csv")


# ============================================================================
# correlation matrix
# ============================================================================
# correlations
library(xtable)
library("PerformanceAnalytics")
chart.Correlation(workingDataClean[,c("age","Fitzpatrick",
                                      "wavelength","hz","ss","fluence",
                                      "Treatment.Date")], histogram=TRUE, pch=19)
title("Correlation Plot",line=3, adj=0.45)
# create new response
library(xtable)
xtable(summary(workingDataClean[,c("age","Fitzpatrick", "wavelength","hz")]))
xtable(summary(workingDataClean[,c("ss","fluence", "Treatment.Date")]))
library(geepack)
library(nlme)
library(lme4)

# ============================================================================
# geeglm
# ============================================================================
geeglm.model <- geeglm(Side.Effects ~ Colors+Gender+Fitzpatrick
                 + ss+fluence+wavelength+hz +Ethnicity+age +Treatment.Date
                 , data=workingDataClean, id = id,
                 family=binomial(), corstr="exchangeable")

mysum <- summary(geeglm.model)

if(0){
# ============================================================================
# mixed effects logistic regression
```

```
# ============================================================================
mixed.model <- glmer(Side.Effects ~ Colors+Gender+Fitzpatrick
                  + ss+fluence+wavelength+hz +Ethnicity+age +Treatment.Date
                  +(1|id)
                  , data=workingDataClean,
                  family=binomial(link="logit"))
summary(mixed.model)
}
mysum
xtable(mysum$coefficients, digits=rep(4,5))

L <- matrix(c(
0,1,0,0,0,0,0,0,0,0,0,
0,0,0,1,0,0,0,0,0,0,0,
0,0,0,0,0,0,1,0,0,0,0,
0,0,0,0,0,0,0,0,0,1,0
), nrow=4,byrow = TRUE)
m1coef <- summary(geeglm.model)$coef$Estimate
Lb <- L%*%matrix(m1coef,nrow=11)
CLb <- L%*%vcov(geeglm.model)%*%t(L)
chi2 <- t(Lb)%*%solve(CLb)%*%Lb
options(digits=5)

dfnum <- 4
chistat <- chi2/dfnum
dft <- length(unique(workingDataClean$id)) - dfnum # degrees of freedom
1-pf(chi2/2,df1=dfnum,df2=dft) #chi-square test, p-value
#L=
#\begin{pmatrix}
#0 1 0 0 0 0 0 0 0 0 0\\
#0 0 0 1 0 0 0 0 0 0 0\\
#0 0 0 0 0 0 1 0 0 0 0\\
#0 0 0 0 0 0 0 0 0 1 0
#\end{pmatrix}.

# ============================================================================
# likelihood ratios
# ============================================================================
exp(m1coef)
# confused about "nested"
# uniform lags?
```