Final Project

Wisdom Takumah and Yohannes Tecleab

April 6, 2017

Introduction

According to the National Organization for Rare Disorders (NORD) cervical dystonia is a neurological disorder that cuase neck muscles to contract involuntarily. These spasms lead to pain and abnormal posturing of the neck and head. Prevalence of this disorder is higher in women and in older individuals. Botulinum toxin injection to the affected area is the common method of alleviating pain and disability in patients. In this project we analyze the effect Botolium injection on pain and disability. The data is from an experiment where subjects from nine different locations in the USA were randomly assigned to three groups. A placebo group, a 5000U units dose of botulinum group and a 10000U dose of botulinum group. Within each site between 18 - 41 subjects were assigned to the three treatments. After the adminstration of the treatment patients were followed for six time points including the baseline. Time points were 0 (baseline), 2, 4, 8, 12, and 16 weeks. At each time point a measure of pain based on the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS) was recorded. TWSTRS is a composite score of severity, disability, and pain.

The dataset is orginally from the book Statistical Methods for the Analysis of Repeated Measurements by Charles S. Davis, pp. 161-163 (Springer, 2002). Previously, the data has been analyzed using response profile analysis in the book Regression modeling strategies: with applications to linear models, logistic regression and ordinal regression, and survival analysis Frank E Harrel. They analyze the data using Generalized Least Squares.

```
## `geom_smooth()` using method = 'loess'
```

Table 1: Placebo group: estimated covariance matrix of pain from Cervical Dystonia at different time points

	week0	week2	week4	week8	week12	week16
week0	86.82964	83.57863	87.73992	91.47581	74.04133	75.85887
week2	83.57863	134.99597	94.80242	114.82258	110.96976	90.26613
week4	87.73992	94.80242	131.18952	139.53226	105.93750	100.58871
week8	91.47581	114.82258	139.53226	179.35484	130.23387	117.48387
week12	74.04133	110.96976	105.93750	130.23387	151.34577	139.04435
week16	75.85887	90.26613	100.58871	117.48387	139.04435	189.08065

Table 2: Placebo group: estimated correlation matrix of pain from Cervical Dystonia at different time points

	week0	week2	week4	week8	week12	week16
week0	1.0000000	0.7719708	0.8220792	0.7330198	0.6458848	0.5920370
week2	0.7719708	1.0000000	0.7123763	0.7379218	0.7763528	0.5649902
week4	0.8220792	0.7123763	1.0000000	0.9096378	0.7518220	0.6386694
week8	0.7330198	0.7379218	0.9096378	1.0000000	0.7904643	0.6379669
week12	0.6458848	0.7763528	0.7518220	0.7904643	1.0000000	0.8219486
week16	0.5920370	0.5649902	0.6386694	0.6379669	0.8219486	1.0000000

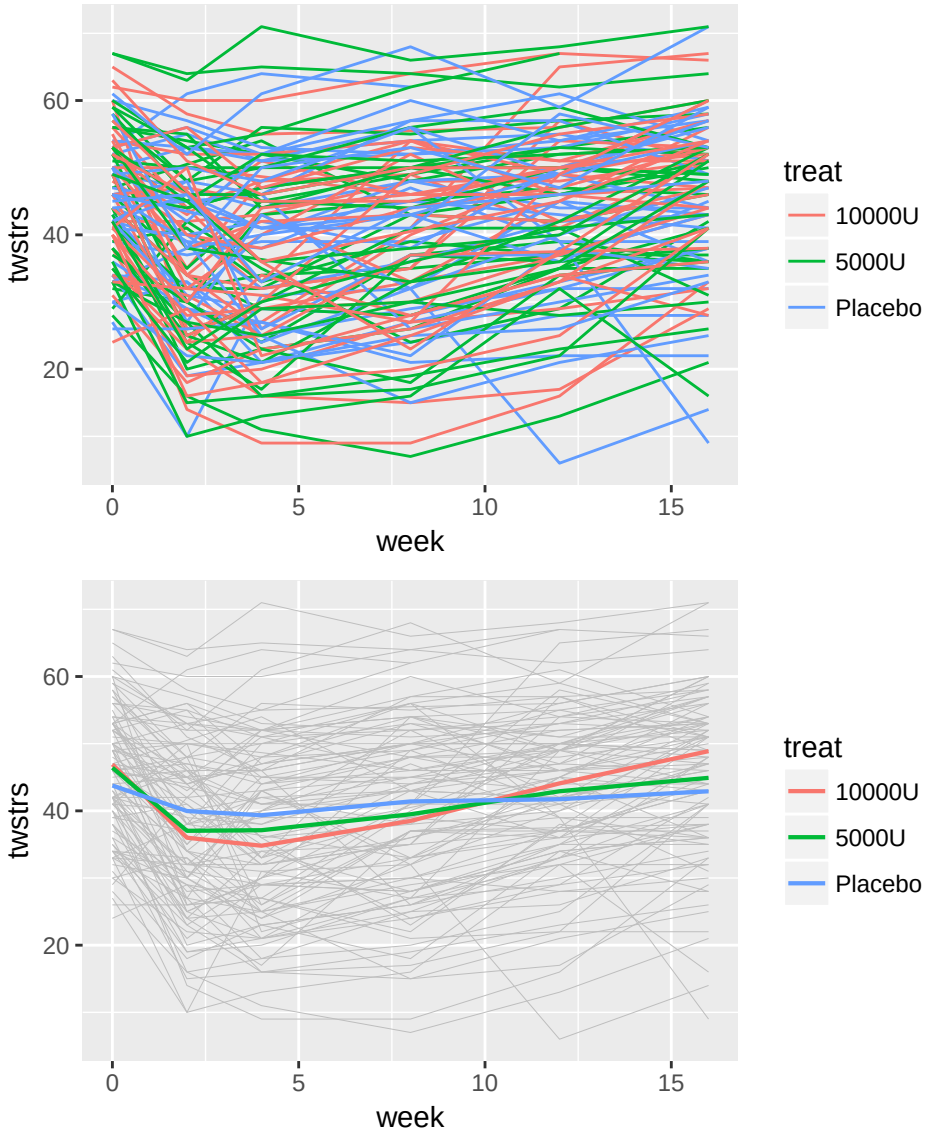


Figure 1: Profile of mean pain by sex and treatment.

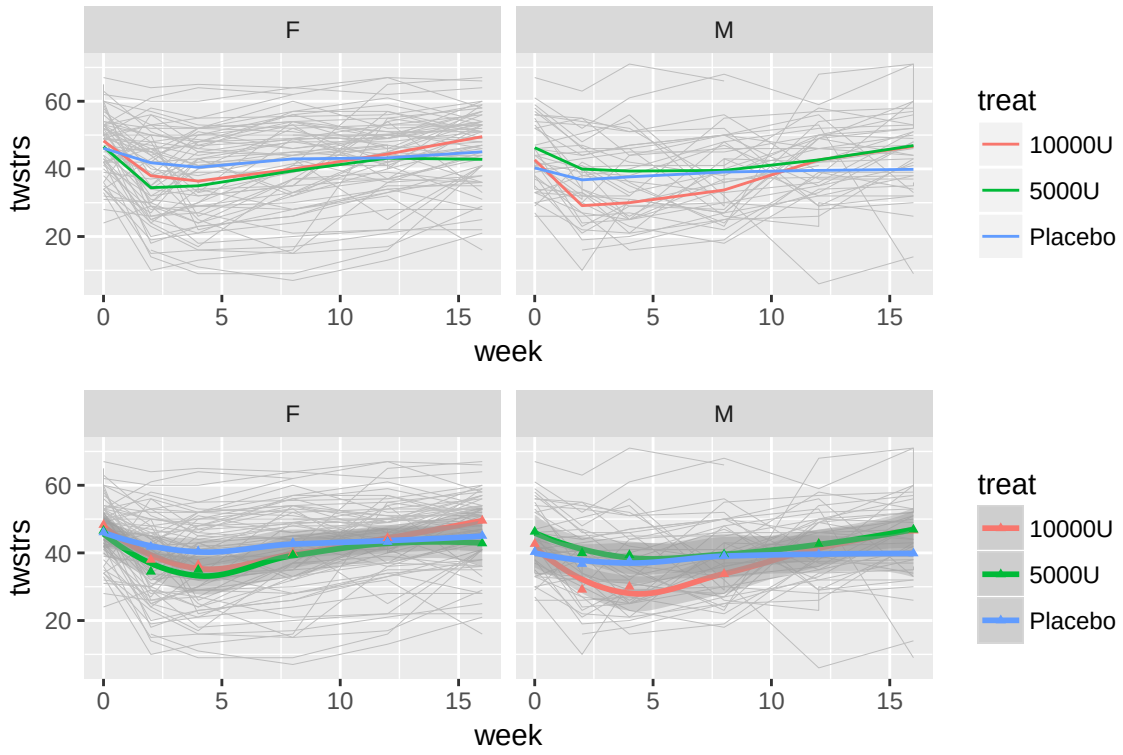


Figure 2: Profile of mean pain (top) and loess fits with mean points (bottom) for treatment and control groups.

Exploratory analysis

The data is balanced with all subjects being measured at the same time points. Exploration of the data revealed that there were several missing values (see Figure 3). In one case a subject had only one measurement at the start. This subject was removed from further analysis. For analysis of response profile and auc, the missing values were inputed using last observation forward method (LOCF).

Covariance and correlation matrices of the three groups were estimated (Tables 1- 6). In both treatment groups variance tends to initially increase in time then decrease. For the placebo group there is no clear trend in the variance. It appears that following injection the variance increases then stabilized towards the end of the experiment. the correlation matrix reveals strong positive correlation between timepoints as is expected in longitudinal studies. In addition, the correlation declines in time.

Table 3: Treatment (5000u) group: estimated covariance matrix of pain from Cervical Dystonia at different time points

	week0	week2	week4	week8	week12	week16
week0	115.2989	124.1075	143.6495	132.2516	119.9215	101.3538
week2	124.1075	205.4473	201.8538	171.5387	146.6828	137.2237
week4	143.6495	201.8538	238.4581	202.2925	176.4774	162.8935
week8	132.2516	171.5387	202.2925	202.7892	172.4677	148.4860
week12	119.9215	146.6828	176.4774	172.4677	158.6366	133.4247
week16	101.3538	137.2237	162.8935	148.4860	133.4247	151.9785

Table 4: Treatment (5000u) group: estimated correlation matrix of pain from Cervical Dystonia at different time points

	week0	week2	week4	week8	week12	week16
week0	1.0000000	0.8063717	0.8663340	0.8649000	0.8867124	0.7656604
week2	0.8063717	1.0000000	0.9119701	0.8404066	0.8125074	0.7765827

Table 6: Treatment (10000U) group: estimated correlation matrix of pain from Cervical Dystonia at different time points

	week0	week2	week4	week8	week12	week16
week0	1.0000000	0.5848859	0.6259442	0.8034396	0.7071619	0.6496213
week2	0.5848859	1.0000000	0.7613707	0.7067162	0.5730499	0.4538820
week4	0.6259442	0.7613707	1.0000000	0.8628464	0.6530755	0.5591347
week8	0.8034396	0.7067162	0.8628464	1.0000000	0.7774048	0.6703704
week12	0.7071619	0.5730499	0.6530755	0.7774048	1.0000000	0.8699948
week16	0.6496213	0.4538820	0.5591347	0.6703704	0.8699948	1.0000000

Analysis based on summary statistics

Analysis of slope and intercept

The scatter plot of intercept and slope by treatment and sex shows no clear pattern. There is a slight tendency for the slopes of hte higher dose to be higher. In addition, there is no correlation between slope and intercept. Comparison of the mean slope in relation to treatment and sex indicates that there is no effect of both teatment and sex. MANOVA analysis on slope and intercept, also reveals that there is no effect of treatment and sex.

Table 7: Anova test of mean slope by treatment and sex

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treat	2	0.9165637	0.4582818	1.1988160	0.3056832
sex	1	0.2056032	0.2056032	0.5378358	0.4649810
Residuals	104	39.7569882	0.3822787	NA	NA

Table 8: MANOVA test of mean slope and intercept by treatment and sex

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
(Intercept)	1	0.9353072	744.5699035	2	103	0.0000000
treat	2	0.0249074	0.6557595	4	208	0.6234389
sex	1	0.0185638	0.9741169	2	103	0.3809760
Residuals	104	NA	NA	NA	NA	NA

Area under the curve (auc) based analysis

Here we used the auc() function from the R package MESS. The auc() computes the area under the curve using linear or natural spline interpolation for two vectors. In the present case one vector is the weeks (x) and the corresponding y vector is twrstrs (pain) values. The function has linear interpolation and natural cubic polynomial spline options for specifiying the type of smoothing.The use of natural spline makes the auc estimate to be more accurate.

```
##
## Call:
## glm(formula = auc ~ treat + age + sex, data = auc)
```

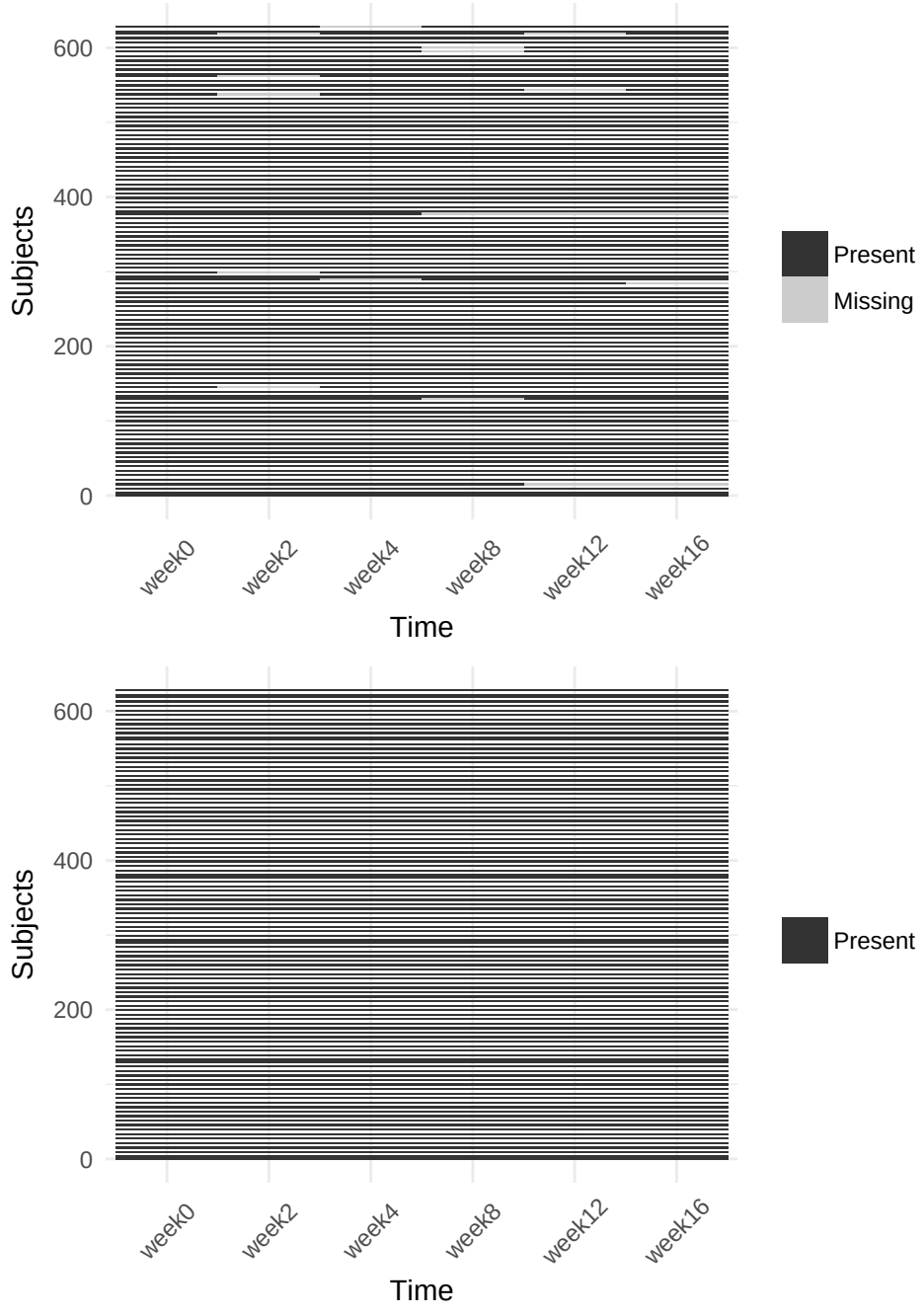


Figure 3: Missing and incomplete observations (top) prior to LOCF (bottom) after LOCF.

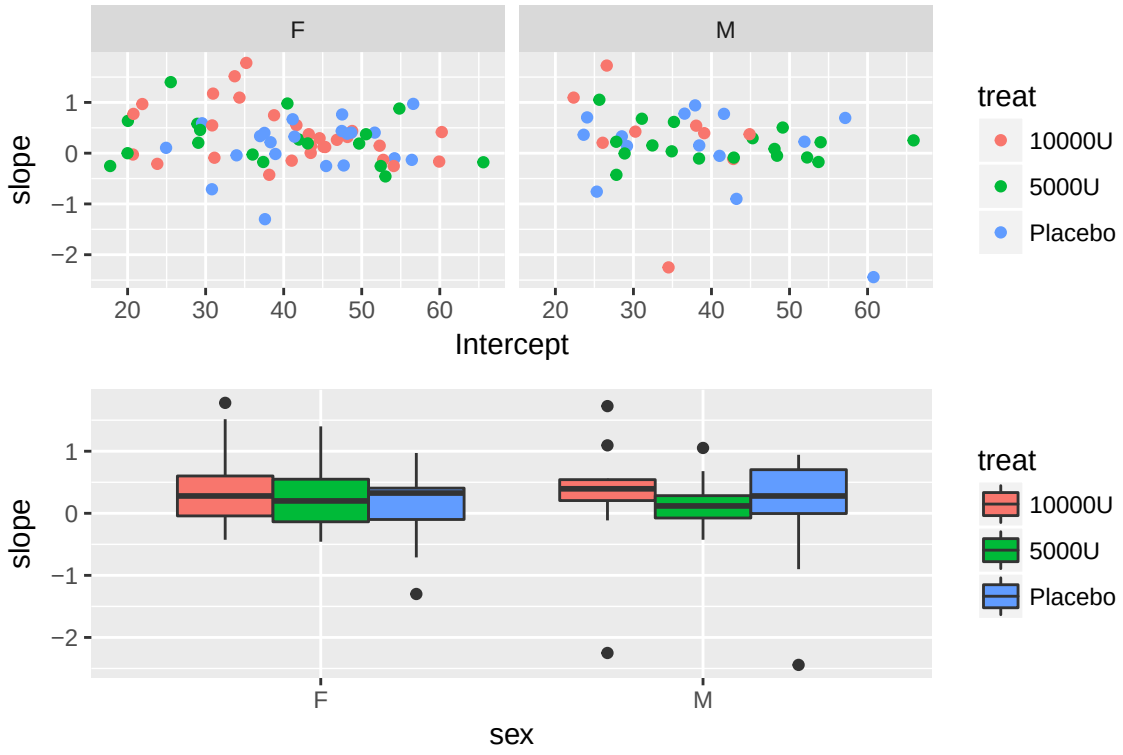


Figure 4: Box plot of slope (top) by sex and treatment (bottom) by site and treatment.

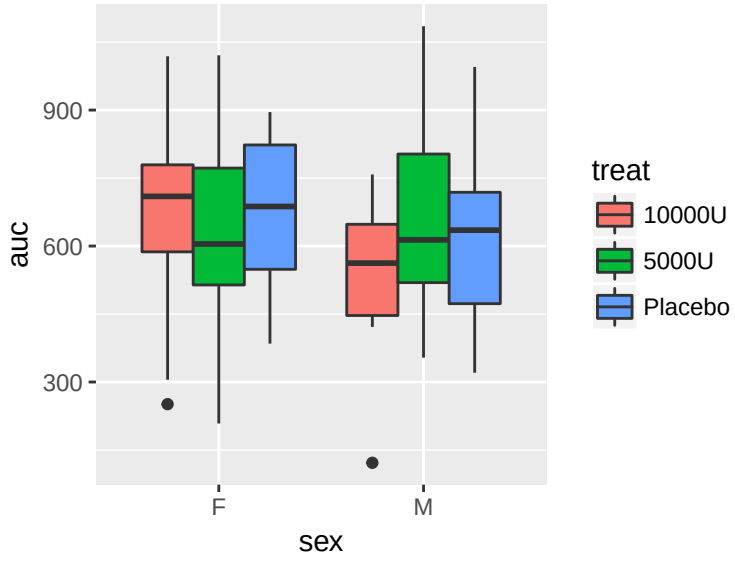


Figure 5: Boxplot of of auc (top) by sex and treatment.

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -483.28  -122.25    26.94   137.29   458.41
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  654.8315    89.8933   7.285 6.71e-11 ***
## treat5000U    16.9194    45.0641   0.375    0.708
## treatPlacebo  19.5129    44.7174   0.436    0.663
## age           -0.1981     1.5068  -0.131    0.896
## sexM         -36.5325    38.1444  -0.958    0.340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 35226.49)
##
##     Null deviance: 3664835  on 107  degrees of freedom
## Residual deviance: 3628328  on 103  degrees of freedom
## AIC: 1444.1
##
## Number of Fisher Scoring iterations: 2
```

Response profile analysis

Advantages of Response Profile Analysis

1. It allows arbitrary patterns in the mean response overtime and arbitrary patterns in the covariances of the responses, hence the method has robustness since the potential risks of bias due to misspecification of the model for means and covariances are minimal.
2. Response profile analysis can be applied to incomplete data resulting from missing response data.
3. Response profile analysis does not coerce the analyst to test certain hypothesis that are not scientifically meaningful unlike the "traditional profile analyses".
4. It permits alternative approaches for making adjustment for baseline response unlike the traditional profile analyses.
5. In response profile analyses, individuals can be group according to more than a single factor.

Disadvantages of Response Profile Analysis

1. It cannot be applied when vectors of repeated measures are obtained at different sequences of time, except by moving an observation to the nearest planned measurement time. Hence, the method is not well suited to handle mistimed measurements.
2. Response profile analyses ignores the time ordering of the repeated measures in a longitudinal study.
3. Response profile analyses produces overall test of effects, it may have low power to detect group differences in specific trends in the mean response overtime. Single degree of freedom tests of specific time trends are more powerful.
4. In analyses of response profiles, the estimated parameters grows rapidly with the number of measurement occasions.

```
FALSE 'data.frame': 648 obs. of  9 variables:
FALSE  $ site  : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
FALSE  $ id    : Factor w/ 19 levels " 1"," 2"," 3",..: 1 2 3 4 5 6 7 8 9 10 ...
```

```
FALSE  $ treat : Factor w/ 3 levels "10000U","5000U",..: 2 1 2 3 1 1 2 3 2 3 ...
FALSE  $ age   : num  32 37 31 26 42 26 38 7 19 14 ...
FALSE  $ sex   : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 2 2 1 2 ...
FALSE  $ week  : Factor w/ 6 levels "week0","week2",..: 1 1 1 1 1 1 1 1 1 1 ...
FALSE  $ twstrs: num  32 60 44 53 53 49 42 34 41 27 ...
FALSE  $ ID2   : int  1 2 3 4 5 6 7 8 9 10 ...
FALSE  $ week2 : num  1 1 1 1 1 1 1 1 1 1 ...
```

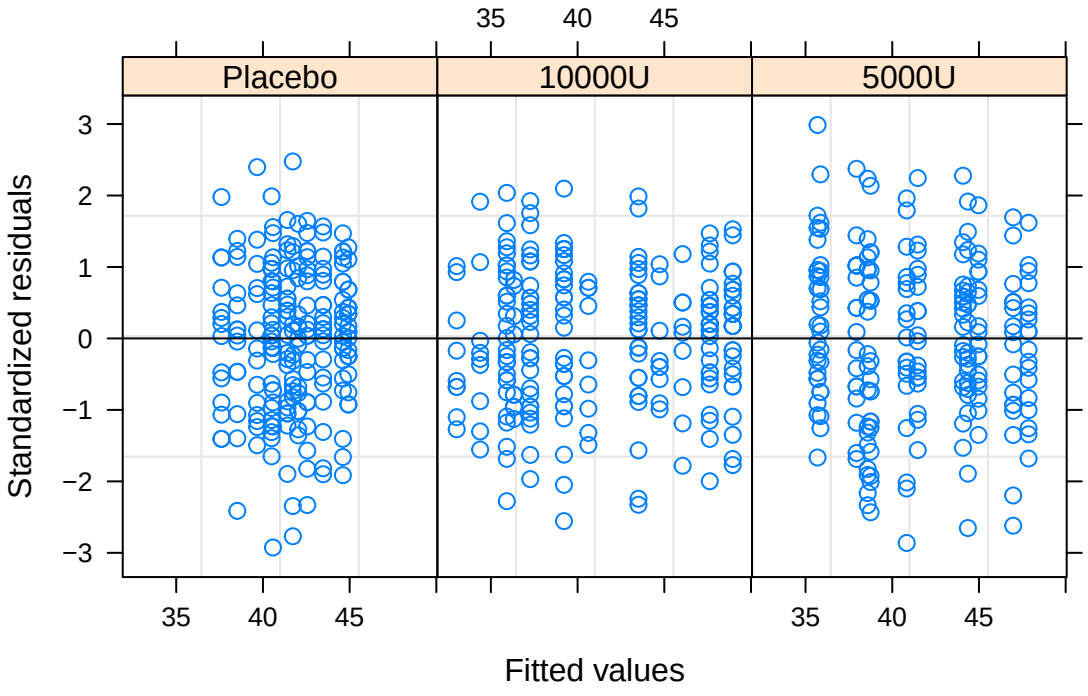
Table 9: AIC: of different model with different covariance structure

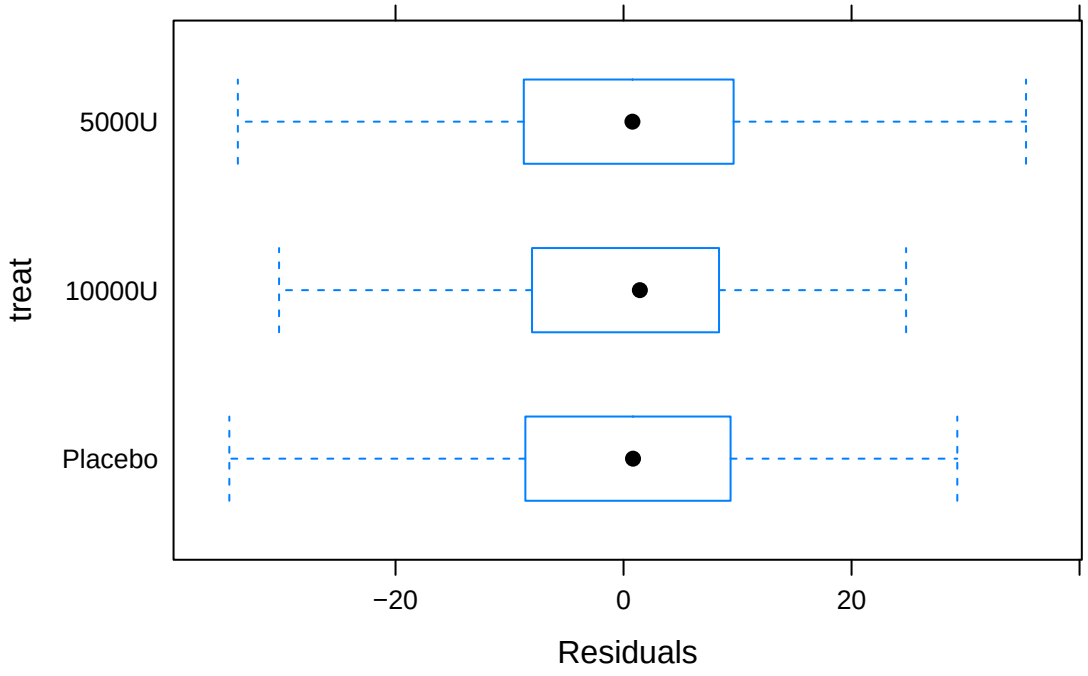
Covariance models	AIC
Unstructured covariance	4440.75314771645
Compund symmetry	4499.3331613689
Autoregressive order 1	4462.39566358877
Exponential	4462.39566358876

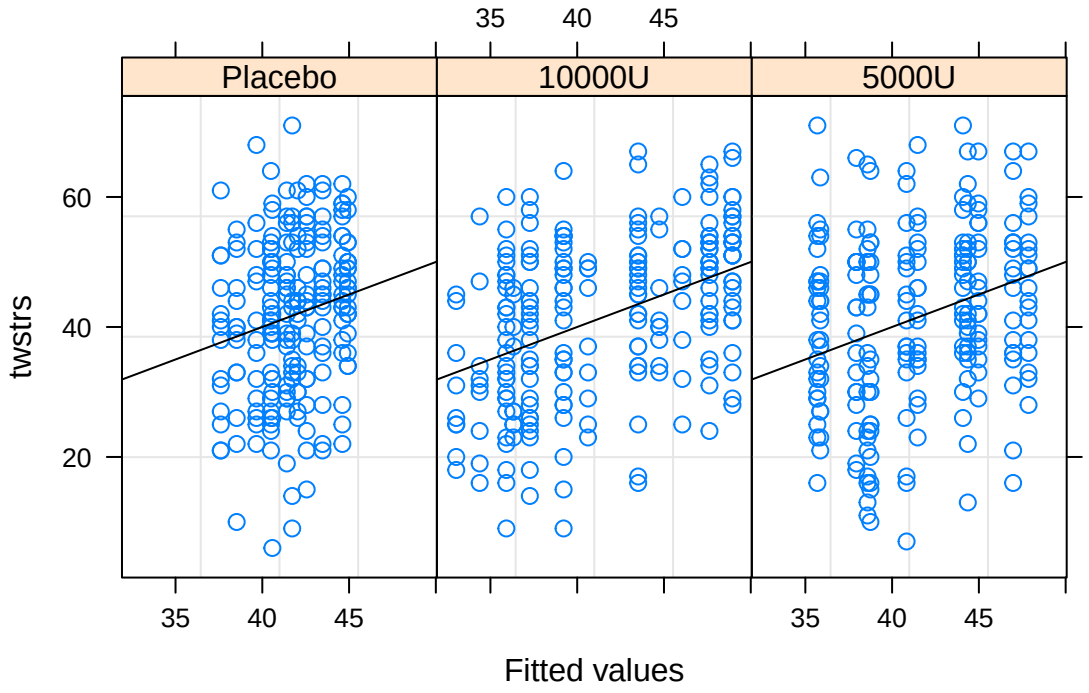
Table 10: Model coefficients

	Value	Std.Error	t-value	p-value
(Intercept)	44.9274715	2.141870	20.9758180	0.0000000
treat10000U	2.6944465	2.803591	0.9610697	0.3368864
treat5000U	2.9342488	2.812614	1.0432461	0.2972351
as.factor(week2)2	-3.5142857	1.259012	-2.7913038	0.0054090
as.factor(week2)3	-4.4285714	1.689835	-2.6207129	0.0089866
as.factor(week2)4	-2.3714286	1.968187	-1.2048798	0.2287028
as.factor(week2)5	-1.4571429	2.165603	-0.6728577	0.5012849
as.factor(week2)6	-0.3142857	2.311688	-0.1359550	0.8919003
sexM	-2.8901073	1.928996	-1.4982442	0.1345715
treat10000U:as.factor(week2)2	-6.8370656	1.756287	-3.8929103	0.0001096
treat5000U:as.factor(week2)2	-5.5968254	1.768104	-3.1654384	0.0016233
treat10000U:as.factor(week2)3	-7.2741313	2.357271	-3.0858268	0.0021190
treat5000U:as.factor(week2)3	-4.8492063	2.373133	-2.0433772	0.0414313
treat10000U:as.factor(week2)4	-6.0339768	2.745565	-2.1977178	0.0283330
treat5000U:as.factor(week2)4	-4.6563492	2.764039	-1.6846174	0.0925585
treat10000U:as.factor(week2)5	-2.6509653	3.020955	-0.8775255	0.3805363
treat5000U:as.factor(week2)5	-2.0428571	3.041283	-0.6717090	0.5020156
treat10000U:as.factor(week2)6	1.6386100	3.224740	0.5081371	0.6115353
treat5000U:as.factor(week2)6	-0.5746032	3.246439	-0.1769949	0.8595694

Model diagnostics







Mixed effects model

Our study did not analyze linear mixed effect model for the data because profile plots shows that the model is highly nonlinear. Hence we resort to semi-parametric regression which assumes that the mean function depends on parameters and nonparametric functions through a known nonlinear functional. Semi-parametric nonlinear regression models are natural extensions of both parametric and nonparametric regression models. It is sometimes difficult, if not impossible, to obtain a specific functional form. Semi-parametric regression methods such as smoothing splines (Wahba, 1990) provide flexible alternatives in these situations.

Semi-paramteric regression

We fited penalized splines to the twstrs (pain) intensity with 4 knots located at the measurement weeks (2, 4, 8,12, 16). The fiitted model is:

$$Y_{ij} = \beta_1 + \beta_2 week_{ij} + \sum_{m=1}^{4} a_m (t_{ij} - k_m)_+ + b_{1i} + b_{2i} + \epsilon_{ij}$$

```
## The following object is masked _by_ .GlobalEnv:
##
##     ID1
```

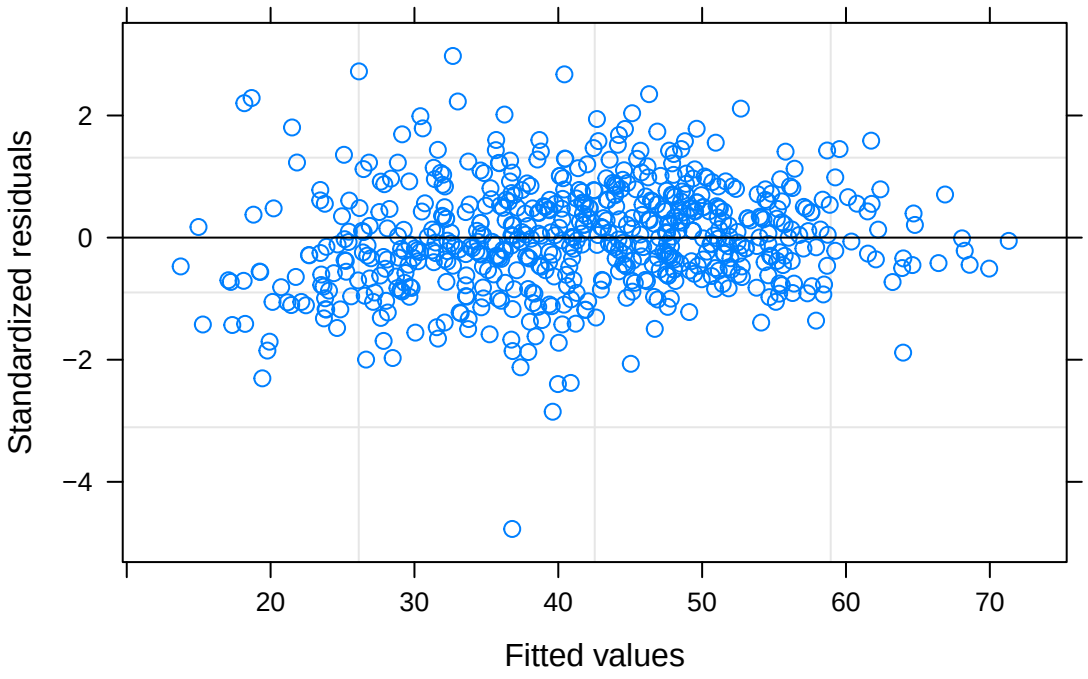
```
## Linear mixed-effects model fit by REML
##  Data: NULL
##        AIC      BIC    logLik
##   4422.671 4475.885 -2199.335
##
## Random effects:
##  Formula: ~-1 + df1 + df2 + df3 + df4 | Const
```

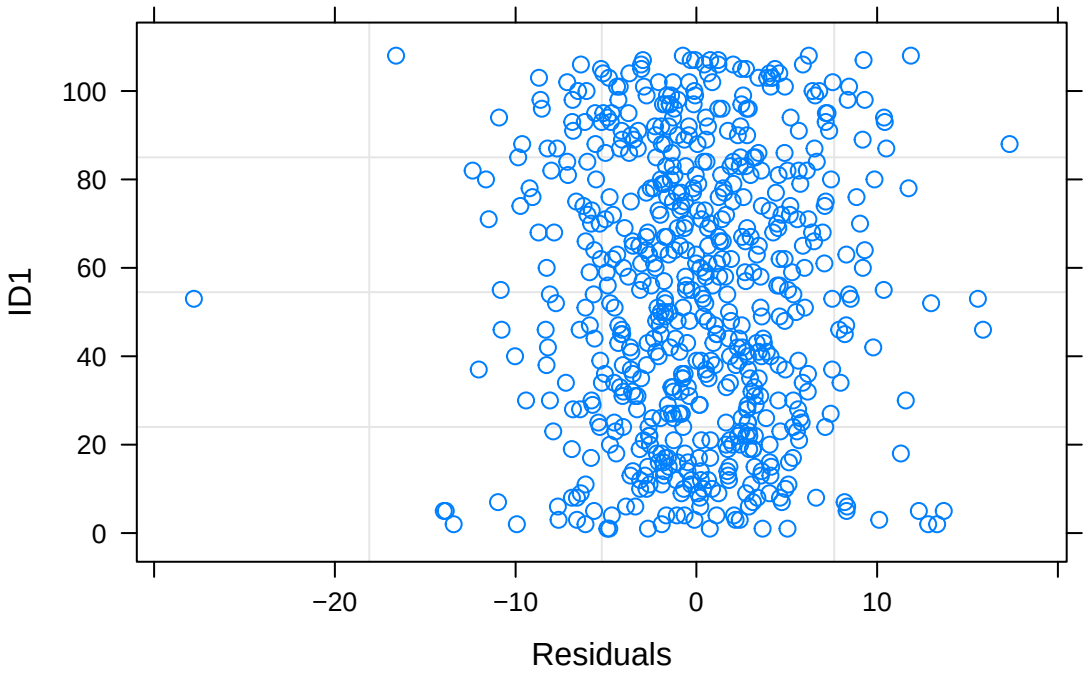
```
##  Structure: Multiple of an Identity
##              df1      df2      df3      df4
## StdDev: 1.070919 1.070919 1.070919 1.070919
##
##  Formula: ~week | ID1 %in% Const
##  Structure: General positive-definite
##             StdDev     Corr
## (Intercept) 10.5791610 (Intr)
## week         0.3635298 -0.079
## Residual     5.8252965
##
## Fixed effects: twstrs ~ week + treat + sex + treat * week
##                       Value Std.Error  DF   t-value p-value
## (Intercept)       45.03142 1.9717587 519 22.838198  0.0000
## week              -3.75489 0.4005207 519 -9.375015  0.0000
## treat5000U         1.93099 2.6955543 104  0.716362  0.4754
## treatPlacebo       2.78002 2.6761601 104  1.038810  0.3013
## sexM              -2.59555 2.2159344 104 -1.171311  0.2442
## week:treat5000U   -0.19793 0.1321569 519 -1.497663  0.1348
## week:treatPlacebo -0.31252 0.1334134 519 -2.342474  0.0195
##  Correlation:
##                   (Intr) week   t5000U trtPlc sexM   w:5000
## week              -0.234
## treat5000U        -0.587  0.040
## treatPlacebo      -0.613  0.041  0.502
## sexM              -0.274  0.007 -0.210 -0.129
## week:treat5000U    0.166 -0.166 -0.243 -0.120 -0.005
## week:treatPlacebo  0.164 -0.165 -0.118 -0.247 -0.005  0.493
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.77194502 -0.53441505 -0.02765926  0.54182788  2.97457249
##
## Number of Observations: 630
## Number of Groups:
##          Const ID1 %in% Const
##              1            108
```

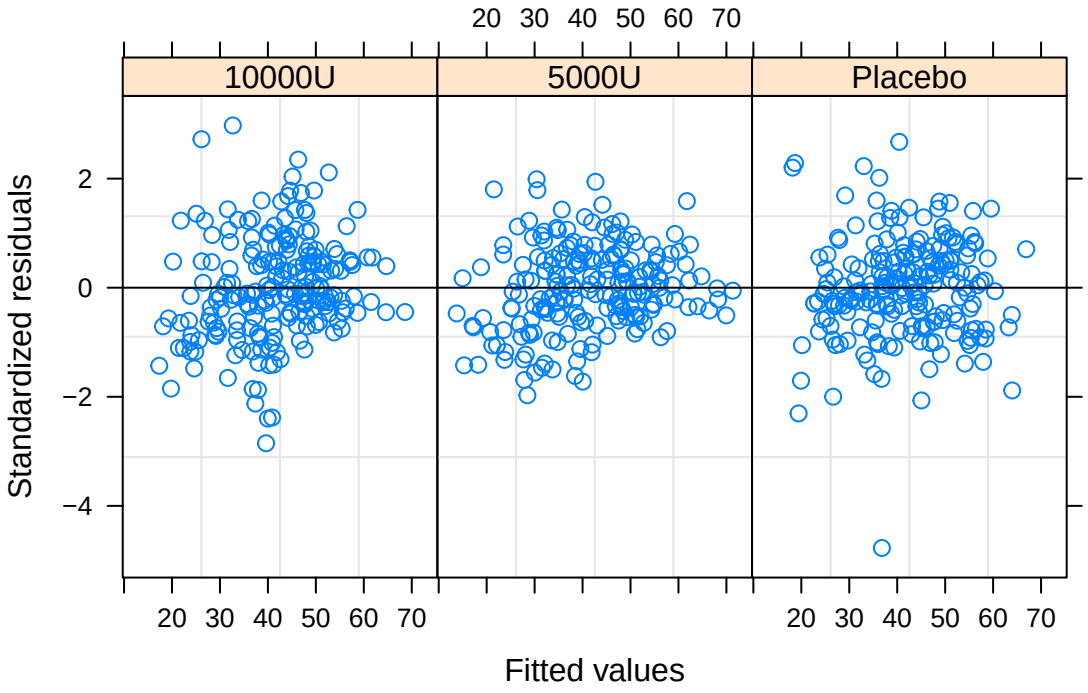
Table 11: Estimated coefficients and their standard errors

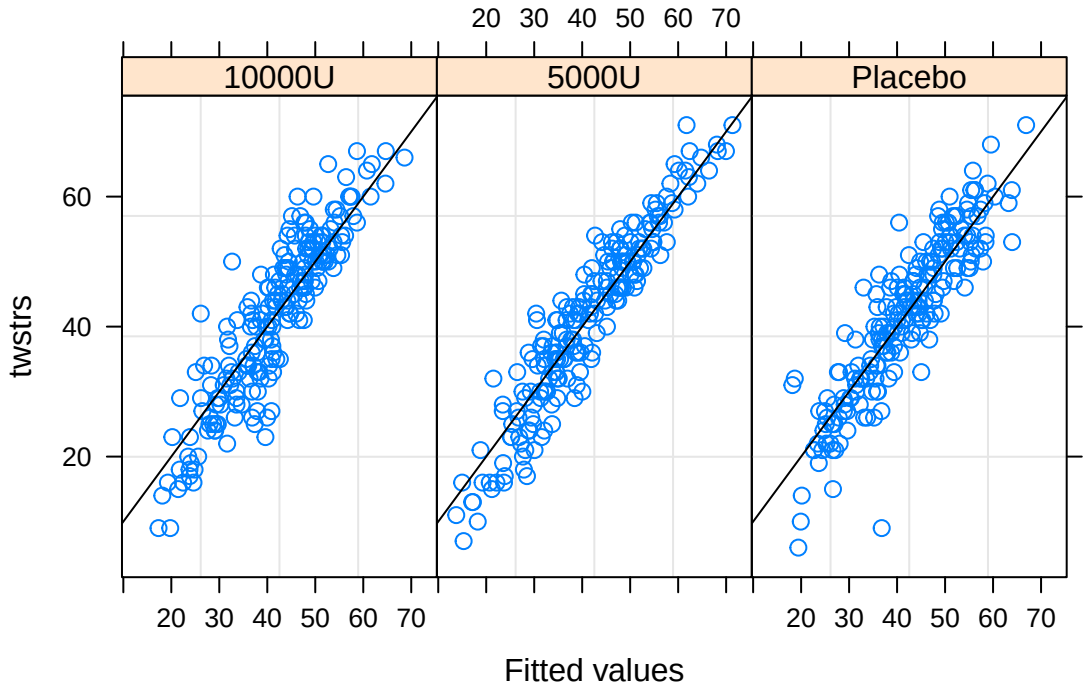
	Value	Std.Error	DF	t-value	p-value
(Intercept)	45.0314165	1.9717587	519	22.8381980	0.0000000
week	-3.7548873	0.4005207	519	-9.3750153	0.0000000
treat5000U	1.9309921	2.6955543	104	0.7163618	0.4753735
treatPlacebo	2.7800213	2.6761601	104	1.0388098	0.3013024
sexM	-2.5955482	2.2159344	104	-1.1713109	0.2441504
week:treat5000U	-0.1979265	0.1321569	519	-1.4976633	0.1348287
week:treatPlacebo	-0.3125174	0.1334134	519	-2.3424736	0.0195333

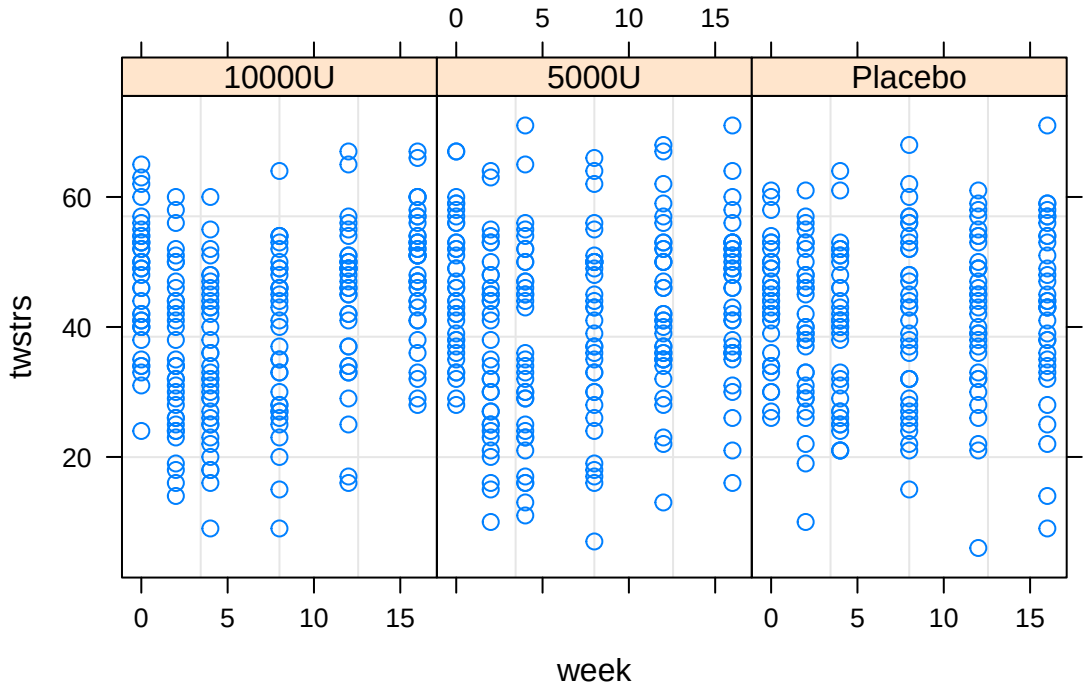
model diagnostics

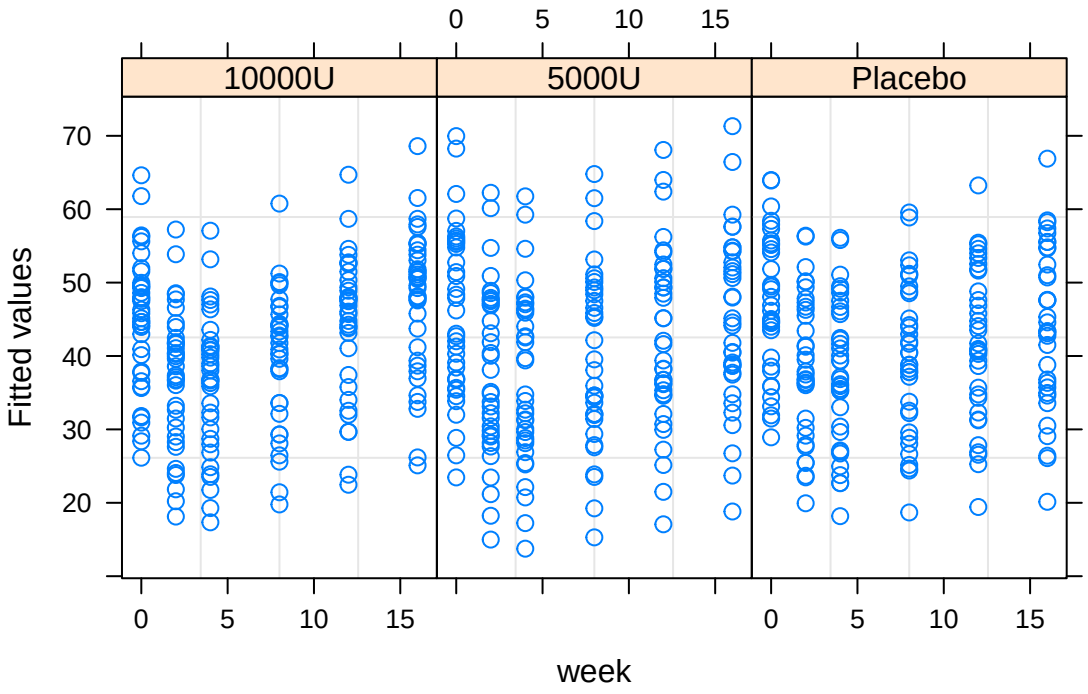





10000U
5000U
Placebo
Standardized residuals
Fitted values


10000U
5000U
Placebo
twstrs
Fitted values


10000U
5000U
Placebo
twstrs
week


10000U
5000U
Placebo
Fitted values
week

Conclusion