**The HELP Study data analysis:**

**Determine the significance of the predictor variables by using Poisson GLMM**

**Introduction:**

The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients who did not have primary care physician were randomized receiving either a multidisciplinary assessment or a brief motivational intervention or usual care assessment and intervention in order to linking them to primary medical care.

Eligible subjects were adults who spoke Spanish or English were reported to be addicted to alcohol, heroin or cocaine. They were homeless and resided in proximity to the primary care clinic. Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded.

**Methods:**

An experiment with five-randomized controlled trials and two-arm Phases were conducted on 470 eligible patients. Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every six months for two years.

Statistical Method

Statistical analysis was done on R statistical software.

Inference statistics was done using Poisson GLMM (Generalized linear mixed effect model) to assess the number of times patients entered a detox program. Poisson GLMM statistical approach was used to decide whether the treatment variable affects the number of times that patients entered a detox program and whether the effect of homelessness or substance type moderates the effect of the treatment on the number of times patients entered a detox program. Unadjusted estimates are reported for every single model.

Model Fitting

At first, I omitted missing values when analyzing the HELP data for the models. Then I handled missing values by using Multiple Imputation (MI) method in sensitivity analysis. During the analysis, I focused on the p-values of the coefficients. If the p-value is less than 0.05, the predictor variable is significant. Finally, I compared the unadjusted model without MI with that with MI.

Diagnostics

As response variable was a count, I checked whether response variable follows the Poisson distribution. The assumption for Poisson regression is that the mean and the variance of an outcome variable are equal. If the response variable satisfies the assumption, Poisson regression will be appropriate. But if the variance is greater than the mean, we have the problem of overdispersion. Therefore, Poisson regression model is not appropriate. Fortunately, the problem of overdispersion can be solved by either using quasi-likelihood estimation or negative binomial distribution instead. If there exists some zero count values in unit time, we should adjust the model by using zero-inflated model.

Handling missing values

As there are many missing values in the data, the pattern of the missing values should be checked. It shows that the missing values of the outcome variable are not missing completely at random (MCAR). So I applied the multiple imputation by chained equations method to calculate the missing values. After handling missing values, I applied Poisson GLMM to fit the unadjusted models.

**Results:**

The descriptive table on baseline (Table 1) shows that the baseline characteristics, such as age, Center for Epidemiologic Studies Depression scale, average number of drinks consumed per day, are equally distributed across treatment groups by random assignment. The average age of the study population is 35.7 years old. The male to female ratio is 3:1. It seems that the males were more likely to abuse drug and entered the detox program.

Table 1. Descriptive Statistics on Baseline

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Continuous Variables** | | | | | |
|  | | **Treatment=0** | | **Treatment=1** | |
| Mean | Std Dev | Mean | Std Dev |
| Age | | 36.53 | 7.95 | 35.28 | 7.23 |
| Center for Epidemiologic Studies Depression scale | | 33.13 | 12.45 | 33.14 | 12.56 |
| Average number of drinks consumed per day | | 17.87 | 20.10 | 16.52 | 19.94 |
| Index of Drug Abuse Consequences total score | | 36.32 | 6.90 | 35.31 | 7.30 |
| Risk-Assessment Battery drug risk score | | 4.57 | 2.92 | 4.78 | 2.58 |
| **Categorical variables** | | | | | |
|  |  | Control Group | | Treatment Group | |
| Gender | Female | 47(24.9%) | | 48(24.4%) | |
| Male | 142(75.1%) | | 149(75.6%) | |
| 1 or more nights on the street (homeless) | Yes | 89(47.1%) | | 89(45.2%) | |
| No | 100(52.9%) | | 108(54.8%) | |
| Primary substance of abuse (substance) | Alcohol | 82(43.4%) | | 63(32%) | |
| Cocaine | 58(30.7%) | | 77(39.1%) | |
| Heroin | 49(25.9%) | | 57(28.9%) | |

Model 1:

+(1|id)

The reference category in the treatment condition is the control group; the reference category in time variable is the baseline. (1|id) here is the random effect term.

Model 2:

+(1|id)

The reference category in treatment condition is the control group; the reference category in time variable is the baseline; the reference category in homeless variable is the people who had no homeless experience. (1|id) here is the random effect term.

Model 3:

+(1|id)

The reference category in treatment variable is the control group; the reference category in time variable is the baseline; the reference category in substance variable is the people who abused cocaine. (1|id) here is the random effect term.

According to the 95% confidence interval, there is no difference between the treatment group and the control group in each of the time points for model 1 (See Table 2). In model 2, the number of times people who had homeless experience entered a detox program in the treatment group is the same as people who had homeless experience in the control group (95% CI: 0.84, 1.17). The number of times people who did not have homeless experience in the treatment group entered a detox program is the same people who did not have homeless experience in the control group (95% CI: 0.75, 1.35). From model 3, the 95% confidence intervals indicates that there is no difference between treatment group and control group no matter what they abused.

As shown in Table 3 model 1 in the first appendix, the treatment variable does not significantly affect the response variable (p-value=0.39, 95% CI: -0.19, 0.07). As reported by the second model, The interaction between treatment and homeless does not significantly affect the response variable (p-value=0.85, 95% CI: -0.21, 0.28). The homeless variable significantly affects the response variable (p-value=0.009, 95% CI: 0.02,0.55). The interaction between treatment and people who abused alcohol does not significantly affect the response variable (p-value=0.36, 95% CI: -0.65, 0.29). The interaction between treatment and people who abuse heroin does not significantly affect the response variable (p-value=0.27, 95% CI: -0.43, 0.25).

Sensitivity Analysis

In order to see how accurate the Poisson GLMM method is, a multiple imputation (MI) sensitivity analysis was conducted. According to Table 2, the Poisson GLMM with MI does not have very similar estimates for the effect of the treatment. It seems that these three models have the same results in each of the time points even though they have different predictor variables. These results are different from the Poisson GLMM without MI. The treatment group is different from the control group at time point 1(95% CI: 1.18,1.74) in the model 1. For the model 2, the number of times people who had homeless experience in the treatment group is 1.15 times more likely than people who had 1 or more nights on the street in the control group to enter a detox program (95% CI: 1.01, 1.23). For the model 3, there are significant differences between treatment group and control group when people abused cocaine (95% CI: 1.01,1.31).

Table 2. Relative Rate between treatment group and control group

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Unadjusted Model without Multiple Imputation** | | | | | | | | | | |
|  | **RR**  **(CI)** | **Time Points** | | | | **Homeless**  **(Yes)** | **Homeless**  **(No)** | **Substance** | | |
|  | **1** | **2** | **3** | **4** | **Cocaine** | **Alcohol** | **Heroin** |
| **Model 1** | 0.88  (0.84,1.19) | 0.78  (0.75,1.10) | 0.86  (0.79,1.19) | 0.84  (0.95,1.37) | **-** | **-** | **-** | **-** | **-** |
| **Model 2** | 0.86  (0.66,1.28) | 0.93  (0.7,1.33) | 0.88  (0.81,1.46) | 0.93  (0.81,1.34) | 0.97  (0.84,1.17) | 0.94  (0.75,1.35) | **-** | **-** | **-** |
| **Model 3** | 0.99  (0.77,1.19) | 1.40  (0.75,1.24) | 0.84  (0.79,1.22) | 0.84  (0.98,1.33) | **-** | **-** | 1.03  (0.87,1.08) | 1.03  (0.62,1.24) | 0.96  (0.70,1.22) |
| **Unadjusted Model with Multiple Imputation** | | | | | | | | | | |
|  | **RR**  **(CI)** | **Time Points** | | | | **Homeless**  **(Yes)** | **Homeless**  **(No)** | **Substance** | | |
|  | **1** | **2** | **3** | **4** | **Cocaine** | **Alcohol** | **Heroin** |
| **Model 1** | **1.65**  **(1.18,1.74)** | 1.02  (0.87,1.15) | 0.89  (0.8,1.07) | 0.88  (0.79,1.14) | **-** | **-** | **-** | **-** | **-** |
| **Model 2** | **1.67**  **(1.15,1.78)** | 1.04  (0.88,1.11) | 0.90  (0.84,1.01) | 0.90  (0.94,1.16) | **1.15**  **(1.01,1.23)** | 1.11  (0.94,1.16) | **-** | **-** | **-** |
| **Model 3** | **1.68**  **(1.15,1.70)** | 1.04  (0.91,1.14) | 0.90  (0.85,1.05) | 0.90  (0.85,1.05) | **-** | **-** | **1.15**  **(1.01,1.31)** | 1.05  (0.87,1.22) | 1.17  (0.92,1.23) |

As reported by Table 4 model 1 in the first appendix, the coefficient of treatment is 0.12 (p-value=0.11, 95% CI: -0.02,0.28). It indicates that the treatment variable does not significantly affect the outcome variable. For the second model, the interaction between homeless and treatment significantly affects the number of times patients entered a detox program (p-value=0.00, 95% CI: -0.13,0.09). For the third model, neither the interaction of treatment with substance\_alcohol (p-value=0.46, 95% CI: -0.28, 0.12) nor the interaction between treatment and substance\_heroin is not significant (p-value=0.88, 95% CI: -0.17, 0.15). These results indicate that the interaction between treatment and substance does not significantly affect the number of times patients to enter a detox program.

**Discussion:**

    The main objective of this analysis is to test whether the treatment variable affects the number of times patients entering a detox program. The secondary objective of this analysis is to test whether the effect of homelessness or substance type moderates the effect of the treatment on the number of times patients entering a detox program. According to the results, the intervention treatment is not as effective as the standard treatment in terms of affecting the number of times patients entering a detox program. Poisson GLMM without MI results implies that predictor variables are statistically insignificant. The small sample size resulting from the missing values might be the main cause of statistically insignificance of predictor variables. These results are not consistent with the results from applying Poisson GLMM with MI.

The missing data poses a statistical challenge for the analysis of the outcome variable. It might be the cause why the primary analysis cannot find any significant results.

**Conclusion:**

There is no difference between treatment group and control group at each of the time points. The number of times people who had homeless experience entered the detox program has difference between treatment group and control group. However, the number of times people who did not have homeless experience entered the detox program has difference between treatment group and control group. The number of times people who abused alcohol and heroin entered the detox program has no difference between treatment group and control group, but there are differences between treatment group and control group when people abuse cocaine.

**References:**

[1] Alan Agresti. Categorical Data Analysis. Jone wiley & Sons, 2002.

[2] R.D. Cook. Residuals and Influence in Regression. Chapman and Hall, 1982

[3] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. Applied Longitudinal Analysis. Wiley, 2004

[4] N. J. Horton, E. Kim, and R. Saitz. A cautionary note regarding count models of alcohol consumption in randomized controlled trials. BMC Medical Research Methodology, 7(9), 2007

[5] N. J. Horton and S. R. Lipsitz. Multiple Imputation in practice: comparison of software packages for regression models with missing variables. The American Statistician, 55(3):244-254,2001

[6] K-Y Liang and S L Zeger. Longitudinal data analysis using generalizing linear models. Biometrika, 73:13-22,1986

[7] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and the R core team. Nlme: linear and nonlinear Mixed Effects Models, 2008. R package version 3.1-90

**Appendix 1:**



Figure 1: Bar Plot for the response variable (e2b)

From Figure 1, the frequency figure showed that there were no zero values in the response variable. It seemed that the response variable was not distributed as Poisson distribution. The Good-of-Fit test told us that the fit is poor (p-value=3.6e-39). It seemed that Poisson model did not appear to be tenable.

Macintosh HD:Users:mlu:Courses:CPH576C:IndividualProject:interactionplot.pdf

Figure 2: Interaction Plot: Treatment VS. Time Points

From the Figure 2, it indicated that the treatment group was different from the control group at each of the time points.

Table 3: Unadjusted GLMM Estimates without MI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Predictor Variables | Estimates | P-value | 95% CI | |
| Lower | Upper |
| **Model 1** | Intercept | 0.80 | 0 | 0.69 | 1.03 |
| Treatment1 | -0.10 | 0.39 | -0.19 | 0.07 |
| Time1 | -0.14 | 0.24 | -0.35 | 0.05 |
| Time2 | -0.17 | 0.16 | -0.38 | 0.07 |
| Time3 | -0.09 | 0.44 | -0.27 | 0.13 |
| Time4 | -0.07 | 0.56 | -0.24 | 0.14 |
| Treatment1\*Time1 | -0.03 | 0.85 | -0.15 | 0.28 |
| Treatment1\*Time2 | -0.15 | 0.60 | -0.38 | 0.11 |
| Treatment1\*Time3 | -0.05 | 0.78 | -0.28 | 0.28 |
| Treatment1\*Time4 | 0.27 | **0.11** | **0.09** | **0.56** |
| **Model 2** | Intercept | 0.62 | 0 | 0.37 | 0.88 |
| Treatment1 | -0.03 | 0.86 | -0.36 | 0.32 |
| Time1 | -0.13 | 0.26 | -0.32 | 0.03 |
| Time2 | -0.15 | 0.21 | -0.25 | 0.10 |
| Time3 | -0.10 | 0.43 | -0.38 | 0.09 |
| Time4 | -0.04 | 0.75 | -0.27 | 0.24 |
| Homeless | 0.31 | **0.009** | **0.02** | **0.55** |
| Homeless\*Treatment1 | -0.03 | 0.85 | -0.21 | 0.28 |
| Treatment1\*Time1 | -0.12 | 0.52 | -0.47 | 0.18 |
| Treatment1\*Time2 | -0.04 | 0.84 | -0.35 | 0.25 |
| Treatment1\*Time3 | 0.25 | 0.14 | -0.07 | 0.43 |
| Treatment1\*Time4 | -0.11 | 0.51 | -0.45 | 0.26 |
| **Model 3** | Intercept | 0.82 | 0 | 0.78 | 0.96 |
| Treatment1 | 0.03 | 0.83 | -0.28 | 0.15 |
| Time1 | -0.15 | 0.21 | -0.31 | 0.06 |
| Time2 | -0.17 | **0.15** | **-0.39** | **-0.04** |
| Time3 | -0.09 | 0.44 | -0.34 | 0.08 |
| Time4 | -0.08 | 0.52 | -0.20 | 0.07 |
| Substance1 | -0.19 | 0.21 | -0.48 | 0.05 |
| Substance2 | 0.10 | 0.45 | -0.16 | 0.21 |
| Substance1\*Treatment1 | -0.001 | 0.36 | -0.65 | 0.29 |
| Substance2\*Treatment1 | -0.09 | 0.27 | -0.43 | 0.25 |
| Treatment1\*Time1 | -0.04 | 0.99 | -0.23 | 0.20 |
| Treatment1\*Time2 | 0.31 | 0.63 | -0.28 | 0.29 |
| Treatment1\*Time3 | -0.20 | 0.82 | -0.18 | 0.25 |
| Treatment1\*Time4 | -0.21 | **0.07** | **0.24** | **0.43** |

Table 4. Unadjusted GLMM Estimates with MI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Predictor Variables | Estimates | P-value | 95% CI | |
| Lower | Upper |
| **Model 1** | Intercept | 0.73 | 0.00 | 0.62 | 0.72 |
| Treatment1 | 0.12 | **0.11** | -0.02 | 0.28 |
| Time1 | -0.11 | **0.09** | -0.17 | -0.001 |
| Time2 | 0.17 | 0.004 | 0.06 | 0.30 |
| Time3 | 0.43 | 0.00 | 0.36 | 0.52 |
| Time4 | 1.05 | 0.00 | 0.93 | 1.10 |
| Treatment1\*Time1 | 0.38 | 4.7e-06 | 0.36 | 0.51 |
| Treatment1\*Time2 | -0.10 | **0.20** | -0.26 | -0.007 |
| Treatment1\*Time3 | -0.24 | 0.002 | -0.42 | -0.14 |
| Treatment1\*Time4 | -0.25 | 0.0004 | -0.46 | -0.02 |
| **Model 2** | Intercept | 0.65 | 0.00 | 0.55 | 0.72 |
| Treatment1 | 0.14 | **0.1** | 0.01 | 0.21 |
| Time1 | -0.11 | **0.08** | -0.20 | -0.01 |
| Time2 | 0.17 | 0.004 | 0.07 | 0.25 |
| Time3 | 0.43 | 0.00 | 0.35 | 0.50 |
| Time4 | 1.05 | 0.00 | 0.99 | 1.10 |
| Homeless | 0.16 | 0.03 | 0.05 | 0.25 |
| Homeless\*Treatment1 | -0.04 | 0.00 | -0.13 | 0.09 |
| Treatment1\*Time1 | 0.37 | **0.19** | 0.27 | 0.44 |
| Treatment1\*Time2 | -0.10 | 0.001 | -0.26 | 0.009 |
| Treatment1\*Time3 | -0.24 | 0.0003 | -0.35 | -0.20 |
| Treatment1\*Time4 | -0.25 | **0.68** | -0.13 | 0.09 |
| **Model 3** | Intercept | 0.69 | 0.00 | 0.58 | 0.78 |
| Treatment1 | 0.14 | **0.16** | 0.01 | 0.27 |
| Time1 | -0.10 | **0.09** | -0.21 | -0.02 |
| Time2 | 0.17 | 0.005 | 0.06 | 0.24 |
| Time3 | 0.43 | 0.00 | 0.32 | 0.50 |
| Time4 | 1.05 | 0.00 | 0.95 | 1.08 |
| Substance1 | 0.05 | **0.55** | -0.007 | 0.21 |
| Substance2 | 0.09 | **0.28** | -0.04 | 0.32 |
| Substance1\*Treatment1 | -0.09 | **0.46** | -0.28 | 0.12 |
| Substance2\*Treatment1 | 0.02 | **0.88** | -0.17 | 0.15 |
| Treatment1\*Time1 | 0.38 | 0.0004 | 0.27 | 0.49 |
| Treatment1\*Time2 | -0.10 | **0.22** | -0.19 | -0.008 |
| Treatment1\*Time3 | -0.24 | 0.003 | -0.34 | -0.17 |
| Treatment1\*Time4 | -0.25 | 0.0005 | -0.34 | -0.19 |

**Appendix 2:**

mydata <-read.table("/Users/mlu/Courses/CPH576C/Data/help\_data.csv",header=T,sep=",")

data1 <- mydata[1:386,]

data2 <- data1[,c("age","anysubstatus","cesd","cesd1","cesd2","cesd3","cesd4","d1","e2b","e2b1","e2b2","e2b3","e2b4","female","g1b","g1b1","g1b2","g1b3","g1b4","homeless","i1","i11","i12","i13","i14","id","indtot","indtot1","indtot2","indtot3","indtot4","linkstatus","satreat","sexrisk","sexrisk1","sexrisk2","sexrisk3","sexrisk4","substance","treat")]

long.data1 <- reshape(data2,idvar=c("id"),varying=list(c("cesd","cesd1","cesd2","cesd3","cesd4"),c("i1","i11","i12","i13","i14"),c("g1b","g1b1","g1b2","g1b3","g1b4"),c("e2b","e2b1","e2b2","e2b3","e2b4"),c("indtot","indtot1","indtot2","indtot3","indtot4"),c("sexrisk","sexrisk1","sexrisk2","sexrisk3","sexrisk4")),v.names=c("cesdtv","i1tv","g1btv","e2btv","indtottv","sexrisktv"),timevar="time",times=0:4,direction="long")

long.data1 <- within(long.data1, {

anysubstatus <- as.factor(anysubstatus);female <- as.factor(female)

homeless <- as.factor(homeless);linkstatus <- as.factor(linkstatus)

satreat <- as.factor(satreat);group <- as.factor(treat)

time <- as.factor(time);g1btv <- as.factor(g1btv);id <- as.factor(id)})

selected.data <- long.data1[which(long.data1$time==0),]

require(psych)

describeBy(selected.data,group=selected.data$group)

require(gmodels)

CrossTable(selected.data$group,selected.data$anysubstatus)

CrossTable(selected.data$group,selected.data$female)

CrossTable(selected.data$group,selected.data$homeless)

CrossTable(selected.data$group,selected.data$linkstatus)

CrossTable(selected.data$group,selected.data$substance)

CrossTable(selected.data$group,selected.data$g1btv)

long.data2 <- na.omit(long.data1);require(grDevices)

tN<- table(long.data2$e2btv);r <- barplot(tN, col = rainbow(20))

ggplot(long.data2, aes(e2btv)) + geom\_histogram() + scale\_x\_log10()

par(cex=0.6)

with(long.data2,interaction.plot(time,group,e2btv,lty=c(1,2,3,4),lwd=3,col=c("chartreuse","blue"),trace.label="group"))

require(lme4)

glmres1 <- glmer(e2btv~group\*time+(1|id),data=long.data1,family="poisson")

summary(glmres1);confint(glmres1, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

mySumm <- function(.) {c(beta=fixef(.),sigma=sigma(.), sig01=sqrt(unlist(VarCorr(.))))}

glmres2 <- glmer(e2btv~group\*time+homeless\*group+(1|id),data=long.data1,family="poisson")

summary(glmres2);confint(glmres2, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

glmres3 <- glmer(e2btv~group\*time+substance\*group+(1|id),data=long.data1,family="poisson")

summary(glmres3);confint(glmres3, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

help.data <- mydata[,c("id","e2b","e2b1","e2b2","e2b3","e2b4","i1","i11","i12","i13","i14","pcs","pcs1","pcs2","pcs3","pcs4","mcs","mcs1","mcs2","mcs3","mcs4","sexrisk","sexrisk1","sexrisk2","sexrisk3","sexrisk4","d1","age","substance","racegrp","cesd","cesd1","cesd2","cesd3","cesd4")]

help.data <- subset(mydata,select=-c(anysubstatus,female,g1b,g1b1,g1b2,g1b3,g1b4,homeless,linkstatus,satreat,treat))

attach(mydata);anysubstatus<- as.factor(anysubstatus)

female <- as.factor(female);g1b <- as.factor(g1b);g1b1 <- as.factor(g1b1);g1b2 <- as.factor(g1b2);g1b3 <- as.factor(g1b3);g1b4 <- as.factor(g1b4)

homeless <- as.factor(homeless);linkstatus <- as.factor(linkstatus)

satreat <- as.factor(satreat);treat <- as.factor(treat);detach(mydata)

help.data <- cbind(help.data,anysubstatus,female,g1b,g1b1,g1b2,g1b3,g1b4,homeless,linkstatus,satreat,treat)

require(mice);missing.pattern <- md.pattern(help.data)

imp1 <-mice(help.data,m=5,maxit=5,seed=23109,

method=c(rep("pmm",75),rep("polyreg",2),rep("logreg",11)));summary(imp1)

complete.data1 <- complete(imp1)

selected.data <- complete.data1[,c("age","anysubstatus","cesd","cesd1","cesd2","cesd3","cesd4","d1","e2b","e2b1","e2b2","e2b3","e2b4","female","g1b","g1b1","g1b2","g1b3","g1b4","homeless","i1","i11","i12","i13","i14","id","indtot","indtot1","indtot2","indtot3","indtot4","linkstatus","satreat","sexrisk","sexrisk1","sexrisk2","sexrisk3","sexrisk4","substance","treat")]

stripplot(imp1,pch=20,cex=1.2)

long.data3 <- reshape(selected.data ,idvar=c("id"),varying=list(c("cesd","cesd1","cesd2","cesd3","cesd4"),c("i1","i11","i12","i13","i14"),c("g1b","g1b1","g1b2","g1b3","g1b4"),c("e2b","e2b1","e2b2","e2b3","e2b4"),c("indtot","indtot1","indtot2","indtot3","indtot4"),c("sexrisk","sexrisk1","sexrisk2","sexrisk3","sexrisk4")),v.names=c("cesdtv","i1tv","g1btv","e2btv","indtottv","sexrisktv"),timevar="time",times=0:4,direction="long")

long.data3 <- within(long.data3, {anysubstatus <- as.factor(anysubstatus)

female <- as.factor(female);homeless <- as.factor(homeless);linkstatus <- as.factor(linkstatus);satreat <- as.factor(satreat);group <- as.factor(treat)

time <- as.factor(time);glbtv <- as.factor(g1btv);id <- as.factor(id)})

glmres4<-with(imp1,glmer(e2btv~group\*time+(1|id),data=long.data3,REML=F,family="poisson"))

summary(pool(glmres4))

glmres4 <- glmer(e2btv~group\*time+(1|id),data=long.data3,family="poisson")

summary(glmres4);confint(glmres4, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

glmres5<-with(imp1,glmer(e2btv~group\*time+(1|id),data=long.data3,family="poisson"))

summary(pool(glmres5))

glmres5 <- glmer(e2btv~group\*time+homeless\*group+(1|id),data=long.data3,family="poisson")

summary(glmres5);confint(glmres5, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

glmres6<-with(imp1,glmer(e2btv~group\*time+(1|id),data=long.data3,family="poisson"))

summary(pool(glmres6))

glmres6 <- glmer(e2btv~group\*time+substance\*group+(1|id),data=long.data3,family="poisson")

summary(glmres6);confint(glmres6, level = 0.95,method = "boot",nsim = 10,

boot.type = "perc", quiet = FALSE)

overdisp\_fun <- function(model) {

vpars <- function(m) {nrow(m)\*(nrow(m)+1)/2}

model.df <- sum(sapply(VarCorr(model),vpars))+length(fixef(model))

rdf <- nrow(model.frame(model))-model.df;rp <- residuals(model,type="pearson")

Pearson.chisq <- sum(rp^2);prat <- Pearson.chisq/rdf

pval <- pchisq(Pearson.chisq, df=rdf, lower.tail=FALSE)

c(chisq=Pearson.chisq,ratio=prat,rdf=rdf,p=pval)}

overdisp\_fun(glmres1);overdisp\_fun(glmres2);overdisp\_fun(glmres3)

overdisp\_fun(glmres4);overdisp\_fun(glmres5);overdisp\_fun(glmres6)