

**STATISTICAL ANALYSIS OF A LONGITUDINAL
ASTHMA STUDY**

**CONTINUATION RATIO AND GENERALIZED
ESTIMATING EQUATION ANALYSES OF A
LONGITUDINAL ASTHMA STUDY**

By

DRAGOS DANIEL CAPAN, B.Sc

A Project

Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

© Copyright by Dragos Daniel Capan

April 2005

MASTER OF SCIENCE (2005)
(Statistics)

McMaster University
Hamilton, Ontario

TITLE: Continuation Ratio and Generalized
Estimating Equations Analyses of a
Longitudinal Asthma Study

AUTHOR: Dragos Daniel Capan,

SUPERVISOR: Professor Roman Viveros Aguilera

NUMBER OF PAGES: x, 74

Abstract

Two randomized controlled trials were conducted to find out whether a new treatment for asthma has a significant effect on the patients. These were multi-center trials with a parallel design, the control arm receiving a Placebo. The data were collected over a period of about 20 days before administering the intervention and for almost 80 days after the intervention. Thus, each patient has many observations recorded, making the data longitudinal. The data are summarized using first descriptive statistics and graphical displays. Then, a continuation ratio model with a lagged covariate to account for the longitudinal aspect is used to model the data. Finally, Generalized Estimating Equations methods are used. These methods have acquired popularity in recent years to account for longitudinal correlation structures. To apply the continuation ratio, the data have to be appropriately restructured. Then, the logistic regression is used to model the symptoms. The results of this procedure show that the treatment is statistically significant. However, the goodness of fit tests show that the model is inadequate. This issue is explored in the last subsection of Chapter 3. Using Generalized Estimating Equations to analyze the number of times rescue medication was used, we concluded that there is no statistically significant difference between the Active and Control groups. However, we noticed that the use of rescue medication decreased with time from the start of treatment.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Roman Viveros-Aguilera, my supervisor, for his guidance, encouragement and patients throughout this project work.

I would also like to thank Dr. Peter Macdonald for his guidance and teaching throughout my years at McMaster.

I would also like to thank my friends for their encouragements, friendship, help and the good time that we spent together.

Finally, I would like to thank my family, especially my wife Cristina and my parents for their love, support and understanding.

Contents

1	Introduction	1
1.1	Asthma Condition	1
1.2	Asthma Trial	2
1.3	Study objectives	4
2	Descriptive Summaries of Patient Profiles and Asthma Symptoms	5
2.1	Patient profiles	5
2.2	Symptoms and Treatment Effects	8
3	Continuation Ratio Model	17
3.1	Modelling Approach	17
3.2	Computational implementation	19
3.3	Model Fitting to Asthma Data	20
3.4	Simulations for Assessing Goodness of Fit	32
3.4.1	General Testing Considerations	32
3.4.2	Goodness of fit	33

3.4.3	Simulations	34
4	Generalized Estimating Equations	44
4.1	Generalized Linear Models	45
4.2	GEE Methodology	47
4.3	Models Assuming Poisson Distribution	50
4.3.1	Independent observations	52
4.3.2	Introducing correlations using GEE	54
4.4	Model Assuming the Negative Binomial Distribution	57
4.4.1	Independent observations	58
4.4.2	Introducing correlations using GEE	60
5	Conclusions	62

List of Tables

1.1	Variables in the data set	3
2.1	Patients in participating countries	6
2.2	Baseline characteristics	7
2.3	Withdraw reasons	7
2.4	Frequencies of symptoms	8
2.5	Percentage of symptom-free days	9
3.1	Goodness of Fit statistics	23
3.2	ML estimates for AMWHEEZ model (3.6)	24
3.3	Goodness of Fit statistics	24
3.4	Parameters and odds ratios for model (3.6)	25
3.5	Goodness of fit statistics for models (3.6) and (3.8)	27
3.6	Parameter estimates for model 3.8	28
3.7	Odds ratios for model 3.8	29
3.8	Goodness of fit for model 3.6 with a lagged covariate	30
3.9	Parameters and odds ratios for model (3.6) with a lagged covariate	31

3.10	Partion for the Hosmer and Lemeshow test	35
3.11	AMWHEEZ after day -1	35
3.12	Simulation probabilities TRTMNT significant	36
3.13	Simulation goodness of fit - TRTMMNT significant	37
3.14	Simulation parameter estimates - TRTMMNT significant	37
3.15	Simulation goodness of fit - none significant	38
3.16	Simulation parameter estimates - none significant	39
3.17	Probabilities of 0,1,2,3 - both significant	39
3.18	Simulation goodness of fit - both significant	40
3.19	Simulation parameter estimates -both significant	40
3.20	Reduced sample size results	41
4.1	TAMUSE parameter estimates - independence	52
4.2	Incidence ratios TAMUSE - independence	53
4.3	TPMUSE parameter estimates - independence	54
4.4	Incidence ratios TPMUSE - independence	54
4.5	TAMUSE parameter estimates - dependence	55
4.6	TPMUSE parameter estimates - dependence	56
4.7	Incidence ratios - dependence	57
4.8	Goodness of fit for TAMUSE	58
4.9	Goodness of fit for TPMUSE	59
4.10	TAMUSE and TPAMUSE parameter estimates - negative binomial . .	59

4.11 TAMUSE and TPMUSE parameter estimates - negative binomial GEE	60
4.12 Incidence ratios - independence	61
4.13 Incidence ratios - dependence	61

List of Figures

2.1	AMWHEEZ symptom by DAY and TRTMNT	10
2.2	AMCOUGH symptom by DAY and TRTMNT	11
2.3	PMWHEEZ symptom by DAY and TRTMNT	12
2.4	PMCOUGH symptom by DAY and TRTMNT	13
2.5	AMUSE by DAY and TRTMNT	14
2.6	PMUSE by DAY and TRTMNT	15
5.1	AMBREATH by DAY and TRTMNT	66
5.2	PMBREATH by DAY and TRTMNT	67

Chapter 1

Introduction

1.1 Asthma Condition

Asthma is a chronic lung condition that it is characterized by difficulty in breathing. People with asthma have extra sensitive or hyper-responsive airways. The airways react by narrowing or obstructing when they become irritated. This makes it difficult for the air to move in and out. The narrowing or obstruction is caused by airways inflammation, meaning that the airways in the lungs become red, swollen and narrow and by bronchoconstriction, meaning that the muscles that encircle the airways tighten or go into spasm. This narrowing or obstruction can cause one or a combination of the following symptoms: wheezing, coughing, shortness of breath, chest tightness. Each symptom is usually rated from 0-3: no symptoms, mild, moderate and severe. Symptoms during the day can be different than night symptoms. Also, use of rescue medication is a sign of poor control.

Although anyone may have an asthma attack, it most commonly occurs in children by the age of 5, adults in their 30s, adults older than 65 and people living in urban communities. Other factors include family history of asthma and personal medical

history of allergies. To diagnose asthma and distinguish it from other lung disorders, physicians rely on a combination of medical history, physical examination and laboratory tests, which may include: spirometry, peak flow monitoring, chest x-rays, blood and allergy tests.

As of yet, there is no cure for asthma. However, it can often be controlled with prescription medications that may help prevent or relieve symptoms, and by learning ways to manage episodes. Various devices and formulations have been developed to deliver drugs efficiently, minimize side-effects and simplify use. With the range of devices now available it is possible for nearly all patients to take drugs by inhalation. In some circumstances oral treatment is needed and it can be given by slow acting or sustained release preparations. Some drugs are available only by oral route (Rees and Kanabar, 2000).

1.2 Asthma Trial

The trial was presented as a Case Study at the Statistical Society of Canada 2004 Annual Meeting that took place in Montreal. The Case Study was conducted and provided to SSC by GlaxoSmithKline Inc. (GSK). GSK is a world leading research-based pharmaceutical company with a powerful combination of skills and resources that provides a platform for delivering strong growth in today's rapidly changing healthcare environment. The data from two 12-week large multi-center asthma trials were combined into a single data set. The trials were conducted in 15 countries on 4 continents.

A description of the variables in the data set is given in Table 1.1.

Variable 2, ANA, is the same with ITT, except it excludes all subjects from one centre which collected data in a non-satisfactory way. Variables 4 to 9 measure the patients symptoms at day and at night and they will all be considered as outcome

Table 1.1: Variables in the data set

	Variable	Description
1	ITT	Is subject in Intent-to-treat population ?
2	ANA	Is subject in the analysis population ?
3	TRTMNT	Treatment, either Active or Placebo
4	AMWHEEZ	Day-time wheez symptom score
5	AMCOUGH	Day-time cough symptom score
6	AMBREATH	Day-time shortness of breath symptom score
7	PMWHEEZ	Night-time wheez symptom score
8	PMCOUGH	Night-time cough symptom score
9	PMBREATH	Night-time breath symptom score
10	AMUSE	Day-time rescue medication use
11	PMUSE	Night-time rescue medication use
12	DASMTDT	Date of diary card observation
13	DAY	Day number relative to when subject first received treatment (Day 0), for instance, Days -7 to -1 refer to last week of baseline
14	FREQBASE	Description of subjects severity of asthma prior to the trial, either Chronic persistent or Episodic (less severe)
15	WDW	Did subject withdraw during the study ?
16	WDWREAS	Reason for withdrawal (if applicable)
17	AGE	Age in years
18	CENTRE	Investigational site

variables. The values that can take are from 0 to 3 where 0 = no symptoms, 1 = mild symptoms (not troublesome), 2 = moderate symptoms, 3 = severe (troublesome symptoms). Variables 10 and 11, give the number of occasions on which rescue medication, Ventolin, was used to relieve asthma symptoms.

1.3 Study objectives

The objective of the trial is to see whether there is a difference between the Placebo and the Active treatment in lowering the symptoms and also the frequency of rescue medication use. The study has also been designed such that the difference in time of the symptoms can be observed.

This project is designed to assess whether differing methods of summary and analysis are more or less sensitive to detecting treatment differences, and whether the methods are appropriate for the data.

Chapter 2

Descriptive Summaries of Patient Profiles and Asthma Symptoms

2.1 Patient profiles

The purpose of patient randomization to either Placebo or Active treatment is to ensure that the two groups are comparable. Through randomization the characteristics of patients at baseline should be expected to be similar in the two groups. When the number of patients is high this is very often achieved. However, when there are not so many patients enrolled in the study, a blocked randomization is helpful in achieving the groups similarity. So first we will take a look at the baseline characteristics in the Placebo and Active arms (Altman, 1991).

The original data set has 31261 observations from 305 patients from different countries. An observation in this study is the result of monitoring and recording the asthma symptoms and the frequency of rescue medication use for one day (day-time and night-time). A frequency of the *ANA* variable revealed that 6 subjects coming from the Czech Republic should not be included in the analysis population as the data

were collected in a non-satisfactory way. These 6 subjects were deleted from the data set, thus only 299 patients being included in the analysis. From these 299 patients, 149 are in the Placebo group and the remaining 150 are in the Active treatment group. Table 2.1 summarizes the number of patients from each country. The countries are alphabetically ordered.

Table 2.1: Patients in participating countries

Country	Number of Patients	Country	Number of Patients
Belgium	17	Ireland	20
Canada	23	Israel	13
Czech Republic	15	New Zealand	22
Denmark	19	Poland	23
Germany	14	South Africa	29
Holland	38	Spain	20
Hungary	24	United Kingdom	19
Iceland	3		

Table 2.1 is interesting but not informative with respect to the baseline characteristics of the patients. The variables of most interest from this point of view are *GENDER*, *FREQBASE*, *WDW*, *STUDY* and *WDWREAS*. It is important the two arms have comparable number of patients with respect to the above variables, especially if they are believed to influence the outcome. Table 2 shows the distribution of patients in the Placebo and Active groups with respect to these variables.

Table 2.2 shows that the baseline characteristics are nearly the same in the two groups. Notice that the number of males participating in the study is about twice the number of females. It is also useful to look at the reasons patients withdraw from the study in the two groups. We would like to see no pattern. The results are in Table 2.3.

Table 2.2: Baseline characteristics

Variable		Active	Placebo
FREQBASE	Chronic	87	84
	Episodic	63	65
GENDER	Male	98	98
	Female	52	51
WDW	Yes	13	16
	No	137	133
STUDY	A	76	81
	B	74	68

Table 2.3: Withdraw reasons

Variable	Reasons	Active	Placebo
WDWREAS	A	5	8
	F	2	3
	L	3	1
	N	1	2
	O	2	2
Total		13	16

Looking at the numbers the same conclusion can be reached, that the two groups have similar characteristics. This assures their comparability.

2.2 Symptoms and Treatment Effects

Next we take a look at the symptoms. Running the frequency procedure in SAS we obtain the following table showing the number of symptom-free days for each variable.

Table 2.4: Frequencies of symptoms

Score	AM			PM		
	Wheez	Cough	Breath	Wheez	Cough	Breath
0	24,095	15,781	24,966	25,864	18,554	27,061
1	4,781	9,908	4,420	3,655	7,833	2,840
2	1,693	4,212	1,154	1,304	3,724	887
3	255	902	205	230	909	180

The high numbers in the table 2.4 should not be of surprise since, on average, each patient has about 100 observations. The number of missing observations was not shown in the table. We can see that the number of symptom-free days is much higher than the number of days with symptoms. We can also notice that frequencies decrease with the severity of symptoms. Patients seem to feel worse during the day than during the night. However, this may have been caused by the fact that the patients may have not recorded the symptoms they had at night as accurate as they recorded the ones they had during daytime. Looking at the numbers we can also say that the most frequent symptom is cough which seems to bother patients both during the day and night. Now, let's look at the percentage of symptom-free days in the two groups before and after receiving the treatment. We should keep in mind though that neither of the above summaries takes into account the correlated aspect of the data. The symptom-free days are defined as the days that have the total sum of the six symptoms equal to zero.

We can see that before day 0 the percentage of symptom free days is similar in the

Table 2.5: Percentage of symptom-free days

%	Active	Placebo
Before	20.40	20.35
After	50.98	40.74

two groups. After day 0, the percentage is increasing in both groups but the increase is bigger in the Active group. The percentage of missing observations in this variable is 1.82% and 2.38% respectively. The variable analyzed, called “al”, had missing values on the rows where at least one of the symptoms had a missing value. Looking solely at these values we would conclude that the treatment seems to have an effect.

Another way to summarize the data is by graphing the percentage of a symptom score in each day against days for each group. To do that, 6 data sets had to be created, one for each symptom, using a SAS macro. In all figures that follow, the active group is represented by a solid line whereas the placebo by a dashed one. Figure 2.1 plots AMWHEEZ.

We can notice that before day 0 the two lines intersect showing no difference between the two groups. In the first plot (upper left corner) the line representing the active group is well above the dashed line after day 0. Since it represents the percentage of AMWHEEZ symptom-free days we can conclude that the treatment is effective in relieving this symptom and that it acts better than the Placebo. The second plot tells us that the percentage of mild-symptom days is smaller in the Active group than in the Placebo. In the last two plots (for scores 2 and 3, bottom row) there is no obvious improvement for either of the groups.

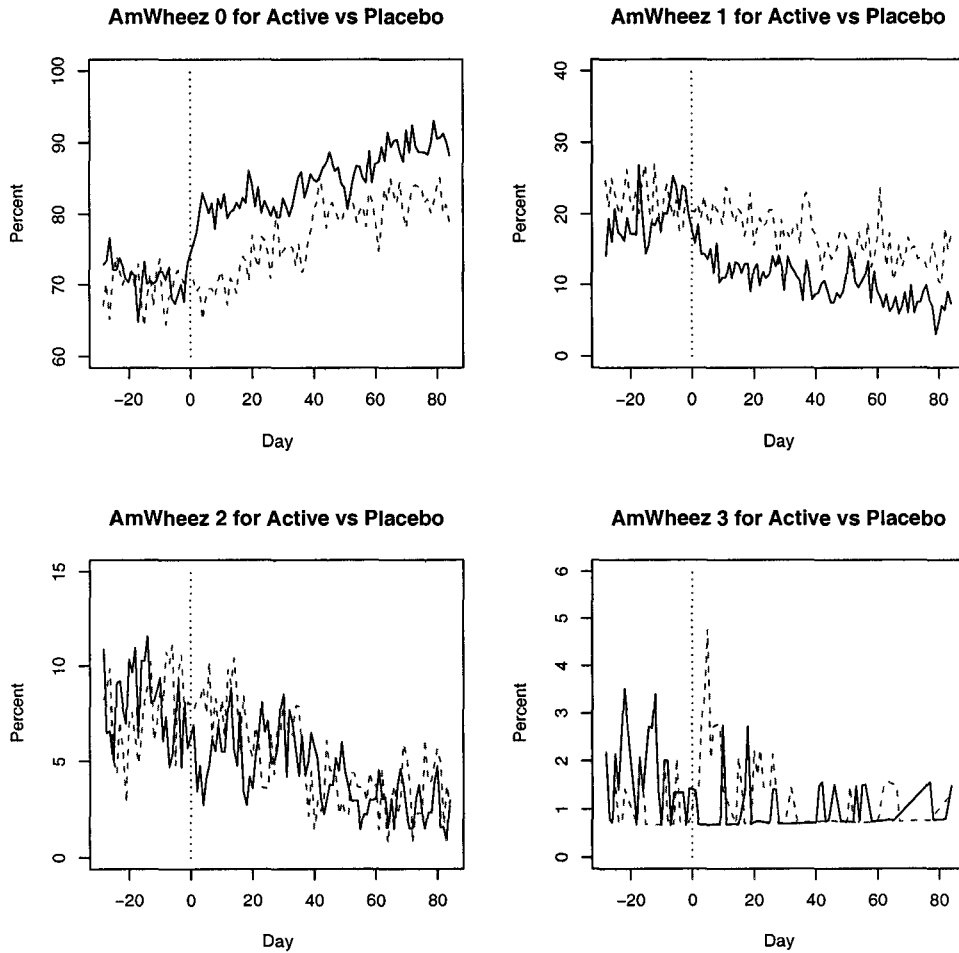


Figure 2.1: AMWHEEZ symptom by DAY and TRTMNT

We will now look at the same type of plot, for the AMCOUGH symptom (Figure 2.2).

The same conclusions can be drawn for this variable. One difference however is that for AMCOUGH=3 there seems to be a difference in favor of the active group. In all six figures, the two plots at the bottom (responses 2 and 3) do not show a clear difference between the two groups. As we noticed before, the frequency of symptoms

scored as high as 2 or 3 is not too large. Therefore, even if the treatment would make a difference, this may not be visible on such a plot. We also have to take into consideration that the plots show aggregate percentage scores and do not take into account individual change. What is to notice however is that as time goes by the trends seem to decrease. In the next two plots we will take a look at the same type of symptoms but during the night, that is, PMWHEEZ and PMCOUGH. The graphs

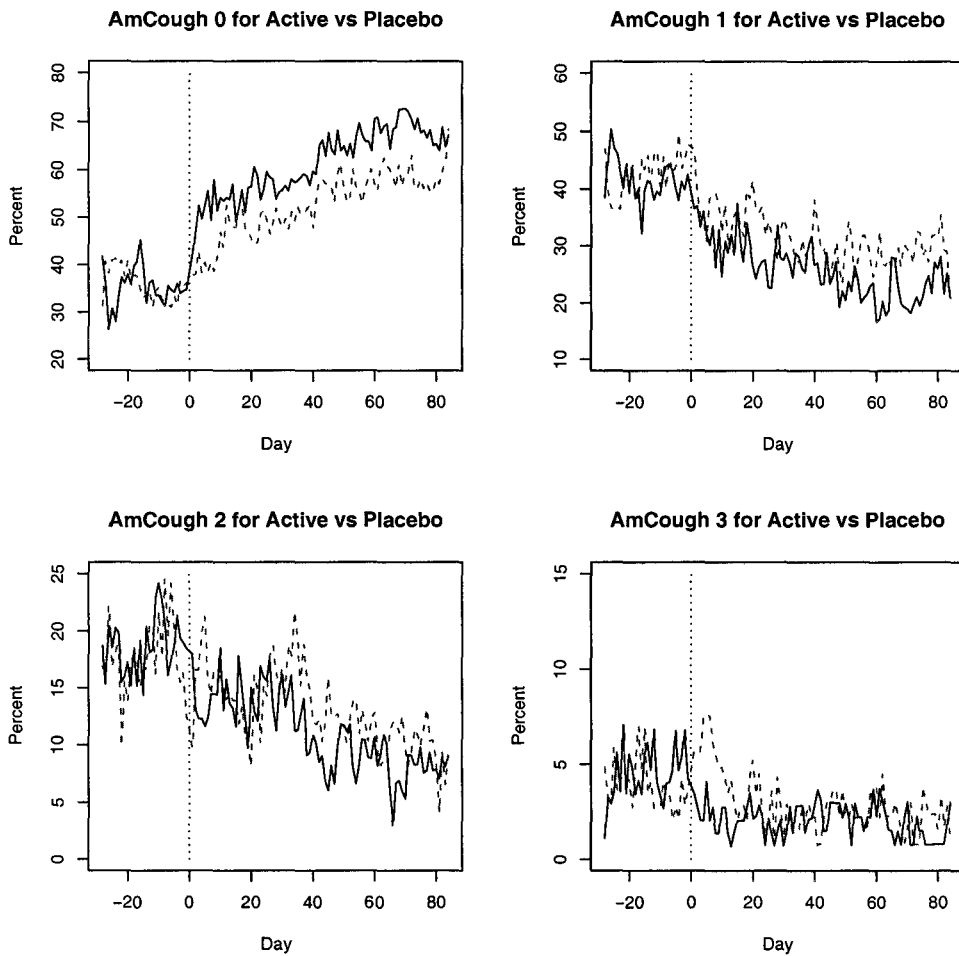


Figure 2.2: AMCOUGH symptom by DAY and TRTMNT

for AM and PM breath are in the Appendix 1.

The Figure 2.3 below shows almost the same trend for the variable PMWHEEZ as for the variable AMWHEEZ. Again we can see that the most obvious difference is for “no symptom” and “mild symptom” whereas for the other two there is no clear difference between the groups.

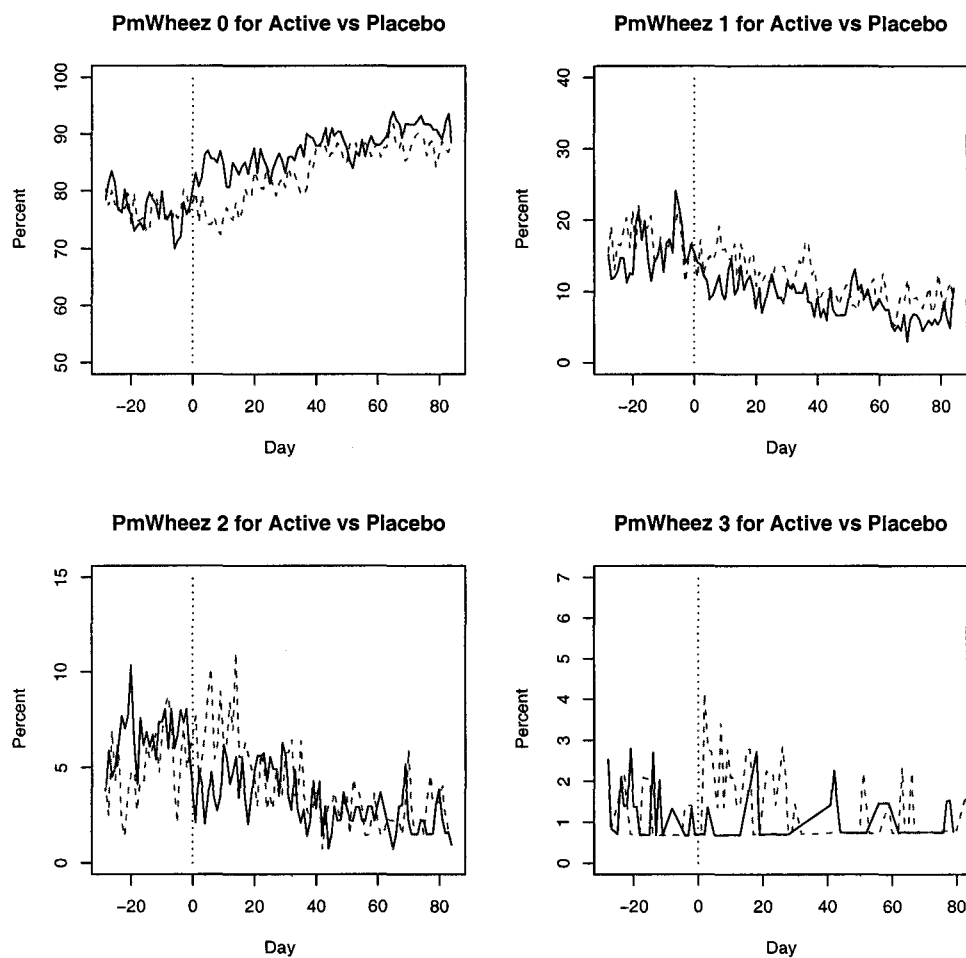


Figure 2.3: PMWHEEZ symptom by DAY and TRTMNT

The Figure 2.4 which plots PMCOUGH is similar to Figure 2.2, for AMCOUGH. We can also see here that for AMCOUGH=2 there is a difference in the two groups. If we compare the graphs for this symptom to the graphs for AMWHEEZ we can see by looking at the Y axis that this symptom is more present. Therefore, the facts

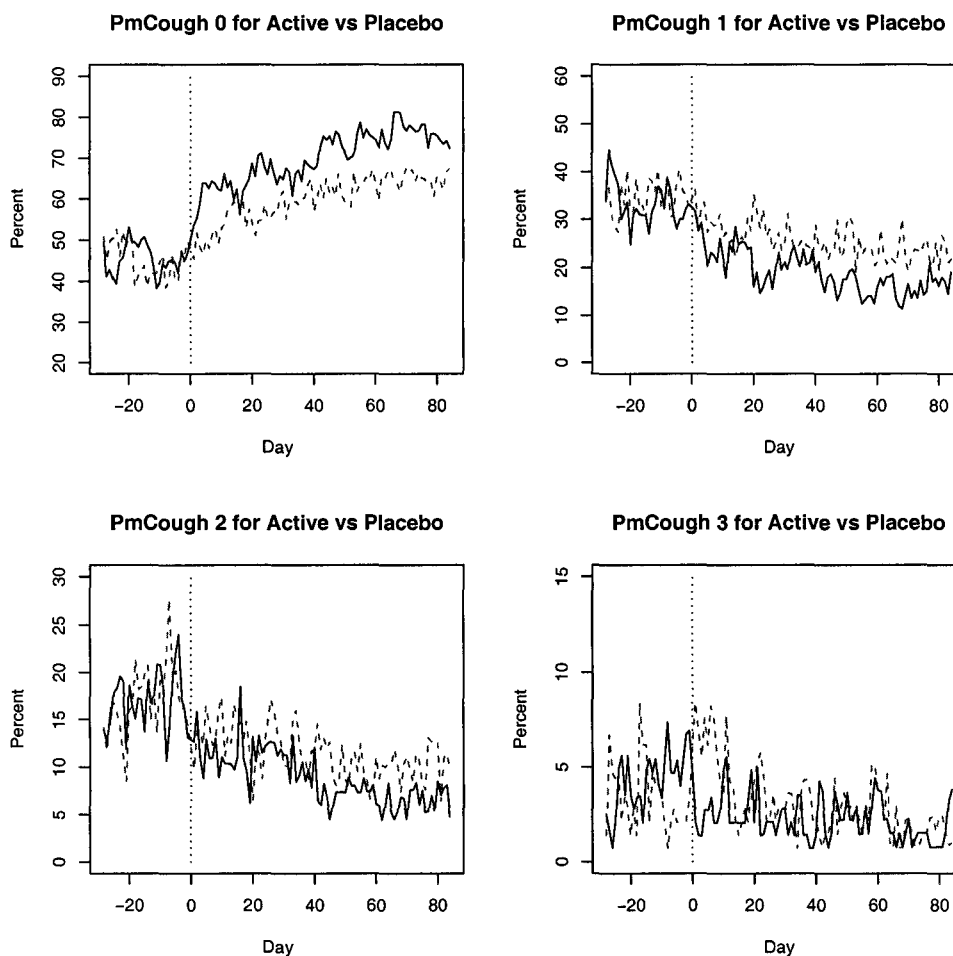


Figure 2.4: PMCOUGH symptom by DAY and TRTMNT

that the frequencies of mild and moderate symptoms decrease in time and that there is a difference between the active and the placebo groups, make us believe that the

treatment is beneficial. It is possible that the treatment is only relieving this symptom. This in fact may be the extent to which the treatment is superior to the Placebo.

As stated earlier, the patients were not followed for the same number of days. In the graphs plotting the 6 symptom variables, the percentages were used instead of frequencies for this very reason. The other two outcome variables, AMUSE and PMUSE recorded the number of occasions the patients used Ventolin, a rescue medication. Therefore their values are not categories as before, but counts and hence percentages cannot be obtained.

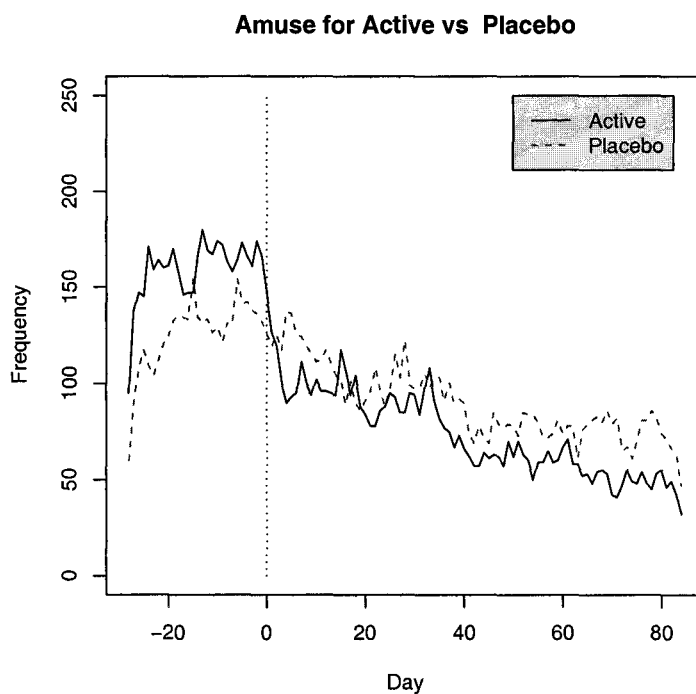


Figure 2.5: AMUSE by DAY and TRTMNT

Figure 2.5 plots the number of occasions Ventolin was used in each day against the variable DAY. To do that, the missing values had to be replaced by zeros because in

SAS arithmetic operations with missing values result in missing values. The percentage of missing values in AMUSE is 1.5% and in PMUSE is 1%.

Notice in Figure 2.5 that the frequencies patients used Ventolin prior to day 0 are not comparable in the two groups. It rather seems that the patients randomized to Placebo felt better prior to day 0 than the patients randomized to FP. However, as shown on the graph, the number of occasions Ventolin was used in the Active group drops after day 0 and it is less than the one in the Placebo group.

Figure 2.6 graph plots the frequency of night-time rescue medication use against DAY.

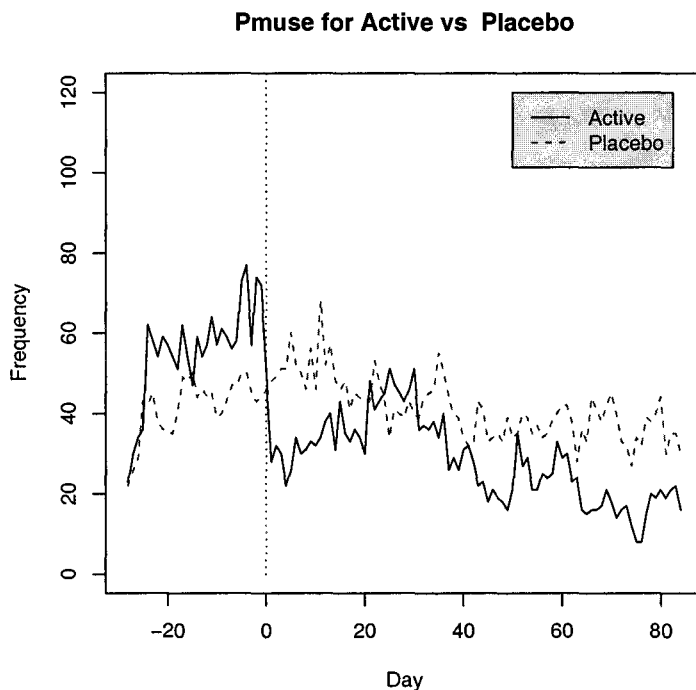


Figure 2.6: PMUSE by DAY and TRTMNT

Similar conclusions as for AMUSE can be drawn. Of notice is that the frequency

of Ventolin use is much lower during the night than during the day for both groups.

In conclusion the descriptive statistics and the associated graphs showed that the two groups are well balanced at the start of the trial and that the Active group seems to improve after intervention.

Chapter 3

Continuation Ratio Model

3.1 Modelling Approach

Ordinal responses are very commonly found in clinical trials. Most often, they are the result of evaluating the state of a patient. In our case, the six symptoms are measured on a scale from 0 to 3 where 0 is no symptom and 3 is the most severe symptom. The three commonly used models for ordinal responses are the log multiplicative, proportional odds and continuation ratio models. All three are generally applied to frequency data in contingency tables, and hence are not directly applicable to longitudinal data. However, continuation ratio model can be fitted by standard logistic regression techniques after restructuring the data.

The continuation ratio model makes a series of comparisons of all lower categories on a scale to the next succeeding one. Thus, the first category is compared to the second, the first two to the third, then the first three to the fourth and so on. Therefore this model describes the probability of moving one step further on the scale given the present situation (Lindsey, Jones and Ebbutt, 1997).

Suppose we have k ordered categories. The continuation ratio model is based upon

the conditional probabilities of being in category j among all subjects who are in category j or lower for a given patient (covariate) profile. That is,

$$\begin{aligned}\delta_j(\mathbf{x}) &= P(Y = j | Y \leq j, \mathbf{X} = \mathbf{x}) = \frac{P(Y = j | \mathbf{X} = \mathbf{x})}{P(Y \leq j | \mathbf{X} = \mathbf{x})} \\ &= \frac{\pi_j(\mathbf{x})}{\pi_0(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}, \quad j = 1, \dots, k.\end{aligned}\tag{3.1}$$

The δ_j 's are called *continuation ratios*. In this project, the backward formulation of the continuation ratio model is used. There is also a forward formulation of this model as:

$$\phi_j(\mathbf{x}) = P(Y = j | Y \geq j, \mathbf{X} = \mathbf{x}).\tag{3.2}$$

The use of forward continuation ratios makes sense if the response Y represents a discrete survival time. As high risk patients have short survival times, forward continuation ratios represent a comparison of high risk patients with low risk patients. Since in our case, Y represents a symptom status given by ordered categories with higher values belonging to more severe disease states, the opposite is true. The two models for the backward and forward continuation ratios are not equivalent and yield different results (Bender and Benner, 2002).

The continuation ratio models are obtained using a generalized linear model:

$$f\{\delta_j(x)\} = \alpha_j + \beta_j \mathbf{x}, \quad j = 1, \dots, k.\tag{3.3}$$

where $f(\cdot)$ is an appropriate link function.

However, the basic assumption in most applications is the “equal slopes” assumption, that is a homogenous coefficient for X is considered. The model in this case becomes:

$$f\{\delta_j(x)\} = \alpha_j + \beta \mathbf{x}, \quad j = 1, \dots, k.\tag{3.4}$$

If the logit link is used, the generalized linear model (3.4) becomes:

$$\log \left\{ \frac{\delta_j(x)}{1 - \delta_j(x)} \right\} = \alpha_j + \beta \mathbf{x} \Leftrightarrow \delta_j(x) = \frac{\exp(\alpha_j + \beta \mathbf{x})}{1 + \exp(\alpha_j + \beta \mathbf{x})}. \quad (3.5)$$

which is called continuation ratio model or proportional logit hazard model. If the more general model (3.3) is used or a mixture of (3.3) and (3.4) in which the equal slopes assumption is relaxed for some of the covariates, the resulting model is called "extended continuation ratio" (Bender and Benner, 2002).

3.2 Computational implementation

Even though most of the software packages provide binary logistic regression at least for the logit link, regression models for ordinal data are not so often implemented. However, it is possible to use any software for binary logistic regression to estimate the parameters of continuation ratio models. If the original data set is appropriately restructured by repeatedly including the corresponding data subset, the cut points and the dichotomous response at that cut-point, binary logistic regression can be applied to the restructured data.

Let now Y be the response variable with $k + 1$ ordered categories, $0, 1, \dots, k$ and the corresponding frequencies n_0, n_1, \dots, n_k . We will define two new variables called BR (binary response) and CP (cut-point). The first subset of data will be used to compare category k with all lower categories, that is, $\delta_k(x) = P(Y = k | Y \leq k)$. The BR will take the value 1 when $Y = k$ and 0 otherwise and at this step CP will equal k . It can be seen that this first subset contains all the original observations. In the second subset, we will compare category $k - 1$ with all lower categories. Now, BR will be 1 when $Y = k - 1$ and 0 when $Y \leq k - 2$, the category $Y = k$ being left out. Now, CP will equal $k - 1$. This subset will have $n_0 + n_1 + \dots + n_{k-1}$ observations. The last subset will only compare category 1 with category 0 and will have $n_0 + n_1$ observations.

Let's denote by N the total number of observations in the restructured data. Then, N will be:

$$N = k(n_0 + n_1) + (k - 1)n_2 + (k - 2)n_3 + \dots + n_k$$

Applying binary logistic regression to the restructured data set with BR as the response variable and $k - 1$ dummy variables representing the cut-point variable yields the regression coefficients of the continuation ratio model. The use of binary logistic regression to the restructured data has the advantage that it is possible to work with the more general models (3.3) without the usual but restrictive equal slopes assumption. By including interactions between the dummy variables describing the strata and the considered explanatory factor X , a model with different regression coefficients for each cut-point can be obtained.

Up until now the model has not taken into account the dependence among successive responses of an individual in repeated measurement situations. With the restructured data this dependencies can be accounted for by conditioning on the lagged value of the binary response, but interpretation is usually easier if the lagged ordinal response is used directly as an explanatory variable instead of recoding it.

3.3 Model Fitting to Asthma Data

The data set has 6 ordinal response variables: AMWHEEZ, AMCOUGH, AMBREATH, PMWHEEZ, PMCOUGH and PMBREATH. For each of these variables we apply the method described above in order to restructure the data. The implementation is shown only for variable AMWHEEZ. A macro in SAS (see Appendix 2) was used to restructure the original data set for each variable. As described, the new data set has two new variables, BR and CP. CP takes here three values, 1, 2 and 3. From this variable I created another two dummy variables called CP2 and CP3.

As an example, let's consider the model:

$$\text{logit}(\delta_{ijkl}) = \alpha + \beta_1 \text{CP2}_i + \beta_2 \text{CP3}_j + \beta_3 \text{TRTMNT}_k + \beta_4 \text{FREQBASE}_l \quad (3.6)$$

For $i, j, k, l \in \{0, 1\}$ we have:

$$i = 0 \text{ and } j = 0 \Rightarrow \text{CP}=1$$

$$i = 0 \text{ and } j = 1 \Rightarrow \text{CP}=3$$

$$i = 1 \text{ and } j = 0 \Rightarrow \text{CP}=2$$

So i and j refer to the cut-points, whereas k refers to treatment (active=1 and placebo=0) and l to frequency of occurrence (chronic=1, episodic=0). For instance,

$$\delta_{00..} = P(Y = 1 \mid Y \leq 1, \mathbf{X} = \mathbf{x})$$

$$\delta_{01..} = P(Y = 3 \mid Y \leq 3, \mathbf{X} = \mathbf{x})$$

$$\delta_{10..} = P(Y = 2 \mid Y \leq 2, \mathbf{X} = \mathbf{x})$$

To facilitate interpretation of coefficients, we write the model in matrix form as:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha & & & +\beta_4 \\ \alpha & & +\beta_3 & \\ \alpha & & +\beta_3 & +\beta_4 \\ \alpha & +\beta_2 & & \\ \alpha & +\beta_2 & & +\beta_4 \\ \alpha & +\beta_2 & +\beta_3 & \\ \alpha & +\beta_2 & +\beta_3 & +\beta_4 \\ \alpha & +\beta_1 & & \\ \alpha & +\beta_1 & & +\beta_4 \\ \alpha & +\beta_1 & +\beta_3 & \\ \alpha & +\beta_1 & +\beta_3 & +\beta_4 \end{bmatrix}$$

Interpretation of δ_{ijkl} can be made as follows. From the matrix display we can see that,

$$\text{logit}(\delta_{0000}) = \alpha \Leftrightarrow \delta_{0000} = \frac{e^\alpha}{1 + e^\alpha}.$$

Therefore, the probability of going up on the scale, from 0 to 1 for a patient who is receiving a placebo and that has episodic symptoms will be $e^\alpha/(1 + e^\alpha)$. In order to estimate the same probability but for moving from 0 or 1 to 2 we need CP=2 so we calculate δ_{1000} :

$$\text{logit}(\delta_{1000}) = \alpha + \beta_1 \Leftrightarrow \delta_{1000} = \frac{e^{\alpha+\beta_1}}{1 + e^{\alpha+\beta_1}}.$$

The above probabilities are interesting to look at but the main purpose is to compare the Active and Placebo groups. Therefore we first compute the probability of moving from 0 to 1 for a patient who is receiving the Active treatment and has episodic symptoms. This probability is given by δ_{0010} :

$$\text{logit}(\delta_{0010}) = \alpha + \beta_3 \Leftrightarrow \delta_{0010} = \frac{e^{\alpha+\beta_3}}{1 + e^{\alpha+\beta_3}}.$$

Now we are able to compare the Placebo and Treatment groups by comparing δ_{0000} to δ_{0010} . However, instead of comparing the probabilities it is more usual to compare the odds by computing the odds ratio. In this case, the odds of δ_{0000} is equal to e^α whereas the odds of δ_{0010} is $e^{\alpha+\beta_3}$. Therefore, the odds ratio of Active vs Placebo will be:

$$\text{odds ratio of Active vs Placebo} = \frac{e^{\alpha+\beta_3}}{e^\alpha} = e^{\beta_3}. \quad (3.7)$$

For this reason the parameters are also called “log odds ratios”. This odds ratio will not change if the patient instead of having episodic symptoms has chronic ones. Nor will it change if we consider odds of moving from 0,1 to 2 because of the “equal slopes assumption”. This assumption will be relaxed later by incorporating interaction terms for the cut-points and treatment.

The results of fitting the logistic procedure on the restructured data set are reported in Tables 3.1-3.4. First let's take a look at the Goodness of Fit statistics reported in Table 3.1:

Table 3.1: Goodness of Fit statistics

Criterion	DF	Value	Value/DF	Pr > ChiSq
Deviance	7	123.6355	17.6622	<.0001
Pearson	7	129.9370	18.5624	< .0001

Since the p-values are less than 0.05 there is evidence that the logistic model is not a good fit for the data. The fact that the Value/DF is greater than 1 suggests overdispersion. To account for the problem of overdispersion, a scale parameter was included in the analysis. When this was done, the test was no longer valid so the Homer and Lemeshow test was used but it still suggested lack of fit. When lack of fit occurs, the standard errors of the parameters are underestimated so they should not be trusted. The lack of fit may be due to violation of one or more of the assumptions (Bender and Benner, 2002). For instance we know for sure that the assumption of independence is not valid. In the next subsection we will take a look at some simulation results trying to explore in some detail this issue.

Table 3.2 shows the maximum likelihood estimates of the parameters. Notice that all of them are statistically significant.

We can now compute the estimated odds ratios for Active vs Placebo and also for Episodic vs Chronic:

$$\text{odds ratio of Active vs Placebo} = e^{\hat{\beta}_3} = e^{-0.5033} = 0.604532$$

$$\text{odds ratio of Chronic vs Episodic} = e^{\hat{\beta}_4} = e^{0.8533} = 2.34738$$

Since the odds ratio of Active vs Placebo is less than one it means that the odds and hence the probability of getting worse is higher in the Placebo group than in the

Table 3.2: ML estimates for AMWHEEZ model (3.6)

Parameter	DF	Estimate	St. Error	Wald ChiSq	Pr > ChiSq
Intercept (α)	1	-2.1122	0.0347	3701.0130	< .0001
CP2 (β_1)	1	-1.2535	0.0370	1146.3282	< .0001
CP3 (β_2)	1	-3.2073	0.0815	1547.0118	< .0001
TRTMNT (β_3)	1	-0.5033	0.0330	232.4881	< .0001
FREQBASE (β_4)	1	0.8533	0.0364	550.3063	< .0001

Active. As expected, the patients who experience Chronic symptoms are more likely to get worse than those who experience episodic symptoms.

Table 3.3 summarizes the goodness of fit results for the other variables:

Table 3.3: Goodness of Fit statistics

Variable	Criterion	DF	Value	Value/DF	Pr > ChiSq
AMBREATH	Deviance	7	38.8940	5.5563	< .0001
	Pearson	7	37.6835	5.3834	< .0001
AMCOUGH	Deviance	7	39.7612	5.6802	< .0001
	Pearson	7	40.9420	5.8489	< .0001
PMWHEEZ	Deviance	7	92.4901	13.2129	< .0001
	Pearson	7	91.3819	13.0546	< .0001
PMBREATH	Deviance	7	54.7258	7.8180	< .0001
	Pearson	7	50.4316	7.2045	< .0001
PMCOUGH	Deviance	7	59.3044	8.4721	< .0001
	Pearson	7	62.2214	8.8888	< .0001

The results suggest that the model does not provide a good fit for any of the response variables. However, except for AMWHEEZ and PMWHEEZ, the Values/DF

are less than 10.

Table 3.4 shows the parameters for TRTMNT and FREQBASE and the computed odds ratios for each symptom. The parameters for CP2 and CP3 and for the two variables mentioned above were found to be significant for all symptoms.

Table 3.4: Parameters and odds ratios for model (3.6)

Variable	Parameter	Estimate	Odds Ratio
AMBREATH	TRTMNT	-0.2700	0.763
	FREQBASE	0.5834	1.792
AMCOUGH	TRTMNT	-0.3214	0.725
	FREQBASE	0.5558	1.743
PMWHEEZ	TRTMNT	-0.3459	0.708
	FREQBASE	0.7561	2.130
PMBREATH	TRTMNT	-0.2203	0.802
	FREQBASE	0.6511	1.918
PMCOUGH	TRTMNT	-0.4101	0.664
	FREQBASE	0.4785	1.614

Looking at the odds ratios for TRTMNT we can observe that they are all less than one. This suggests that the treatment has a positive effect compared to the Placebo. As noticed for AMWHEEZ, the odds ratios of the other variables for FREQBASE are greater than one indicating that patients with Chronic symptoms have a higher probability of moving up on the symptom scale.

Up until now we have examined a simple model that assumes independent observations and equal slopes. Next we will introduce another model that includes interaction between the cut-points CP2, CP3 and TRTMNT. If the coefficients of the interactions are significant, we will have different odds ratios for each cut-point.

The model can be written as:

$$\text{logit}(\delta_{ijkl}) = \alpha + \beta_1 \text{CP2} + \beta_2 \text{CP3} + \beta_3 \text{TRT} + \beta_1^* \text{CP2}^* \text{TRT} + \beta_2^* \text{CP3}^* \text{TRT} + \beta_4 \text{FREQ} \quad (3.8)$$

The model in matrix form is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_2^* \\ \beta_1^* \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha & & & & & & +\beta_4 \\ \alpha & & +\beta_3 & & & & \\ \alpha & & +\beta_3 & & & & +\beta_4 \\ \alpha & +\beta_2 & & & & & \\ \alpha & +\beta_2 & & & & & +\beta_4 \\ \alpha & +\beta_2 & +\beta_3 & & +\beta_2^* & & \\ \alpha & +\beta_2 & +\beta_3 & & +\beta_2^* & +\beta_4 & \\ \alpha & +\beta_1 & & & & & \\ \alpha & +\beta_1 & & & & & +\beta_4 \\ \alpha & +\beta_1 & & +\beta_3 & +\beta_1^* & & \\ \alpha & +\beta_1 & & +\beta_3 & +\beta_1^* & & +\beta_4 \end{bmatrix}$$

The odds ratios of moving up on the scale for the variable TRTMNT depend on the cut-points. The odds ratios are given by:

$$\text{OR}(0 \rightarrow 1) = e^{\beta_3}$$

$$\text{OR}(0, 1 \rightarrow 2) = e^{\beta_3 + \beta_1^*}$$

$$\text{OR}(0, 1, 2 \rightarrow 3) = e^{\beta_3 + \beta_2^*}$$

Table 3.5: Goodness of fit statistics for models (3.6) and (3.8)

Model	Criterion	DF	Value	Value/DF	Pr> ChiSq
Main effects	Deviance	7	123.6455	17.6622	<.0001
	Pearson	7	129.9370	18.5624	<.0001
Interaction	Deviance	6	90.1709	15.0285	<.0001
	Pearson	6	94.6914	15.7819	<.0001

However, if for instance the interaction between CP3 and TRTMNT is not significant then $OR(0,1,2 \rightarrow 3)$ will only be equal to e^{β_3} .

To obtain the estimates, the macro used for model (3.6) was modified to include in the logistic procedure the interactions. Backward logistic regression was used. First, we will look at the Pearson goodness of fit test. In the table 3.5 we can compare these statistics obtained for AMWHEEZ in the two models.

We can see that when the interaction terms are included the value of the Pearson statistic is smaller suggesting a better fit. However, the value is still large and hence the test rejects the null hypothesis that the model is adequate.

In assessing the model the AIC (Akaike Information Criterion) score is also useful. Its purpose is to penalize a model that has too many parameters. AIC statistic is widely used to select the best model among alternative parametric models. The model with the smallest AIC is considered to be the best.

$$AIC = -2\text{Log}L + 2(k + s)$$

where k is the total number of response levels minus one, and s is the number of explanatory effects. The AIC score for the main affects, model (3.6) is 27824.837 which is a little bigger than the one for model (3.8) that includes interaction terms, which is 27793.372.

We conclude that model (3.8) is a little better than 3.6. Next we will look at the

parameter estimates and the odds ratios.

Table 3.6 shows the parameter estimates $\beta_3, \beta_1^*, \beta_2^*, \beta_4$ for all the symptoms.

Table 3.6: Parameter estimates for model 3.8

Symptom	Parameter	Estimate	Pr>ChiSq
AMWHEEZ	FREQBASE (β_4)	0.8560	<.0001
	TRTMNT (β_3)	-0.6192	<.0001
	CP2*TRTMNT (β_1^*)	0.4285	<.0001
AMBREATH	FREQBASE (β_4)	0.5832	<.0001
	TRTMNT (β_3)	-0.2529	<.0001
	CP3*TRTMNT (β_2^*)	-0.4257	0.0207
AMCOUGH	FREQBASE (β_4)	0.5564	<.0001
	TRTMNT (β_3)	-0.3696	<.0001
	CP2*TRTMNT (β_1^*)	0.1458	0.0207
PMWHEEZ	FREQBASE (β_4)	0.7557	<.0001
	TRTMNT (β_3)	-0.3205	<.0001
	CP3*TRTMNT (β_2^*)	-0.5133	0.0033
PMBREATH	FREQBASE (β_4)	0.6504	<.0001
	TRTMNT (β_3)	-0.3205	<.0001
	CP2*TRTMNT (β_1^*)	-0.2211	0.0255
	CP3*TRTMNT (β_2^*)	-0.9463	<.0001
PMCOUGH	FREQBASE (β_4)	0.4788	<.0001
	TRTMNT (β_3)	-0.4552	<.0001
	CP2*TRTMNT (β_1^*)	0.1392	0.0091

It can be seen that all β_4 coefficients are strictly positive. This implies that the corresponding odds ratios will be greater than 1 which means that patients with Chronic symptoms are more likely to have higher scores than patients with Episodic symptoms.

This is consistent with our beliefs and agrees with the previous model. We can also see that all β_3 coefficients are negative. This means that the treatment performs better than the placebo in lowering the symptoms.

All the estimated odds ratios are presented in Table 3.7. As anticipated, the numbers in the FREQBASE odds ratio column are all greater than one whereas the odds ratios corresponding to TRTMNT are less than one. Notice that for AMWHEEZ, AMCOUGH and PMCOUGH, $OR(0 \rightarrow 1) = OR(0,1,2 \rightarrow 3)$. This is so because for these symptoms the interaction between CP3 and TRTMNT was not significant and therefore excluded from the final model.

Table 3.7: Odds ratios for model 3.8

Symptom	FREQBASE	TRTMNT		
		OR(0 \rightarrow 1)	OR(0,1 \rightarrow 2)	OR(0,1,2 \rightarrow 3)
AMWHEEZ	2.3537	0.5383	0.8263	0.5383
AMBREATH	1.7917	0.7765	0.7765	0.5073
AMCOUGH	1.7443	0.6910	0.7994	0.6910
PMWHEEZ	2.1291	0.7257	0.7257	0.4343
PMBREATH	1.9163	0.8809	0.7061	0.3419
PMCOUGH	1.6141	0.6343	0.7290	0.6343

So far the two models that we looked at assumed that the observations are independent. However, around 100 observations were recorded in consecutive days for each patient. The observations are therefore likely to be correlated within patients but independent between patients. To take into the account this dependence one can condition on the lagged value of the binary response. The interpretation however is usually easier if the lagged ordinal response is used directly as an explanatory variable (Muenz and Rubinstein, 1985 and Lindsey, 1997). By lagging the variable of interest the last observation of patient i becomes the first observation of patient $i + 1$ and

hence it should be deleted. There are two approaches to this. As stated before, for this model we are only interested in the observations recorded after day -1. If we only keep the observations recorded after day -1 and then lag the variable, we would have to delete for each patient the first observation after this day. A better approach is to first lag the variable and then delete the observations recorded prior to day 1. In this way we will still have an observation for day 1 which will come from day -1 (there is no day 0). The second approach is used.

The model is:

$$\text{logit}(\delta_{ijklm}) = \alpha + \beta_1 \text{CP2}_i + \beta_2 \text{CP3}_j + \beta_3 \text{TRTMNT}_k + \beta_4 \text{FREQBASE}_l + \beta_5 \text{LAGG}_m \quad (3.9)$$

As we did for the other models, we will first look at the goodness of fit statistics for variable AMWHEEZ in Table 3.8.

Table 3.8: Goodness of fit for model 3.6 with a lagged covariate

Criterion	DF	Value	Value/DF	Pr> ChiSq
Deviance	42	799.6658	19.0397	<.0001
Pearson	42	4664.0656	111.0492	<.0001

Looking at the Value/DF we can see that there is a big difference between these values for the two criteria. This indicates that the test statistics should not be trusted. The AIC score for this model is 16520.208 which is smaller than in the previous models. By introducing the lagged variable, called LAG, as a covariate, the number of response profiles augmented from 12 to 48. This may be the cause for the difference above.

To obtain the parameter estimates the same macro with a small modification was used. In order to create the new variable the function “lagg” was applied to the symptom variable. It was afterwards included in the model in *proc logistic*. This variable was treated as continuous. The interpretation of the coefficient is that the

odds ratio will increase by a factor of e^{β_5} for an increase in LAGG of 1 unit. The parameter estimates along with the odds ratio estimates are presented in Table 3.9.

Table 3.9: Parameters and odds ratios for model (3.6) with a lagged covariate

Variable	Parameter	Estimate	Odds Ratio
AMWHEEZ	TRTMNT	-0.2933	0.746
	FREQBASE	0.4181	1.519
	LAGG	3.1016	22.232
AMBREATH	TRTMNT	-0.1546	0.857
	FREQBASE	0.3318	1.393
	LAGG	3.0685	21.509
AMCOUGH	TRTMNT	-0.1766	0.838
	FREQBASE	0.2602	1.297
	LAGG	2.6325	13.909
PMWHEEZ	TRTMNT	-0.1613	0.851
	FREQBASE	0.4073	1.503
	LAGG	3.1916	24.327
PMBREATH	TRTMNT	-0.1108	0.895
	FREQBASE	0.4227	1.526
	LAGG	2.8854	17.911
PMCOUGH	TRTMNT	-0.2334	0.792
	FREQBASE	0.2437	1.276
	LAGG	2.4103	11.138

Again the odds ratios of Active vs Placebo are less than one suggesting that the patients receiving the treatment are doing better than those receiving the placebo. The odds ratios corresponding to the variable LAGG are very large. They tell us that going from i to $i + 1$ increases the odds of moving one step on the scale.

3.4 Simulations for Assessing Goodness of Fit

3.4.1 General Testing Considerations

In broad terms, goodness of fit tests may fail to validate the logistic regression as a viable model for a data set for the following reasons:

1. the trials (observations) are not independent;
2. the link function does not accurately capture the relationship between covariates and the response's probability of "success";
3. the sample size is too large that minor model departures are detected as being statistically significant even though they are not scientifically relevant; and
4. data sparsity invalidates the approximate (in many cases asymptotic) null distribution of the test statistic.

It was noted before that the longitudinal nature of the asthma experiment would likely assure the presence of (1). When the groups defined by quantitative covariates are large, informal plots of link-transformed observed response proportions against covariate values may shed some light on the possible deficiencies noted in (2). It is apparent in the following subsections that (3) is likely to be the reason since a large number of trials is involved in the asthma study. The restructuring of the data makes the number of observations to be even larger.

A detailed exploration of the goodness of fit statistics is presented in this section. Particular features of the data being analyzed in conjunction with simulations are used to get insights into the goodness of fit of the logistic model in the present problem.

3.4.2 Goodness of fit

The results of the Goodness of fit test in all three models presented in the previous sections suggest that the logistic model is not adequate. In this section we examine in detail the goodness of fit statistics applied. Once we had the model we needed to assess how well it fits the data, or how close the model-predicted values are to the corresponding observed values. Test statistics that assess fit in this manner are known as goodness of fit statistics. Following standard practice, the test statistics used here are based on comparing in some appropriate manner observed and expected counts under the model. Departures of the predicted proportions from the observed proportions should be essentially random. The test statistics have approximate chi-square distributions when the number of observations in each group is sufficiently large. For ease, assume that we have only two covariates with indexes h and i both with two levels and that the response variable is dichotomous, with index j . Then, the Pearson chi-square, Q_P , and the likelihood ratio chi-square or deviance, Q_L , are given by:

$$Q_P = \sum_{h=1}^2 \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{hij} - m_{hij})^2}{m_{hij}}$$

$$Q_L = \sum_{h=1}^2 \sum_{i=1}^2 \sum_{j=1}^2 2n_{hij} \log\left(\frac{n_{hij}}{m_{hij}}\right)$$

where the m_{hij} 's are the model predicted counts.

If the model fits, both Q_P and Q_L are approximately distributed as chi-square with degrees of freedom equal to the number of profiles minus the number of parameters. The sample size guidelines for these statistics to be approximately chi-square are:

- each of the groups has at least 10 subjects ($n_{hi+} \geq 10$)
- 80% of the predicted counts (m_{hij}) are at least 5
- all other expected counts are greater than 2, with essentially no 0 counts

Sufficient replication within sub-populations is required to make the Pearson and deviance goodness-of-fit tests valid. When one or more continuous predictors are in the model, the data are often too sparse to use these statistics. Hosmer and Lemeshow (1989) proposed a statistic that they show, through simulation, is distributed as chi-square when there is no replication in any of the sub-populations. The “lackfit” option in the model statement in SAS performs the Hosmer and Lemeshow goodness-of-fit test. The subjects are divided into approximately ten groups of roughly the same size based on the percentiles of the estimated probabilities. The discrepancies between the observed and expected number of observations in these groups are summarized by the Pearson chi-square statistic, which is then compared to a chi-square distribution with t degrees of freedom, where t is the number of groups minus n . Performing this test for AMWHEEZ using model (3.6) yields the results of Table 3.10:

The chi-square test statistic is 129.9370 with 10 degrees of freedom. Since the p -value is less than 0.0001 we reject the adequacy of the logistic model.

3.4.3 Simulations

Simulations will be used to explore various aspects of goodness of fit. First, let’s examine the data. The symptoms are rated on a scale from 0 to 3, thus these variables are multinomially distributed. By restructuring the data, the variable BR takes only two values, 0 and 1 where 1 can be considered as a “success” and 0 as a failure. Table 2.4 showed the frequencies of the symptom scores and we noticed that there were a lot of zeros. Table 3.11 shows the frequencies and percents of 0, 1, 2 and 3 for AMWHEEZ by treatment, after day -1. There were 181 missing observations in the Treatment group and 207 in the Placebo. By looking at the percents we may say that patients receiving the treatment are doing better.

Notice the high percentage of observations of 0 and 1 in both groups. After the

Table 3.10: Partion for the Hosmer and Lemeshow test

Group	Total	BR=1		BR=0	
		Observed	Expected	Observed	Expected
1	4647	27	13.71	4620	4633.29
2	5048	15	24.59	5033	5023.41
3	7010	38	48.36	6972	6961.64
4	6545	81	74.35	6464	6470.65
5	4620	157	94.49	4463	4525.51
6	5033	112	168.02	4921	4864.98
7	6972	341	325.74	6631	6646.26
8	4463	346	304.14	4117	4158.86
9	6464	463	484.75	6001	5979.25
10	4921	479	531.06	4442	4389.94
11	6631	849	971.57	5782	5659.43
12	6001	1460	1327.23	4541	4673.77

Table 3.11: AMWHEEZ after day -1

Score symptom	Active		Placebo	
	Frequency	Percentage	Frequency	Percentage
0	9718	84.68	8776	77.08
1	1195	10.41	1939	17.03
2	498	4.34	575	5.05
3	65	0.57	96	0.84

data are restructured for AMWHEEZ, 6% of the observations are 1's and 97% are 0's, the probability of "success" being very low. This may be a reason for which the tests for goodness of fit show that the logistic model is inadequate.

Since all the symptoms have a similar distribution of scores, another “symptom” variable called “Y” was generated that has the score probabilities in the two groups given in Table 3.12.

Table 3.12: Simulation probabilities TRTMNT significant

Symptom scores	Probabilities	
	Active	Placebo
0	0.25	0.15
1	0.50	0.45
2	0.20	0.30
3	0.05	0.10

The probabilities were chosen so that there would be a difference between Placebo and Active groups. Thus, the generation of the random numbers took into account the two interventions but not the severity (episodic or chronic) of the symptoms. That is why we would expect the variable TRTMNT to be significant and FREQBASE not to be significant.

A macro in SAS was created to generate the variable Y, restructure the data and apply the logistic procedure, n times (See Appendix 2). The macro calculated the goodness of fit statistics and the parameters for each of the n data sets, creating at each step two new data sets containing the results of interest. In the end, after the $2n$ data sets had been created, they were merged into two data set. So in the end, two data sets were created, one containing the n goodness of fit tests and the other the n parameter estimates results. Some of the default variables created were renamed. A total of $n = 100$ data sets were generated. We now want to see if the tests show us whether the logistic model is a good fit for the data. Since we have 100 test results on 100 different variables we need to summarize the data obtained.

Table 3.13 shows the mean, standard deviation, minimum and maximum for the

Value/DF and for the p -value. Remember that for a good fit we need the Value/DF to be very close to 1 and the p -value to be larger than 0.05.

Table 3.13: Simulation goodness of fit - TRTMMNT significant

Criterion	Variable	Mean	Std Dev	Min	Max
Deviance	Value/DF	6.0208	1.8285	2.5342	11.1351
	Pr > Chi-Square	0.00038	0.0017	3.611E-14	0.0132
Pearson	Value/DF	6.0119	1.8268	2.5208	11.0907
	Pr > Chi-Square	0.00040	0.0018	4.176E-14	0.01367

Looking at the values in the table we can conclude that the logistic model is not adequate for the data. The mean Value/DF for both Deviance and Pearson is about 6, and the minimum 2.5. The largest p -values are 0.01 which correspond to the smallest Value/DF.

We will now look at the parameter estimates. First, we want to see if all are significant for each step of the simulation, that is, for each data set previously created. Table 3.14 summarizes the results for TRTMNT and FREQBASE. The intercept, CP2 and CP3 were all found to be significant.

Table 3.14: Simulation parameter estimates - TRTMMNT significant

Variable	Parameter/P-value	Mean	Std Dev	Min	Max
TRTMNT	Estimate	-0.5706	0.0205	-0.6192	-0.4890
	Pr > Chi-Square	3.85E-119	3.85E-118	8.33E-182	3.85E-117
FREQBASE	Estimate	0.0051	0.0222	-0.0341	0.0579
	Pr > Chi-Square	0.4712	0.2898	0.0065	0.9990

We can see that the variable TRTMNT was significant in all 100 replications since both the minimum and maximum p -value are less than 0.05. The estimates are all less

than zero indicating that the treatment has a better effect than the placebo.

As expected, FREQBASE was not significant in the majority of cases. In 93 out of 100 cases the p-value was larger than 0.05, the mean p-value being 0.47.

So even if a better distribution of the score symptoms is assured, the goodness of fit test suggests that the logistic model is not adequate. The only improvement is that the value/DF is closer to 1 than before. However, the logistic method employed still shows the TRTMNT to be significant and recognizes the FREQBASE not to be significant.

Let's now simulate a symptom variable whose scores are no different for the placebo and active groups. So instead of generating the scores with different probabilities, the scores for Placebo are generated the same way as the scores for Active. The goodness of fit results are in Table 3.15. Most of the values are concentrated between 0.4 and 1.2. It seems that now, there is underdispersion.

Table 3.15: Simulation goodness of fit - none significant

Criterion	Variable	Mean	Std Dev	Min	Max
Deviance	Value/DF	1.0473	0.5937	0.0951	2.8259
	Pr > Chi-Square	0.4871	0.3042	0.0060	0.9985
Pearson	Value/DF	1.0467	0.5933	0.0951	2.8199
	Pr > Chi-Square	0.4874	0.3043	0.0061	0.9985

Table 3.16 below shows the parameter estimates:

We can see that neither TRTMNT nor FREQBASE are significant.

Let's now try to make both TRTMNT and FREQBASE significant. To do this, the macro was modified so that it generates 0, 1, 2, 3 with different probabilities in each of the four groups that are created with the two variables. The probabilities used

Table 3.16: Simulation parameter estimates - none significant

Variable	Parameter/P-value	Mean	Std Dev	Min	Max
TRTMNT	Estimate	0.0008	0.0231	-0.0667	0.0537
	Pr > Chi-Square	0.5058	0.3097	0.0019	0.9973
FREQBASE	Estimate	0.0011	0.0209	-0.0497	0.0448
	Pr > Chi-Square	0.5061	0.2877	0.0229	0.9962

are in Table 3.17 and the goodness of fit results in Table 3.18.

Table 3.17: Probabilities of 0,1,2,3 - both significant

	Score	Active	Placebo
Episodic	0	0.30	0.20
	1	0.50	0.30
	2	0.15	0.35
	3	0.05	0.15
Chronic	0	0.25	0.15
	1	0.40	0.25
	2	0.25	0.45
	3	0.10	0.15

In this setting the results of the goodness of fit tests are in Table 3.18.

The value over degrees of freedom in this case is very high indicating lack of fit.

The parameter estimates from the simulated data appear in Table 3.19. Both variables are now significant as all p-values are very small. However the logistic model seems to be inadequate.

To further investigate the reasons of lack of fit we will now reduce the sample size. We must remember that the logistic regression is originally run on around 68,000

Table 3.18: Simulation goodness of fit - both significant

Criterion	Variable	Mean	Std Dev	Min	Max
Deviance	Value/DF	115.2341	7.2682	97.7686	135.0950
	Pr > Chi-Square	<0.0001	<0.0001	<0.0001	<0.0001
Pearson	Value/DF	115.0906	7.2150	97.5704	134.8313
	Pr > Chi-Square	<0.0001	<0.0001	<0.0001	<0.0001

Table 3.19: Simulation parameter estimates -both significant

Variable	Parameter/P-value	Mean	Std Dev	Min	Max
TRTMNT	Estimate	-0.6867	0.0217	-0.7409	-0.6337
	Pr > Chi-Square	< 0.0001	< 0.0001	< 0.0001	< 0.0001
FREQBASE	Estimate	0.3168	0.0224	0.2665	0.3859
	Pr > Chi-Square	< 0.0001	< 0.0001	< 0.0001	< 0.0001

observations. This large number of observations is due to the longitudinal aspect and to the restructuring of the original data set. To reduce the sample size we will now only use the symptom scores recorded on days that are multiple of 5. That is, we select the records that have the variable DAY equal to $5k$, where $k = 1, 2, 3, \dots$. The same data restructuring algorithm is applied to new data set. The restructured data set has around 13,000 observations. For reasons of comparison 3 other data sets are created by selecting the records that have variable DAY equal to $5k + 1$, $5k + 2$ and $5k + 3$. The results after running the logistic procedure for AMWHEEZ for all data sets are in Table 3.20 that shows the Values/DF for the Pearson statistic, the parameter estimates of TRTMNT and FREQBASE and their odds ratios. All covariates were significant.

The model does not adequately fit the data as the Value/DF are much greater than 1 and the p -values are smaller than 0.05. However, we can notice that with a reduced

Table 3.20: Reduced sample size results

DAY	Pearson Value/DF	Covariate	Parameter estimates	Odds ratios
5k	4.7833	TRTMNT	-0.5351	0.5856
		FREQBASE	0.8373	2.3101
5k + 1	5.1101	TRTMNT	-0.4797	0.6189
		FREQBASE	0.9250	2.5218
5k + 2	3.6049	TRTMNT	-0.4721	0.6236
		FREQBASE	0.8650	2.3750
5k + 3	4.0278	TRTMNT	-0.4255	0.6534
		FREQBASE	0.8363	2.3078
10k	3.6420	TRTMNT	-0.5106	0.6001
		FREQBASE	0.8966	2.4512
ALL	18.5624	TRTMNT	-0.5033	0.6050
		FREQBASE	0.8533	2.3470

but still large sample size the goodness of fit improves.

Next, a data set with 300 observations was generated. The data set has 3 variables: TRTMNT, FREQBASE, and SYMPTOM. We can look at this data set as a copy of the original one only that here each patient has only one observation. The values for SYMPTOM were generated with the probabilities found in Table 3.17. Running the same macro as before, we obtain that 65% of the times the goodness of fit test shows that the model is adequate. The mean Value/DF in the 100 results generated is 2.56. As for the parameter estimates, TRTMNT is significant 97% of the times whereas FREQBASE only 39%.

Making the sample size smaller may improve the results of the goodness of fit test but it is not to be recommended. A smaller sample size is unable to detect small differences in the two groups.

The simulations showed us that even though there may be evidence that it is not adequate, the logistic model seems to perform well showing a significant variable when indeed there is one.

The fact that Value is much larger than DF, making the ratio bigger than 1 can be the result of:

- missing covariates and/or interaction terms
- presence of non-linear effects
- wrong link function
- existence of large outliers

Adding more covariates such as gender and country, and also interaction terms seemed to increase the goodness of fit, Value/DF decreasing from 17 to 12. However it is still large. The model assumes logistic regression and there seems to be no hint or reason to use a link function other than *logit*. We can also exclude the existence of large outliers.

Overdispersion can be explained by:

- variation among the success probabilities
- correlation between the binary responses

Since we know for a fact that the data are correlated, this perhaps explains the presence of overdispersion. The logistic model is:

$$Y_i \sim \text{Bin}(n_i, p_i), \text{ independent}$$

$$p_i = e^{x_i^t \beta} / (1 + e^{x_i^t \beta})$$

$$E(Y_i) = n_i p_i$$

$$\text{Var}(Y_i) = n_i p_i (1 - p_i)$$

Now, in the case of correlation we have:

$$\begin{aligned}
Y_i &= \sum_{j=1}^{n_i} R_{ij}, \text{ where } R_{ij} \text{ equals 1 for success and 0 for failure} \\
P(R_{ij} = 1) &= p_i \\
\Rightarrow \text{Var}(Y_i) &= \sum_{j=1}^{n_i} \text{Var}(R_{ij}) + \sum_{j=1}^{n_i} \sum_{k \neq j}^{n_i} \text{Cov}(R_{ij}, R_{ik}) \\
&= \sum_{j=1}^{n_i} p_i(1 - p_i) + A, \quad A \neq 0 \\
&= n_i p_i(1 - p_i) + A
\end{aligned}$$

So we can see that Y_i does not have a binomial distribution which is a critical assumption in the logistic regression model.

Chapter 4

Generalized Estimating Equations

Correlated data come from many sources such as longitudinal studies on health care outcomes, crossover studies concerned with drug comparisons or clinical trials investigating new treatments with baseline and follow-up visits. An important consideration in each of these situations is how to account for the correlated measurements in the analysis. Within subject factors are likely to have correlated measurements, while between-subject factors like age, gender, are likely to have independent measurements. Unless the correlation is taken into account, parameter estimates and standard errors will not be correct.

The generalized estimating equations (GEE) approach, first introduced by Liang and Zeger (1986), is an extension of generalized linear models that provides a semi-parametric approach to longitudinal data analysis. The GEE methodology models a known function of the marginal expectation of the dependent variable as a linear function of one or more explanatory variables. With quasi-likelihood methods, statistical models are created by making assumptions about the link function and the relationship between the first two moments, but without fully specifying the complete distribution of the response. GEEs describe the random component of the model for

each marginal response with a common link and variance function.

The GEE methodology provides consistent estimators of the regression coefficients and their variances under weak assumptions about the actual correlation among a subject's observations. This approach avoids the need for multivariate distributions by assuming only a functional form for the marginal distribution at each time point or condition. The covariance structure across time or conditions is managed as a nuisance parameter. The method relies on the independence across subjects to estimate consistently the variance of the proposed estimators even when the assumed working correlation structure is incorrect. Since GEEs are an extension of generalized linear models (GLM), GLM will be first shortly introduced (Stokes, Davis and Koch, 2000).

4.1 Generalized Linear Models

Generalized linear models are a broad class of regression models suitable for analyzing diverse types of univariate responses (continuous, binary, counts). It has a three part specification:

1. a distributional assumption
2. a systematic component
3. a link function

Distributional assumption

Generalized linear models extend many of the basic concepts and ideas of standard linear regression analysis to settings where the response variable is discrete and can no longer be assumed to have a normal distribution. They assume that the response

variable has a probability distribution belonging to the exponential family of distributions, that includes the normal, binomial, poisson, gamma and others. The probability distribution function of this family can be expressed as follows:

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}. \quad (4.1)$$

With this notation, θ is called location parameter, $a(\phi)$ is a scale parameter whereas $c(y, \phi)$ is a normalizing term. It is important to notice that:

$$\begin{aligned} E(y) &= b'(\theta) = \mu, \\ V(y) &= b''(\theta)a(\phi). \end{aligned}$$

We can see that the variance is a function of the expected value of the distribution and a function of the possibly unknown scale parameter $a(\phi)$.

Systematic component

The systematic component of the generalized linear model specifies that the effects of the covariates, \mathbf{x}_i , on the mean distribution of Y_i can be expressed via the following linear predictor:

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}.$$

The term “linear”, as used here, means that η_i , must be linear in the regression parameters.

Link function

The final way in which a generalized linear models extend the standard linear regression model is by taking a suitable transformation of the mean response and relating

the transformed mean response to the covariates. This is achieved by introducing a *link function*, denoted here by $g(\cdot)$:

$$g(\mu_i) = \eta_i = \sum_{k=1}^p \beta_k x_{ki} = \boldsymbol{\beta}' \mathbf{x}_i.$$

where $\mu_i = E(Y|\mathbf{x}_i)$. The link function will be a known function such as “log”. This implies that it is the transformed mean response that changes linearly with changes in the values of the covariates. Thus, while in the standard linear regression model the mean response is related directly to a linear combination of the covariates, in generalized linear models, it is some appropriate transformation of the mean response, that is related to a linear combination of the covariates. The choice of the link function should be made such that the model produces predictions of the mean response that are within the allowable range.

There are two main types of link functions, *canonical* link functions and *non-canonical* link functions. The former are unique and can be derived for any selected distribution. The latter are somewhat arbitrary and bear no direct relation to the selected distribution. Example of a canonical link is logit which is associated to the Bernoulli distribution. The GLM with a logit link function is also called logistic regression. To notice is that the standard linear regression is obtained if the link function is the identity (Fitzmaurice, Laird and Ware, 2004).

4.2 GEE Methodology

Let's consider the density function of the exponential family (4.1). The joint density for a set on n independent outcomes subscripted from y_1, y_2, \dots, y_n is the product of the densities for the individual outcomes:

$$f(y_1, \dots, y_n, \theta, \phi) = \prod_{i=1}^n \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\}.$$

Then, the likelihood is simply a restatement of the joint density where we consider the outcomes as given, and model the parameters as unknown.

$$L(\theta, \phi | y_1, \dots, y_n) = \prod_{i=1}^n \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\}.$$

Now, the log-likelihood of the exponential family is

$$\mathcal{L}(\theta, \phi | y_1, \dots, y_n) = \sum_{i=1}^n \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\}.$$

The goal is to obtain a maximum likelihood estimator for θ . Since the focus is only on θ , we derive an estimating equation where we treat the dispersion parameter $a(\phi)$ as ancillary.

Our estimating equation is then, $\Psi(\Theta) = \partial \mathcal{L} / \partial \theta$ where we derive

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^n \frac{y_i - b'(\theta)}{a(\phi)}.$$

Using the GLM result that in canonical form $b'(\theta) = \mu$, we may write:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)}.$$

Since the goal is to introduce covariates that model the outcome, we included a subscript on μ allowing the mean to reflect a dependence on a linear combination of the covariates and their associated coefficients. We can now use the chain rule to obtain a more useful form of the estimating equation $\Psi(\theta) = 0$ for $\theta = \beta_{p \times 1}$.

$$\frac{\partial \mathcal{L}}{\partial \beta} = \left[\left(\frac{\partial \mathcal{L}}{\partial \theta} \right) \left(\frac{\partial \theta}{\partial \mu} \right) \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial \eta}{\partial \beta_j} \right) \right]_{(p \times 1)} \quad (4.2)$$

$$= \left[\sum_{i=1}^n \left(\frac{y_i - b'(\theta_i)}{a(\phi)} \right) \left(\frac{1}{V(\mu_i)} \right) \left(\frac{\partial \mu}{\partial \eta} \right)_i (x_{ji}) \right]_{(p \times 1)} \quad (4.3)$$

$$= \left[\sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)V(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i x_{ji} \right]_{(p \times 1)}. \quad (4.4)$$

So the general estimating equation for the exponential family is given by:

$$\left[\left\{ \frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)V(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i x_{ji} \right\}_{j=1, \dots, p} \right]_{p \times 1} = [0]_{p \times 1}. \quad (4.5)$$

The result is an estimating equation derived from the exponential family of distributions where the expected value of the outcome variable is parameterized in terms of a linear predictor.

The assumption behind the derivation of the above estimating equations is that the observations are independent. This assumption however, may not always hold as in the case of our data set. Let's assume that we have $i = 1, \dots, n$ patients (clusters) where each patient has $t = 1, \dots, n_i$ correlated observations. With this notation, the equations become:

$$\Psi(\beta) = \left[\left\{ \frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^n \sum_{t=1}^{n_i} \frac{y_{it} - \mu_{it}}{a(\phi)V(\mu_{it})} \left(\frac{\partial \mu}{\partial \eta} \right)_{it} x_{jit} \right\}_{j=1, \dots, p} \right]_{p \times 1} = [0]_{p \times 1}. \quad (4.6)$$

where p is the dimension of the matrix of covariates \mathbf{X} . Apart from a second subscript, the pooled estimating equation is no different from equation 4.6. The implied likelihood of the estimating equation does not address any second order dependence of the data.

We begin introducing the dependence by rewriting the equation in matrix form:

$$\Psi(\beta) = \left[\left\{ \sum_{i=1}^n X_{ji}^T D \left(\frac{\partial \mu}{\partial \eta} \right) [V(\mu_i)]^{-1} \left(\frac{y_i - \mu_i}{a(\phi)} \right) \right\}_{j=1, \dots, p} \right]_{p \times 1} = [0]_{p \times 1}. \quad (4.7)$$

where $D()$ denotes a diagonal matrix. $V(\mu_i)$ is a diagonal matrix which can be decomposed into:

$$V(\mu_i) = [D(V(\mu_{it}))^{1/2} I_{(n_i \times n_i)} D(V(\mu_{it}))^{1/2}]_{n_i \times n_i}. \quad (4.8)$$

This makes it clear that the estimating equation is treating each observation within a panel as independent. The GEE proposed by Liang and Zeger (1986) is a modification

of the estimating equation for GLMs that simply replaces the identity matrix with a more general correlation matrix, since the variance matrix for correlated data does not have a diagonal form.

$$V(\mu_i) = [D(V(\mu_{it}))^{1/2} R(\gamma)_{(n_i \times n_i)} D(V(\mu_{it}))^{1/2}]_{n_i \times n_i}. \quad (4.9)$$

We write $R(\gamma)$ to emphasize that the correlation matrix is to be estimated through the parameter vector γ . This matrix is called working correlation matrix and may have different forms (Hardin and Hilbe, 2003).

It is reasonable to assume a time dependence for the association of the repeated observations within the panels that have a natural order. The correlation structure is assumed to be $\text{corr}(y_{it}, y_{it'}) = \gamma^{|t-t'|}$. For normally distributed y_{it} , this is analogous to a continuous time autoregressive (AR) process (Hardin and Hilbe, 2003).

In this case, γ is a vector and we estimate the correlation using the Pearson residuals

$$\hat{r}_{it} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{V(\hat{\mu}_{it})}}.$$

from the current fit of the model.

$$\hat{\gamma} = \frac{1}{\hat{\phi}} \left[\sum_{i=1}^n \left(\frac{\sum_{t=1}^{n_i-0} \hat{r}_{i,t} \hat{r}_{i,t+0}}{n_i}, \dots, \frac{\sum_{t=1}^{n_i-k} \hat{r}_{i,t} \hat{r}_{i,t+k}}{n_i} \right) \right]. \quad (4.10)$$

4.3 Models Assuming Poisson Distribution

So far, the response variables analyzed were the three symptoms, wheeze, breath and cough during day and night. Now we will concentrate on the analysis of response variables AMUSE and PMUSE which count the number of times rescue medication is used during the day and night respectively. Since the variables represent discrete counts that can be considered to be distributed as Poisson, the suitable model is a log-linear model. This is obtained from the generalized linear model by considering

the link function to be “log”:

$$\log(\mu) = \alpha + x\beta.$$

We can rewrite the model as:

$$\mu = e^\alpha e^{x\beta}.$$

Frequently, discrete counts represent information collected over time (days, years) and interest lies in modelling rates. If the exposure time is denoted by N , we write the rate as Y/N and the expected value as μ/N . Modelling the rate, the log-linear model becomes:

$$\begin{aligned}\log\left(\frac{\mu}{N}\right) &= \alpha + x\beta \\ \log(\mu) &= \alpha + x\beta + \log(N)\end{aligned}$$

The term $\log(N)$ is called an offset and must be accounted for in the estimation process.

Because there are so many observations per patient, the data are restructured by creating 5 time intervals that are almost equal. Thus, the first interval is represented by the observations prior to day 1, the second contains the observations between day 1 and day 20, the third between day 21 and day 40, the fourth between day 41 and day 60 whereas the last one contains observations after day 60. The counts in all these intervals were added and thus the new data set contains only 1444 observations. Since the number of patients is 299 this means that not all of them had observations in the last or first interval. The new variables created are:

- **OCCASION** - with values from 0 to 4, denotes the five intervals
- **DAYS** - counts the number of observations that corresponds to each interval for each patient
- **TAMUSE** - counts the number of times rescue medication is used in each interval during the day

- **TPMUSE** - counts the number of times rescue medication is used in each interval during the night

4.3.1 Independent observations

We will first treat the observations as being independent. The response variables are TAMUSE and TPMUSE whereas the covariates considered here are TRTMNT, FREQBASE and OCCASION. The procedure in SAS to model these variables is GENMOD with the Poisson distance and log link. Table 4.1 shows the estimates for TAMUSE.

Table 4.1: TAMUSE parameter estimates - independence

Parameter		DF	Estimate	Pr>ChiSq
Intercept		1	-0.8902	<.0001
TRTMNT	ACTIVE	1	-0.0573	<.0001
TRTMNT	PLACEBO	0	-	-
FREQBASE	CHRONIC	1	0.2862	<.0001
FREQBASE	EPISODIC	0	-	-
OCCASION	0	1	0.7410	<.0001
OCCASION	1	1	0.4347	<.0001
OCCASION	2	1	0.3177	<.0001
OCCASION	3	1	0.0602	0.0209
OCCASION	4	0	-	-

We can see that all three variables are significant at all levels. By exponentiating the parameters we obtain the incidence ratio of Active vs Placebo, Chronic vs Episodic, Occasion 0 vs Occasion 4, Occasion 1 vs Occasion 4, etc. For instance, the incidence ratio of Active vs Placebo is 0.9443 which shows us that patients in the Active group make use of rescue medication a little less than patients in the Placebo group.

The incidence ratios are in table 4.2. We can notice that patients that experience

Table 4.2: Incidence ratios TAMUSE - independence

Ratio	Ratio estimate
Active vs Placebo	0.9443
Chronic vs Episodic	1.3313
Occasion 0 vs Occasion 4	2.0900
Occasion 1 vs Occasion 4	1.5444
Occasion 2 vs Occasion 4	1.3739
Occasion 3 vs Occasion 4	1.0620

Chronic symptoms use rescue medication more often than those experiencing episodic symptoms. We can also see that as time goes by the number of times Ventolin was used, decreases. Thus, before receiving the intervention the number of rescue medication use was two times greater than in the last 20 to 25 days. The ratio becomes close to one as the time intervals are closer to each other. Table 4.3 shows the parameter estimates when the response variable is TPMUSE.

The results are almost similar except that OCCASION=3 is not significant. Again the estimate for TRTMNT is negative, suggesting a ratio of incidences smaller than one. Table 4.4 summarizes the ratios for all variables in the model. Covariates that were not significant have a star next to their ratio.

We can see a bigger difference in Active vs Placebo than the one observed for AMUSE. The same pattern as before can be observed for the other variables.

The goodness of fit statistics for both response variables suggested that there is overdispersion. To deal with this problem we will later assume instead of Poisson the Negative Binomial distribution for the outcomes.

Table 4.3: TPMUSE parameter estimates - independence

Parameter		DF	Estimate	Pr>ChiSq
Intercept		1	-1.5887	<.0001
TRTMNT	ACTIVE	1	-0.1818	<.0001
TRTMNT	PLACEBO	0	-	-
FREQBASE	CHRONIC	1	0.1829	<.0001
FREQBASE	EPISODIC	0	-	-
OCCASION	0	1	0.4900	<.0001
OCCASION	1	1	0.3138	<.0001
OCCASION	2	1	0.3378	<.0001
OCCASION	3	1	0.0531	0.1781
OCCASION	4	0	-	-

Table 4.4: Incidence ratios TPMUSE - independence

Ratio	Ratio estimate
Active vs Placebo	0.8337
Chronic vs Episodic	1.2006
Occasion 0 vs Occasion 4	1.6323
Occasion 1 vs Occasion 4	1.3686
Occasion 2 vs Occasion 4	1.4018
Occasion 3 vs Occasion 4	1.0545*

4.3.2 Introducing correlations using GEE

Next we will introduce the dependence between the observations using the generalized estimating equations. The choice for the working correlation matrices is the an autoregressive matrix. The model remains the same, that is, considering the out-

comes distributed as Poisson and keeping TRTMNT, FREQASE and OCCASION as covariates.

Table 4.5 contains the parameter estimates for TAMUSE.

Table 4.5: TAMUSE parameter estimates - dependence

Parameter		DF	Estimate	Pr>ChiSq
Intercept		1	-0.9316	<.0001
TRTMNT	ACTIVE	1	0.0374	0.7033
TRTMNT	PLACEBO	0	-	-
FREQBASE	CHRONIC	1	0.3478	0.0003
FREQBASE	EPISODIC	0	-	-
OCCASION	0	1	0.6980	<.0001
OCCASION	1	1	0.3957	<.0001
OCCASION	2	1	0.2861	<.0001
OCCASION	3	1	0.0446	0.4695
OCCASION	4	0	-	-

We can see that the TRTMNT is no longer significant. The other coefficients however, indicate that patients experiencing chronic symptoms use rescue medication more often and that the use of rescue medication decreases in time.

The estimated correlation matrix used to compute the parameters is:

$$R(\alpha) = \begin{bmatrix} & occ0 & occ1 & occ2 & occ3 & occ4 \\ occ0 & 1.0000 & 0.5540 & 0.3069 & 0.1700 & 0.0942 \\ occ1 & 0.5540 & 1.0000 & 0.5540 & 0.3069 & 0.1700 \\ occ2 & 0.3069 & 0.5540 & 1.0000 & 0.5540 & 0.3069 \\ occ3 & 0.1700 & 0.3069 & 0.5540 & 1.0000 & 0.5540 \\ occ4 & 0.0942 & 0.1700 & 0.3069 & 0.5540 & 1.0000 \end{bmatrix}$$

We can see that observations that are closer in time are more correlated with each other.

The parameter estimates for TPMUSE are in Table 4.6.

Table 4.6: TPMUSE parameter estimates - dependence

Parameter		DF	Estimate	Pr>ChiSq
Intercept		1	-1.6548	<.0001
TRTMNT	ACTIVE	1	0.0360	0.8003
TRTMNT	PLACEBO	0	-	-
FREQBASE	CHRONIC	1	0.1399	0.0003
FREQBASE	EPISODIC	0	-	-
OCCASION	0	1	0.1113	0.0002
OCCASION	1	1	0.1015	0.0123
OCCASION	2	1	0.0965	0.0040
OCCASION	3	1	0.0781	0.7292
OCCASION	4	0	-	-

The situation is almost similar to the one for TAMUSE. To notice is that during the night the use of rescue medication does not decrease as much in time as during the day. The estimated correlation matrix used to compute the estimates is:

$$R(\alpha) = \begin{bmatrix} & occ0 & occ1 & occ2 & occ3 & occ4 \\ occ0 & 1.0000 & 0.6732 & 0.4533 & 0.3052 & 0.2054 \\ occ1 & 0.6732 & 1.0000 & 0.6732 & 0.4533 & 0.3052 \\ occ2 & 0.4533 & 0.6732 & 1.0000 & 0.6732 & 0.4533 \\ occ3 & 0.3052 & 0.4533 & 0.6732 & 1.0000 & 0.6732 \\ occ4 & 0.2054 & 0.3052 & 0.4533 & 0.6732 & 1.0000 \end{bmatrix}$$

Next, in Table 4.7 we can see the incidence ratios for the two variables analyzed. Covariates that were not significant have a star next to their ratio.

Table 4.7: Incidence ratios - dependence

Ratio	Ratio estimates	
	TAMUSE	TPMUSE
Active vs Placebo	1.0381*	1.0366*
Chronic vs Episodic	1.4159	1.1501
Occasion 0 vs Occasion 4	2.0097	1.1177
Occasion 1 vs Occasion 4	1.4854	1.1068
Occasion 2 vs Occasion 4	1.3312	1.1013
Occasion 3 vs Occasion 4	1.0456*	1.0812*

4.4 Model Assuming the Negative Binomial Distribution

The goodness of fit tests for the models that assumed independence, had the Pearson and Deviance scores over the degrees of freedom around 17. Thus, it seems that there was overdispersion. Overdispersion occurs when the observed variance is larger than the nominal variance for a particular distribution. This is not surprising for the assumed distribution because the respective variance is fixed by a single parameter, the mean. With the Poisson distribution, the rate λ is assumed fixed. If however, the rate λ which generates the counts is not the same for all patients, then assuming a Gamma distribution with parameters α and β for the distribution of the rate λ , the counts will be distributed negative binomial with the same parameters α, β . Moreover, the variance as a function of the mean in the case of the negative binomial is $V(\mu) =$

$\mu + k\mu^2$, where k is another parameter that is estimated through maximum likelihood.

4.4.1 Independent observations

Table 4.8 compares the results of the goodness of fit tests when using the Poisson and the Negative Binomial distributions for TAMUSE.

Table 4.8: Goodness of fit for TAMUSE

Distribution	Criterion	DF	Value	Value/DF
Poisson	Deviance	1437	26688.4854	18.5724
	Pearson	1437	28809.4547	20.0483
	Log Likelihood		38751.0179	
Negative Binomial	Deviance	1437	1654.7594	1.1515
	Pearson	1437	951.4073	0.6621
	Log Likelihood		49320.6432	

We can see that in the case of negative binomial, the Deviance over DF is very close to one which tells us that we have goodness of fit. The Pearson over DF is however less than one which implies that we have underdispersion. However the choice of negative binomial distribution seems to be much better than the choice of Poisson.

Table 4.9 shows the goodness of fit for variable TPMUSE. Using the negative binomial distribution, the results of the goodness of fit tests are even better than for TAMUSE.

We can see the estimates for these response variables in Table 4.10.

The interpretation is the same as in the case of a Poisson distribution. For TAMUSE we can see that TRTMNT is not significant but FREQBASE and OCCASION are. The last one suggests that in time, patients need less and less rescue

Table 4.9: Goodness of fit for TPMUSE

Distribution	Criterion	DF	Value	Value/DF
Poisson	Deviance	1437	16520.6832	11.4966
	Pearson	1437	23165.3890	16.1207
	Log Likelihood		6984.0663	
Negative Binomial	Deviance	1437	1457.2827	1.0141
	Pearson	1437	1266.5072	0.8814
	Log Likelihood		13063.9759	

Table 4.10: TAMUSE and TPAMUSE parameter estimates - negative binomial

Parameter			TAMUSE		TPMUSE	
			Estimate	Pr>ChiSq	Estimate	Pr>ChiSq
Intercept		1	-0.8375	<.0001	-1.5022	<.0001
TRTMNT	ACTIVE	1	-0.1013	0.2069	-0.2121	0.0325
TRTMNT	PLACEBO	0	-	-	-	-
FREQBASE	CHRONIC	1	0.2469	0.0023	0.1051	0.2920
FREQBASE	EPISODIC	0	-	-	-	-
OCCASION	0	1	0.7342	<.0001	0.4803	0.0021
OCCASION	1	1	0.4370	0.0006	0.2853	0.0663
OCCASION	2	1	0.3333	0.0093	0.3282	0.0363
OCCASION	3	1	0.0932	0.4698	0.1187	0.4550
OCCASION	4	0	-	-	-	-

medication. On the other hand, for TPMUSE, TRTMNT is significant indicating less use of rescue medication by patients receiving the active drug, whereas FREQBASE isn't significant.

4.4.2 Introducing correlations using GEE

So far the analysis assumed that the observations are independent. Using the negative binomial instead of Poisson, and taking into account the dependence between the observations by using GEEs, we obtain for the two variables the following estimates summarized in Table 4.11:

Table 4.11: TAMUSE and TPMUSE parameter estimates - negative binomial GEE

Parameter			TAMUSE		TPMUSE	
			Estimate	Pr>ChiSq	Estimate	Pr>ChiSq
Intercept		1	-0.8084	<.0001	-1.4675	<.0001
TRTMNT	ACTIVE	1	-0.0538	0.6072	-0.1368	0.3698
TRTMNT	PLACEBO	0	-	-	-	-
FREQBASE	CHRONIC	1	0.2738	0.0091	0.1420	0.3535
FREQBASE	EPISODIC	0	-	-	-	-
OCCASION	0	1	0.6613	<.0001	0.3722	0.0009
OCCASION	1	1	0.3698	<.0001	0.1891	0.0683
OCCASION	2	1	0.2719	0.0002	0.2241	0.0205
OCCASION	3	1	0.0436	0.4803	0.0287	0.6988
OCCASION	4	0	-	-	-	-

We can see that the estimates for TRTMNT are both negative but not statistically significant. On the other hand, both estimates for FREQBASE are positive but only the one for TAMUSE is significant. The OCCASION is significant and shows that the use of rescue medication decreases in time. Table 4.13 shows the incidence ratios.

Using the negative binomial and taking into account the dependence of the observations, we can see that in time the number of times rescue medication is used is decreasing. The variable TRTMNT was not significant.

Table 4.12: Incidence ratios - independence

Ratio	Ratio estimates	
	TAMUSE	TPMUSE
Active vs Placebo	0.9036*	0.8088*
Chronic vs Episodic	1.2800	1.1108*
Occasion 0 vs Occasion 4	2.0832	1.6165
Occasion 1 vs Occasion 4	1.5480	1.3301
Occasion 2 vs Occasion 4	1.3312	1.1013
Occasion 3 vs Occasion 4	1.0976*	1.1260*

Table 4.13: Incidence ratios - dependence

Ratio	Ratio estimates	
	TAMUSE	TPMUSE
Active vs Placebo	0.9476*	0.8721*
Chronic vs Episodic	1.3149	1.1525*
Occasion 0 vs Occasion 4	1.9373	1.4509
Occasion 1 vs Occasion 4	1.4474	1.2081
Occasion 2 vs Occasion 4	1.3124	1.2511
Occasion 3 vs Occasion 4	1.0445*	1.0291*

Chapter 5

Conclusions

The descriptives summaries of asthma symptoms show a general improvement in the patients' state of health. The difference between Active and Placebo is not however obvious at all levels of the symptom response variables. Over time, the graphs show that the use of rescue medication is decreasing and we can detect a difference between the two interventions in favor of the Active group.

A continuation ratio model is used to analyze the symptom response variables assuming first that the observations are independent all six symptoms, TRTMNT and FREQBASE were found to be significant, showing a difference between the Active and Placebo groups in favor of the patients that were receiving the Active intervention. The independent assumption is relaxed by incorporating in the model a lagged covariate. The results are similar. The goodness of fit tests suggests however that this model is not adequate to the data. The simulated data didn't seem to be adequate either, but the model was able to correctly detect a difference between Active and Placebo.

Finally, the generalized estimating equations methodology was used to analyze the remaining two response variables that count the number of times rescue medication is used. The data was restructured in such a way that a patient had only 5 obser-

vations. Assuming independence and a Poisson distribution for the outcomes, the analysis showed TRTMNT, FREQBASE and OCCASION to be statistically significant. The negative sign of the coefficient for TRTMNT shows a difference between Active and Placebo in favor of Active. That is, patients in the Active arm use less rescue medication than patients in the Placebo arm. The incidence ratios for OCCASION tell us that in time, the use of rescue medication decreases. The goodness of fit tests show however that the model is not adequate for the data. Assuming dependence of the observations by using GEE, the results are similar except that the coefficient of TRTMNT even though is negative is not statistically significant.

Because using a Poisson distribution resulted in overdispersion, the next natural choice was the negative binomial distribution. The goodness of fit tests in this case suggest that this distribution is appropriate for the data. Again, the covariate TRTMNT is not significant in neither cases, independent or dependent observations. OCCASION however, is significant and suggests as before that in time the use of rescue medication decreases.

Bibliography

- [1] Agresti, A. (1990). *Categorical Data Analysis*, Wiley.
- [2] Altman, Douglas G. (1991). *Practical Statistics for Medical Research*, Chapman & Hall.
- [3] Azzalini, A. (1994). Logistic Regression for Autocorrelated Data with Application to Repeated Measures, *Biometrika* **81**, 767-775.
- [4] Bender, R. and Benner, (2002). Calculating Ordinal Regression Models in SAS and S-Plus. *Biometrika Journal*, **42**, 677-699.
- [5] Carey, V., Zeger, S.L. and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions, *Biometrika*, **80**, 517-526.
- [6] Conaway, M.R. (1989). Analysis of Repeated Categorical Measurements with Conditional Likelihood, *Journal of the American Statistical Association*, **84**, 53-62.
- [7] Diggle, P.J., Heagerty, P.J., Liang, K. and Zeger, S.L. (2002). *Analysis of Longitudinal Data, Second Edition*, Oxford University Press.
- [8] Dirley, M. Berridge, M. (2000). A continuation ratio random effects model for repeated ordinal responses, *Statistics In Medicine* **19**, 3377-3388.
- [9] Fitzmaurice, G.M., Laird N.M. and Ware, J.H. (2004). *Applied Longitudinal Analysis*, John Wiley & Sons.

- [10] Hardin, J.W. and Hilbe, J.M. (2003). *Generalized Estimating Equations*, Chapman & Hall/CRC.
- [11] Korn, E.L. and Whittemore, A.S. (1979). Methods for analyzing panel studies of acute health effects of air pollution, *Biometrics*, **35**, 795-802.
- [12] Liang, K. and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* **73**, 13-22.
- [13] Lindsey, J.K., Jones, B. and Ebbutt, A.F. (1997). Simple models for repeated ordinal responses with application to a seasonal rhinitis clinical trial, *Statistics In Medicine*, **16**, 2873-2882.
- [14] Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratios as a measure of association. *Biometrika* **78**, 153-160.
- [15] Muenz, L.R. and Rubinstein, L.V. (1985). Markov models for covariate dependence of binary sequences, *Biometrics* **41**, 91-101.
- [16] Rees, J. and Kanabar, D. (2000). *ABC of Asthma*, BMJ Books.
- [17] Stokes, M.E., Davis, C.S., Koch, G.G. (2000). *Categorical Data Analysis Using the SAS System*, Second Edition, Cary, NC: SAS Institute, 2000.

Appendix 1

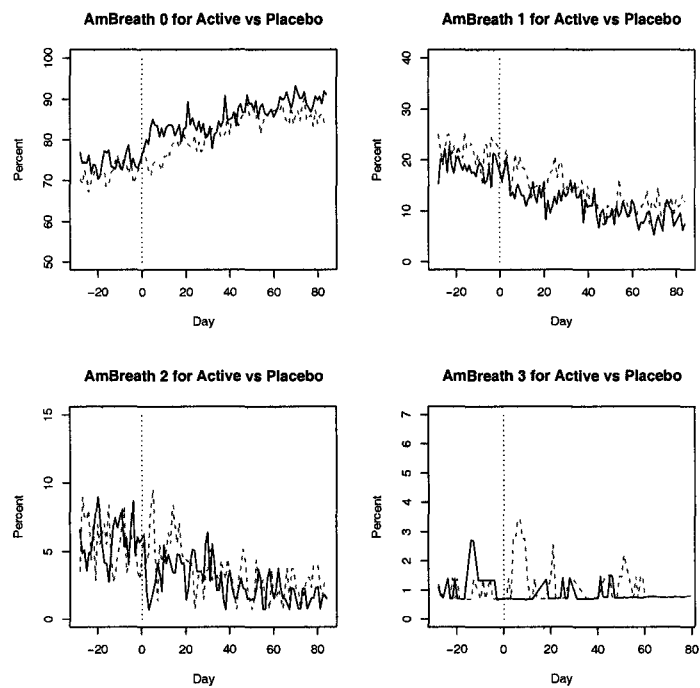


Figure 5.1: AMBREATH by DAY and TRTMNT

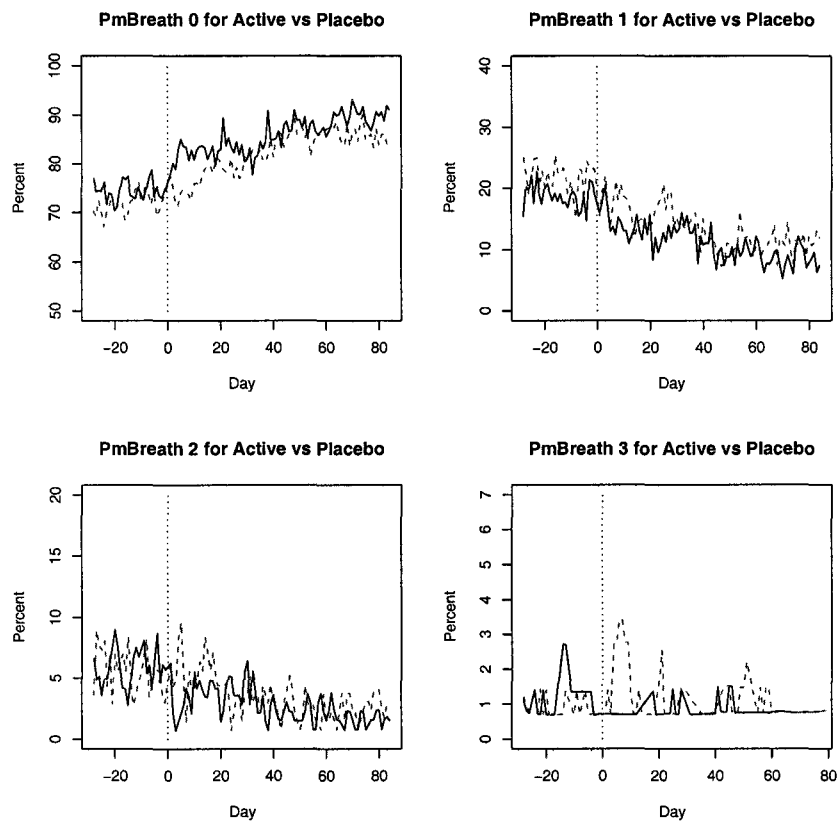


Figure 5.2: PMBREATH by DAY and TRTMNT

Appendix 2

To obtain the number of subjects in each study arm, Placebo and Active:

```
data asthma;
  set asthma;
  if ana='Y';
run;
proc freq data=asthma;
  tables subject*trtmnt/ out=treatment nocol nopercent nocum norow;
run;
proc freq data=treatment;
  tables trtmnt / nocol nopercent nocum norow;
run;
```

To obtain the descriptive tables:

```
data one;
  set asthma;
  by subject;
  if first.subject=1;
run;
proc freq data=one;
  tables trtmnt*freqbase trtmnt*gender trtmnt*wdw
  trtmnt*study trtmnt*wdwreas\ nocol norow nocum nopercent;
run;

proc freq data=asthma;
  tables subject*country/ out=country nocol nopercent nocum norow;
run;
proc freq data=country;
```

```

    tables country / nocol nopercent nocum norow;
run;

```

Code to obtain the percentage of symptom free days:

```

data free;
    set asthma;
    symp=amwheeze+amcough+ambreath+pmwheeze+pmcough+pmbreath;
    if day<0 then d='before';
    else d='after';
run;
data free;
    set free;
    if symp=0 then al='free';
    else if symp='.' then al='NA';
    else al='symptom';
run;
proc sort data=free;
    by trtmnt;
run;
proc freq data=free;
    by trtmnt;
    tables al*d;
run;

```

Macro to obtain the data sets:

```

%macro datasets(varb,dset);
data new1;
    set asthma;
    if trtmnt='ACTIVE';
run;
proc freq data=new1;
    table day * &varb /nocol nocum nopercent outpct out=new1;
run;
data new1;
    set new1;
    if &varb ne '.';
    trtmnt='ACTIVE';

```



```

    keep day &varb count pct_row trtmnt;
run;
data new2;
    set asthma;
    if trtmnt='PLACEBO';
run;
proc freq data=new2;
    table day * &varb /nocol nocum nopercnt outpct out=new2;
run;
data new2;
    set new2;
    if &varb ne '.';
    trtmnt='PLACEBO';
    keep day &varb count pct_row trtmnt;
run;
data &dset;
    set new1 new2;
run;
proc export data = &dset
    outfile="C:\Documents and Settings\Cristina\My
Documents\Dragos\SAS-lucrare\&dset"
    DBMS=CSV replace;
run;
%mend datasets;
%datasets(amwheeze,amwh)
%datasets(amcough,amcogh)
%datasets(ambreath,ambr)
%datasets(pmwheeze,pmwh)
%datasets(pmcough,pmcogh)
%datasets(ambreath,pmbr)

```

Macro to obtain the estimates of the model:

```

\%macro symptoms(varb , dset);
data asthmad;
    set asthma;
    if day>0;
run;
data ast1;
    set asthmad;

```

```

    if &varb<=1; cp=1;
    if &varb=1 then br=1; else br=0;
data ast2;
    set asthmad;
    if &varb<=2; cp=2;
    if &varb=2 then br=1; else br=0;
data ast3;
    set asthmad;
    if &varb<=3;cp=3;
    if &varb=3 then br=1; else br=0;
data &dset;
    set ast1 ast2 ast3;
    if cp=3 then cp3=1; else cp3=0;
    if cp=2 then cp2=1; else cp2=0;
run;
proc logistic data=&dset descending;
    class trtmnt freqbase/param=ref;
    model br=cp2 cp3 trtmnt freqbase/scale=none aggregate;
run;

%mend;
\%symptoms(ambwheeze, damwh) \%symptoms(ambreath, dambr)
\%symptoms(amcough, damch) \%symptoms(pmweeze,dpmwh)
\%symptoms(pmbreath, dpnbr) \%symptoms(pmcough, dpmch)

```

Simulation macro:

```

%macro simulation(varb, dset, n);
data asthmad;
    set asthma;
    if day>0;
run;
data act;
    set asthmad;
    if trtmnt='ACTIVE';
data plc;
    set asthmad;
    if trtmnt='PLACEBO';
run;
%do j=1 %to &n;
data act;

```

```

set act;
  %do i=1 %to %length(trtmnt);
    x=uniform(-1);
  %end;
  if x<0.25 then y=0;
    else if x<0.75 then y=1;
      else if x<0.95 then y=2;
        else y=3;
run;
data plc;
  set plc;
  %do i=1 %to %length(trtmnt);
    x=uniform(-1);
  %end;
  if x<0.15 then y=0;
    else if x<0.60 then y=1;
      else if x<0.90 then y=2;
        else y=3;
run;
data asthmay;
  set act plc;
run;
data ast1;
  set asthmay;
  if &varb<=1 ;cp=1;
  if &varb=1 then br=1; else br=0;
data ast2;
  set asthmay;
  if &varb<=2; cp=2;
  if &varb=2 then br=1; else br=0;
data ast3;
  set asthmay;
  if &varb<=3; cp=3;
  if &varb=3 then br=1; else br=0;
data &dset;
  set ast1 ast2 ast3;
  if cp=3 then cp3=1; else cp3=0;
  if cp=2 then cp2=1; else cp2=0;
run;
proc logistic data=&dset descending;
  class trtmnt freqbase/param=ref;

```

```

    model br=cp2 cp3 trtmnt freqbase/scale=none aggregate;
    ods output GoodnessOfFit=d&j ParameterEstimates=p&j;
run;
%end;
data fin;
set
    %do k=1 %to &n;
        d&k
    %end;
;
run;
data param;
set
    %do k=1 %to &n;
        p&k
    %end;
;
run;
%mend;
%simulation(y,amwh,100)

```

Creating TAMUSE and TPMUSE by restructuring the data and the GEE code

```

data asth;
set asthma;
if day<0 then ocasion=0;
else if day<20 then ocasion=1;
else if day<40 then ocasion=2;
else if day<60 then ocasion=3;
else ocasion=4;
run;
data asth;
set asth;
if amuse='.' then amuse=0;
if pmuse='.' then pmuse=0;
run;
data asth;
set asth;
by subject ocasion;
retain days tamuse tpmuse;

```

```

if first.ocasion then do;
tamuse=0;tpmuse=0;days=0;
end;
tamuse=tamuse+amuse;
tpmuse=tpmuse+pmuse;
days=days+1;
if last.ocasion then output;
run;
data asth;
set asth;
ldays=log(days);
run;

```

```

proc genmod data=asth;
class subject trtmnt freqbase occasion;
model tamuse=trtmnt freqbase occasion /dist=poisson link=log type3 offset=ldays;
repeated subject=subject /type=ar corrw;
run;

```