Using the theory of Longitudinal Data to model the use of antipsychotics in elderly patients with dementia

A Statistical Method for Analysing Drug-use

Master Thesis

Sally Kit Ipsen

Aalborg University
Mathematics

This report is written using LaTeX for the typesetting. Furthermore STATA was used for the preparation of the data, and R was used when performing the analysis including making plots and tables of results.

Department of Mathematical Sciences
Fredrik Bajers Vej 7G
DK-9220 Aalborg Ø
http://math.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Using the theory of Longitudinal Data to model the use of antipsychotics in elderly patients with dementia

Theme:
A Statistical Method for Analysing Drug-use

Project Period:
Spring Semester 2016

Project Group:
G5-105b

Participant:
Sally Kit Ipsen

Supervisor:
Poul Svante Eriksen

Copies: 4

Page Numbers: 62

Date of Completion:
June 10, 2016

Abstract:

This project examines the use of antipsychotics among elderly patients with dementia. In 2007 the Danish Health and Medicines Authority issued a guideline including an upper limit of how much antipsychotic medication an elderly patient with dementia should receive.
The analysis is performed using the theory of longitudinal data. The model used is a linear mixed effects model with random intercepts. The estimates obtained from this model, expresses how much effect the different variables have on the use of medication.
The conclusion is that the patients still use the same amount of antipsychotics regardless the guideline. When adjusted for all other variables, there still is an increase in the amount used over time - even after the guideline was issued.

DANSK RESUMÉ

I Danmark formodes der at eksistere et overforbrug af antipsykotisk medicin blandt ældre patienter med demens. Tidligere studier har vist, at almindelige mængder af denne medicin er meget farlig for denne patientgruppe, samt at netop demente normalt ikke har gavn af medicinen. I 2007 udsendte Sundhedsstyrelsen derfor en vejledning netop rettet mod dette problem. I denne vejledning er der fastsat en øvre grænse for, hvor store mængder antipsykotisk medicin demenspatienter bør modtage. Grænsen er fastsat udfra tidligere studiers erfaringer. Siden vejledningen udkom har ingen undersøgt det faktiske forbrug, men sundhedsministeriet skriver i en rapport, at de formoder et overforbrug eksisterer, men at omfanget af dette indtil videre er ukendt.

Ved hjælp af teorien om longitudinal data skal forbruget af antipsykotisk medicin blandt ældre patienter med demens altså undersøges. Dette gøres med en såkaldt *linear mixed effects model* med random intercepts, som tillader én patients observationer at være korrelerede. Variable inkluderet i modellen omfatter: køn, alder på diagnosetidspunktet, tid, tid siden diagnosen, om personen er diagnosticeret i psykiatrisk regi, somatisk regi eller hos egen læge (dvs. via recept på antidemensmedicin), og om patienten er diagnosticeret før eller efter vejledningen fra 2007. Medicinforbrug måles i DDD (Defined Daily Dosage), og grænsen fra Sundhedsstyrelsen er sat til 7 DDD.

Studiepopulationen består af $149,869$ patienter med demens, hvoraf $56,218$ bruger antipsykotisk medicin på mindst ét tidspunkt i løbet af perioden fra 1. januar 1995 til 31. december 2012. Studieperioden opdeles i kvartaler og består således af 72 kvartaler.

Resultatet af analysen er, at Sundhedsstyrelsens vejledning fra 2007 ikke har

haft nogen effekt på medicinforbruget. Som forventet bruger kvinder generelt mere medicin end mænd, og patienter diagnosticeret i psykiatrisk regi bruger mere end dem diagnosticeret i somatisk regi, som igen bruger mere end dem diagnosticeret ved egen læge. Herudover bruger patienter 10% mindre medicin per 10 år ældre de er på diagnosetidspunktet.

Efter at have justeret for alle andre faktorer, ender vi med at have en kurve for tidens effekt (indeks 100) på medicinforbruget. Grænsen på 7 DDD omregnes og svarer til indeks 40. Gennemsnitsforbruget er steget, og ligger i slutningen af 2012 omkring 20% højere end i starten af 1995, dvs. det er tre gange så højt som anbefalet.

Konklusionen må altså være, at demente i Danmark til stadighed og i højere grad end tidligere forbruger farlige mængder antipsykotisk medicin, samt at Sundhedsstyrelsen bør gribe ind.

CONTENTS

PREFACE

This master thesis is written by Sally Kit Ipsen in the spring of 2016 during the 10th semester of Mathematics at Institute for Mathematical Sciences at Aalborg University in collaboration with the National Center for Registerbased Research at Aarhus University and supervised by Poul Svante Eriksen.

It is assumed that the reader possesses the mathematical qualifications corresponding to the completion of the bachelor education of Mathematical Sciences at Aalborg University as a minimum. If not it is possible to skip the theory descriptions and read only the results of the analysis in chapter **??**. Furthermore it is advised to have the authors project report from 9th semester [Ipsen, 2015] at hand as there will be references to this.

I could not have written this report without National Center for Registerbased Research ($NCRR$) at Aarhus University. At NCRR I was able to be connected to a project, and through that project I gained access to the relevant data at Statistics Denmark. I would like to thank NCRR for providing access to the relevant danish registers, an office and computer equipment, and in particular thank Preben Bo Mortensen, Janne Larsen, Aske Astrup and Christiane Gasse for help when needed.

My supervisor at Aalborg University, Svante, deserves a big thank you for his patience with me and his always competent guidance.

Reading instructions

References throughout the report will be presented according to the Harvard system.
We always refers to the undersigned.
Figures, tables and equations are enumerated in reference to the chapter.

Notation

In general, capital letters represent random variables or matrices, while small letters are for specific observations. Scalars will be in normal type, and vectors and matrices will be in bold type.

Code

All do-files from the data manipulation in Stata, and all R-files used to perform the analysis and making plots are available at `http://github.com/SallyIpsen/ninja-turtles`. Furthermore, [Ipsen, 2015] can be found here.

For my lovely husband, Kim, for his enormous support both mentally and emotionally, for help with github and tech support whenever my computer acted up, and last (but not least) for believing in me every minute of every day.

Aalborg University, June 10, 2016

Sally Kit Ipsen
<smorte06@student.aau.dk>

CHAPTER 1

INTRODUCTION

This master thesis is a continuation of my project from the 9th semester: Using the theory of Longitudinal Data to model the use of antipsychotic in elderly patients with dementia [Ipsen, 2015]. This project is available at `http://github.com/SallyIpsen/ninja-turtles` along with all do-files from the data manipulation in Stata and all R-files used to perform analysis and making plots.

Dementia is a progressive syndrome that is characterized by certain symptoms of failing brain function. For more details about the illness and the problems caused by dementia see [Ipsen, 2015, p.1]. World wide an increasingly large group of elderly citizens live with dementia. In Denmark the number is approximately $80,000$, but it is expected that this number will be doubled by the year 2035 [Socialministeriet, 2010, p.9], as a result from higher incidence and higher prevalence over time [Phung et al., 2007b, p.146].

The Danish Health and Medicines Authority tries to protect this group of patients with regard to the use of antipsychotic medication, because its use is associated with cumulative risk of severe adverse effects including death [Sundhedsstyrelsen, 2013; Gulmann, 2006]. Furthermore several studies show that discontinuation of antipsychotics in nursing homes does not increase the problems with violent behavior, on the contrary it had a positive impact on the function of the elderly [Gulmann, 2006]. The Danish Health and Medicines Authority therefore issued warnings and guidelines for the use of antipsychotic medication in elderly patients in 1991, 2000, 2004, 2005, and 2007. In the beginning of 2005 the FDA issued a warning against these drugs in USA. This point in time will prove to be important as well as latest

guideline from the Danish Health and Medicines Authority in 2007.

The guideline from 2007 states that elderly patients with dementia should not be treated with antipsychotics for more than one week equivalent to 7 DDD [Sundhedsstyrelsen, 2013, p.73]. Another limit is found to be 12 weeks per year, equivalent to 21 DDD per quarter, since use of more than that is associated with cumulative risk of severe adverse events including death [Ballard and Corbett, 2013]. The over-use of antipsychotics is only estimated and its actual extent is currently unknown [Sundhedsstyrelsen, 2005, p.1].

A new nationwide study states, that among patients with dementia, the prevalence of the use of antipsychotics decreased from 31.3% in 2000 to 20.4% in 2012 [Nørgaard et al., 2016, p.211]. This is of course a good thing in the light of the recommendation from the Danish Health and Medicines Authority, but as the number of patients with dementia is increasing, we do not know if the actual number of patients using these drugs is different. Furthermore, it is also a problem if the patients still taking the antipsychotic medication exceeds the limit, which is indicated since the annual median number of DDD increased from 33.3 in 2000 to 42.0 in 2012 [Nørgaard et al., 2016, p.211]. Consequently, a more thorough investigation is both relevant and necessary.

The last theme of motivation for this project is the lack of a statistical method when analyzing drug-use over time. There has not been registers on medication data available for long, only since 1995, so it is a relatively new research topic, but definitely exciting and very important. Scientists have asked, and still do, statisticians to develop a method to analyze the use of medication over time on an individual level.

Hence the aim of this project is to use the theory of longitudinal data to fit a statistical model for use of medication over time on an individual level. Since there seems to be a severe problem with over-use of antipsychotics in elderly patients with dementia, we use this as a case for our analysis.

At Statistics Denmark's server we have had access to the relevant registers and already made data sets, constructed to the project we are connected to. Thus we have used a data set named fstdemed02_2012 containing all patients in Denmark diagnosed with any kind of dementia. All information about medication is featured in a file labeled alldrugs_psykofarm. Because of the unique CPR number (pnr), a kind of personal security number, we are able to link the prescriptions from alldrugs_psykofarm to the patients in fstdemed02_2012. Furthermore we can link the data to a register named stamdata, which contains basic data about every danish

residents, for example date of birth, sex etc.

For the sake of the Danish Act on Processing of Personal Data the data never leaves Statistics Denmark's server. Instead we have worked with the data in Stata and R directly on the server, and then send the results "home" by e-mail.

CHAPTER 2

DATA PRESENTATION

This chapter contains a description of the data available, how they are used to construct the data set, and a presentation of both the variables and the study population. For more information about the danish registers from which these data come, see [Ipsen, 2015, p.5-7].

2.1 Data Available

From fstdemed02_2012 we have all patients identified with any kind of dementia in Denmark from the 1st of January 1969 until the 31st of December 2012. A patient is considered demented according to the following definition: subjects with an ICD-10 code corresponding to any kind of dementia *or* subjects who has received a prescription for antidementia medicine. The ICD-10 codes corresponding to dementia are the following: F00, F01, F02, F03 and G30. [Sundhedsstyrelsen, 2006] With each patient we have their date of diagnosis and information about how they were diagnosed. Dementia can be diagnosed in three different ways; in a psychiatric department in a hospital, in a somatic department in a hospital or by proxy if their general practitioner (GP) has prescribed antidementia medication.

The data set alldrugs_psykofarm gives us information about prescriptions; the patients pnr, date, what kind of medication, number of packages and volume in each package. Volume is measured in DDD, which stand for *Defined Daily Dosage*, and is a measurement defined by the World Health Organization [WHO, 2003].

In the register stamdata we find complementary information about the patients

such as sex, date of birth, if they are still alive and living in Denmark and if not, date of death or immigration.

2.2 Study Period

The study period lasts from 1st of January 1995 till 31st of December 2012. It is divided into quarters, such that it consists of 72 quarters. In the first quarter of the study period we have all patients diagnosed before 1st of January 1995. Then for each new quarter we add the patients diagnosed in the mean time and drop the ones who are now dead.

The separation into quarters is chosen because we wish to balance between two aims. First we would like some continuity in our data, meaning that we would like to avoid too many patients dropping in and out of medication use. If we divided the period into months, many patients would seem to be using a higher dosis but only every other month, because prescriptions provides antipsychotics to use for several weeks. On the other hand we do not just divide the study period into years, as we would like many repeated measurements (observations) over time. Hence quarters seems like an appropriate compromise.

Since the latest guideline from the Danish Health and Medicines Authority came in 2007, the period is separated at the 1st of January 2008 dividing the study period into two; one before and one after the issuing of the guideline. As such, we hope to see that the use of antipsychotics is different in the two time periods.

2.3 Study Population

The patients in the study population are born between the years 1892 and 1952. From date of birth and date of diagnosis the age when diagnosed is easily calculated. Subjects diagnosed before their 60th birthday was excluded from this study, since the validity of those diagnoses is weak [Phung et al., 2007a]. Furthermore the study only includes patients who are alive and residents in Denmark on the first day of the study period, 1st of January 1995.

Some patients in the study population were diagnosed both in a hospital and by proxy; 817 patients were diagnosed both in psychiatric care and by prescription, and 295 were diagnosed both in somatic care and by prescription. To avoid this overlap we defined, that the mentioned patients was only diagnosed in psychiatric or somatic care. There were no overlaps with psychiatric and somatic diagnosis type.

We end up having 149,869 elderly patients with dementia in the population. Table 2.1 presents an overview of the study population with regards to gender, age when diagnosed, diagnosis type and whether or not they use antipsychotics.

	Total		Females		Males	
	n	%	n	%	n	%
Gender	149,869	100	93,501	62.39	56,368	37.61
Age when diagnosed:						
60 – 69 years	14,795	9.9	7,016	47.4	7,779	52.6
70 – 79 years	47,658	31.8	27,265	57.2	20,393	42.8
80 – 89 years	71,308	47.6	46,988	65.9	24,320	43.1
90+ years	16,108	10.7	12,232	75.9	3,876	24.1
Not using antipsychotics	93,651	62.5	58,846	62.8	34,805	37.2
Using antipsychotics	56,218	37.5	34,655	61.6	21,563	38.4
Diagnosis type:						
Psychiatric	53,774	35.9	34,542	64.2	19,232	35.8
Somatic	85,411	57.0	52,907	62.0	32,504	38.0
Prescription	10,684	7.1	6,052	56.6	4,632	43.4
Diagnosis time:						
Before	107,880	72.0	67,974	63.0	39,906	37.0
After	41,989	28.0	25,527	60.8	16,462	39.2

Table 2.1: Overview of the study population with regards to gender, age when diagnosed, if they use medication and how they are diagnosed.

In all the age-groups, except for the youngest, there are more females than males. This is consistent with the danish population, where women in general live longer than men [Larsen, 2016]. Furthermore studies show that more women than men develop dementia [NVD, 2015].

2.4 Constructing the Data sets

Antipsychotic medication has ATC-code $N05A$. All other types of medication was disregarded. For each prescription we calculated DDD equal to number of packs (APK) times one pack's numeric volume (Volume) measured in DDD. The prescriptions were summarized over years and quarters, so the total number of DDD's per patient per quarter was obtained. The data set contains 495,177 rows in total.

In this data set a patient who didn't used antipsychotics in the study period fill only one line (row), and the total_DDD-variable is set to missing. As shown in table 2.1 that is the case for 93,651 of the patients in this study. This leaves 56,218 patients creating the remaining 401,526 lines. A patient who did use antipsychotics fill one line for every quarter he received one or more prescriptions. Hence some patients take up many rows and some only a few. We have 7,262 subjects who takes up 2 lines, meaning that they received prescriptions in two of 72 quarters. The higher number of lines the fewer subjects, and the data set contains 1,720 subjects who fill 10 lines, meaning that they received prescriptions in 10 of the 72 quarters.

With this data set in long format we calculated the mean DDD for those using antipsychotics from the whole population as well as from different sub populations: females, males, those diagnosed in psychiatric care, in somatic care and by prescription, and beforeafter guideline. The study population was divided into two groups according to the time of their diagnosis; all patients diagnosed up and through the year 2007, and those diagnosed after the 31st of December 2007. Hence we will analyze the mean DDD and look for potential differences between the sub populations. Furthermore this data set is used in R, when we perform the linear mixed effects with random intercept in chapter 5.

Another data set is constructed in Stata, when the data set is reshaped to wide format. In wide format there is one row per patient and a DDD-variable for each of the 72 quarters in the study period. When a patient is diagnosed and still alive, the DDD-variable contains a number; zero if the patient does not use medication and the volume (DDD) otherwise. For the quarters before the diagnosis and after a patients death, the DDD-variable is set to missing. That way we are able to calculate and illustrate the prevalence for each quarter i.e. how many percent of the patients with dementia use antipsychotics.

Now that the data and the variables has been described, the next chapter will present the methods used in this project.

CHAPTER 3

METHODS

This chapter begins with a brief illustration of the prevalence of antipsychotic drug use, followed by a description of the splines used on the mean DDD. When making the splines we use weights, which will also be explained here. Lastly the linear mixed effects model with random intercept is introduced along with methods for model diagnostics.

3.1 Decreasing prevalence

Figure 3.1 is a barchart of the number of patients with dementia not using antipsychotics (blue) and using antipsychotics (red) for each quarter in the years from 1995 to 2012.

The prevalence is decreasing over the years, but since the number of patients increases so drastically, the actual number of patients using antipsychotics is actually increased a little. The vertical green lines represents the FDA warning in USA in 2005 and the guideline from the Danish Health and Medicines Authority issued in 2007, respectively. We can see that the red bar tend to grow in length from 1995 to 2007, after which it slightly reduces.

Table 3.1 contains the actual number of patients receiving medication and also the prevalence for each of the 72 quarters.

Because of earlier studies, and because the guideline from the Danish Health and Medicines Authority states a limit of 7 DDD, it is not relevant to dig deeper into this prevalence over time, but instead investigate *how much* medication is used, by those

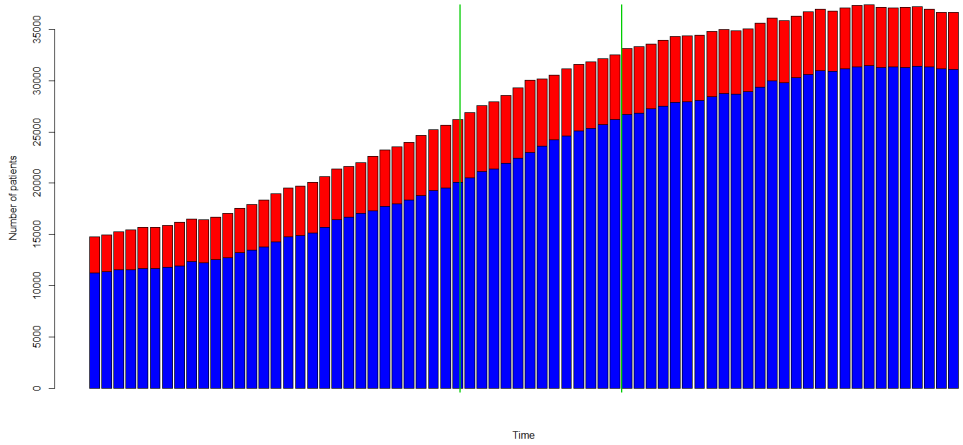

Figure 3.1: Barchart of the number of patients with dementia not using antipsychotics (blue) and using antipsychotics (red) over time from the first quarter of 1995 to the fourth quarter of 2012.

who actually use antipsychotics. For that purpose the DDD is a perfect measure. First we make splines for the mean DDD over time, and later we present a linear mixed effects model with random intercept.

3.2 The mean DDD and Splines

This section contains a description of the splines made in chapter 4 and an explanation of the weights used to make these splines.

A cubic spline is a spline constructed of piecewise third-order polynomials which pass through one or more knots. The knots divides the function in intervals. The cubic spline is continuous, and both the first and second derivative in the knots are continuous as well. A cubic spline is of order 4. In general an order-M spline has continuous derivatives up to order $M-2$, which also fits here. Furthermore the so-called natural cubic spline has restrictions on the end-points; they have to be linear i.e. $a_2 = a_3 = 0$. Thus preventing large deviations near the end-points. [Hastie et al., 2009, p.141-144]

The R-command for making b-splines is named *bs* and it fits a cubic spline with 3 degrees of freedom per default. It is possible to add knots, which simultaneously adds degrees of freedom. We will add two knots representing the two points in time equivalent to the FDA warning in USA in 2004 and the latest guideline issued from the Danish Health and Medicines Authority in 2007, respectively.

Year	1. quarter	2. quarter	3. quarter	4. quarter
1995	3,500 (23.7)	3,610 (24.1)	3,722 (24.4)	3,917 (25.3)
1996	4,003 (25.5)	4,018 (25.6)	4,092 (25.8)	4,280 (26.4)
1997	4,125 (25.0)	4,232 (25.7)	4,139 (24.8)	4,332 (25.4)
1998	4,336 (24.7)	4,442 (24.8)	4,565 (24.9)	4,675 (24.7)
1999	4,789 (24.5)	4,775 (24.3)	4,910 (24.5)	4,999 (24.2)
2000	4,985 (23.3)	4,959 (23.0)	5,000 (22.7)	5,336 (23.6)
2001	5,490 (23.6)	5,559 (23.6)	5,603 (23.4)	5,882 (23.9)
2002	5,952 (23.6)	6,084 (23.7)	6,099 (23.3)	6,343 (23.6)
2003	6,405 (23.2)	6,515 (23.3)	6,607 (23.1)	6,882 (23.5)
2004	7,027 (23.4)	6,519 (21.6)	6,321 (20.7)	6,525 (21.0)
2005	6,466 (20.5)	6,484 (20.4)	6,376 (19.9)	6,323 (19.5)
2006	6,411 (19.3)	6,479 (19.5)	6,330 (18.9)	6,435 (19.0)
2007	6,431 (18.7)	6,407 (18.6)	6,363 (18.5)	6,368 (18.3)
2008	6,254 (17.9)	6,134 (17.6)	6,132 (17.5)	6,227 (17.5)
2009	6,175 (17.1)	6,095 (17.0)	6,001 (16.5)	6,131 (16.7)
2010	5,991 (16.2)	5,876 (16.0)	5,946 (16.0)	5,987 (16.0)
2011	5,955 (15.9)	5,893 (15.9)	5,794 (15.6)	5,832 (15.7)
2012	5,812 (15.6)	5,624 (15.2)	5,502 (15.0)	5,608 (15.3)

Table 3.1: Number of patients using antipsychotics for each quarter during the 18 years.

Figure 3.2 illustrates a cubic spline (green line) with two knots in ξ_1 and ξ_2. The columns in the design matrix defines the curve that is the spline. The basis representing such a spline consists of six equations as follows.

$$\begin{aligned} h_1(X) &= 1, \\ h_2(X) &= X, \\ h_3(X) &= X^2, \\ h_4(X) &= X^3, \\ h_5(X) &= (X-\xi_1)^3_+, \\ h_5(X) &= (X-\xi_2)^3_+. \end{aligned}$$

The b-spline basis is different and allows for fast and efficient computations, also when the number of knots is large. [Hastie et al., 2009, p.144]

From the basis we can write the form for a spline with two knots as follows:

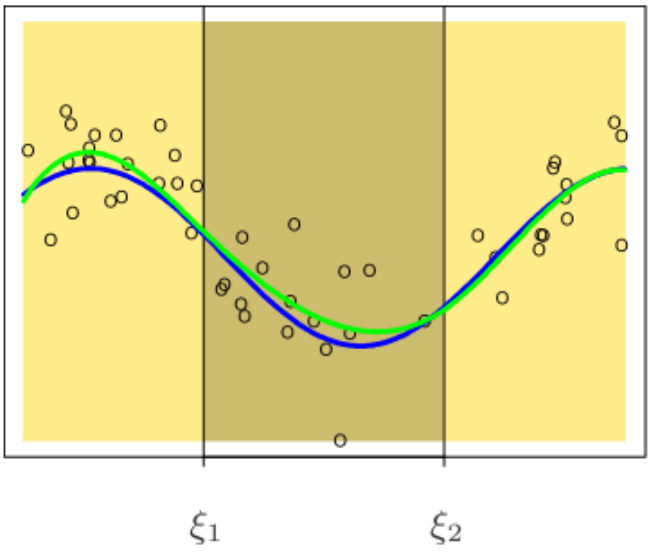

Figure 3.2: Cubic spline with two knots

$$f_0(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \tag{3.1}$$

for $0 < x < \xi_1$,

$$f_1(x) = f_0(x) + b_1(x - \xi_1)^3 \tag{3.2}$$

for $\xi_1 < x < \xi_2$ and

$$f_2(x) = f_1(x) + b_2(x - \xi_2)^3 \tag{3.3}$$

for $\xi_2 < x \leq T$.

It is clear that in total R should estimate 6 parameters; a_0 (the intercept), a_1, a_2, a_3, b_1 and b_2. The *bs* function in R provides exactly 5 estimates + the intercept.

Here we take a look at the derivatives and the limits with the purpose of checking the criteria regarding continuity and smoothness. In the first knot in $x = \xi_1$, we see that the limit for f_1 when x approaches the knot, is in fact f_0:

$$\lim_{x \downarrow \xi_1} f_1(x) = f_0(x).$$

We determine the first derivative of f_1,

$$f_1'(x) = f_0'(x) + 3b_1(x - \xi_1)^2,$$

and look at its limit when x approaches the knot;

$$\lim_{x \downarrow \xi_1} f_1'(x) = f_0'(x),$$

exactly as it should be.

Now we write the second derivative of f_1 and its limit;

$$f_1''(x) = f_0''(x) + 6b_1(x - \xi_1),$$

and

$$\lim_{x \downarrow \xi_1} f_1''(x) = f_0''(x),$$

again, exactly as it should be.

It is clear that the exact same calculations are valid for the second knot.

When making the splines R uses the method of least squares. It is a common method and is all about minimizing the expression:

$$\frac{1}{2\sigma^2} (y - \mathbf{X}\beta)^T W (y - \mathbf{X}\beta), \tag{3.4}$$

where y is the observed mean DDD and $\mathbf{X}$ represents how time affects the use of antipsychotics. The variance for the jth observation is $\frac{\sigma^2}{n_j}$, where n_j is equal to the number of patients using medication for the j'th quarter. Hence we obtain the weight matrix W as diagonal with $\{n_j\}_{j=1}^{72}$. Thus we are able to allow for and neutralise the n's involved in the variance otherwise.

An important feature when making splines for the mean DDD, is a test to determine, if a spline with two knots in $x = 37$ and $x = 53$ fits significantly better than a spline without knots. For that purpose we will use the *anova*-command, which provides an analysis of variance and compare two models directly.

In chapter 4 we first and foremost construct an overall cubic spline for the mean DDD. We use the anova to check, if the spline fits significant better with two knots or without knots. After that, we divide the population in different ways and make the same splines and tests. The 3 groups of sub populations are:

- By gender: females vs. males
- By diagnosis type: psychiatric vs. somatic vs. prescription
- By time of diagnosis: diagnosed before the guideline in 2007 vs. diagnosed after 2007

Here the R-command *drop1* is useful. It is used to compare the overall model with the model resulting from removing terms of maximal order. In this case that would be the interactions between time (the spline) and the relevant factor. In this study we use it to decide, whether or not it is a reasonable assumption that two or more splines for sub populations are parallel. In the output table the *Pr(F)*, which is a measure of how likely the corresponding F-value is, indicates exactly that. Hence, when it is close to zero it tells us, that the term is in some way important to include in a good model, as a model without that term is significantly worse.

The design matrix from the overall spline is used as the time factor in the random intercept model, which is described in the following section and fitted in chapter 5.

3.3 The Random Intercept Model

In this section the linear mixed effects model with random intercept used in chapter 5 is introduced and outlined on matrix form.

The point of origin for the model that we would like to fit, is the linear mixed effects model with random intercept. We let y_{ij} represent the observation on patient i made at time t_j. Here $j = \{1, \ldots, 72\}$ and $i = \{1, \ldots, 149, 869\}$. The model for the observation y_{ij} is

$$y_{ij} = \beta_0 + s_j \cdot \beta_1 + u_i + \epsilon_{ij}, \tag{3.5}$$

where s_j is the j'th row from the design matrix $\mathbf{S}(t_j)$ from the fitted spline with two knots from the previous section. The random intercept denoted u_i is patient-specific and constant over time. The residual ϵ_{ij} varies randomly over time. Added together these two components corresponds to the residual in a typical linear regression model. [Everitt and Hothorn, 2010, p.217]

In the model it is assumed that:

- u_i are normally distributed with zero mean and variance σ_u^2
- ϵ_{ij} are normally distributed with zero mean and variance σ^2
- u_i and ϵ_{ij} are independent of each other and of the time t_j

In R the *lmer* command fits the data to the model i expression 3.5. It uses the unique *pnr*'s to separate the patients, and treat the observations as repeated

measurements. The response variable is the DDD variable, and for the time factor we used the design matrix from the overall spline of the mean DDD. We have tried different optimizers for the *lmer* command, but since they did not improve the results, we used the command with the use of optimizers.

When constructing the model we will add the following explanatory variable as fixed effects:

- Gender (kqn2)
- Diagnosis type; inpsyk, insom or prescpt (diag_how)
- Age when diagnosed (diag_age)
- Years since diagnosis (time_sin_diag)
- Diagnosed before or after the year 2007 (diag_after)

As these are considered fixed effects, we can evaluate their estimates fairly simple by looking at two things:
1) If the t-value is outside ± 2, the estimate is considered significant, and if it is outside ± 3 it is very significant. Because of the large amount of patients, it is likely to obtain significant estimates.
2) The estimate. How small or large is it compared to the intercept, and is it positive or negative according to our assumption. For example with the diagnosed before or after variable, we hope to see a negative estimate, because that would mean, that patients diagnosed after the year 2007 uses less medication than those diagnosed before the guideline.

We end up with this written on vector form:

$$\begin{aligned} \mathbf{Y} \;=\; & \beta_{\mathbf{0}} + \mathbf{S} \cdot \beta_{\mathbf{1}} + \text{ kqn2 } \cdot \beta_{\mathbf{2}} + \text{ diag\_how } \cdot \beta_{\mathbf{3}} \\ & + \text{ diag\_age } \cdot \beta_{\mathbf{4}} + \text{ time\_sin\_diag } \cdot \beta_{\mathbf{5}} + \text{ diag\_after } \cdot \beta_{\mathbf{6}} \\ & + \mathbf{U} + \epsilon. \end{aligned} \tag{3.6}$$

The model described in equation 3.6 is called Model 1.

3.4 Model Diagnostic

We check the model by making qq-plots; one for the residuals for the random intercept and one for all the residuals. A qq-plot illustrates normally distributed residuals, when the points form a straight line.

If the model does not seems to fit, we will try a log-transformation of the mean DDD, and then check if that model fits better. We call it Model 2, and it is defined as

$$\begin{aligned} \log(\mathbf{Y}) &= \beta_{\mathbf{0}} + \mathbf{S} \cdot \beta_{\mathbf{1}} + \text{ kqn2 } \cdot \beta_{\mathbf{2}} + \text{ diag\_how } \cdot \beta_{\mathbf{3}} \\ &\quad + \text{ diag\_age } \cdot \beta_{\mathbf{4}} + \text{ time\_sin\_diag } \cdot \beta_{\mathbf{5}} + \text{ diag\_after } \cdot \beta_{\mathbf{6}} \\ &\quad + \mathbf{U} + \epsilon. \end{aligned} \tag{3.7}$$

If an estimate is not significant, the variable is excluded from the caldulations and the model is refitted. Thus we end up with a final model named Model 3. Lastly, diagnostics with both qq-plots and histograms for the residuals are made on Model 3.

Now with the models described, we move on to the actual analysis. The splines for the mean DDD are performed in chapter 4, while fitting the random intercept model takes place in chapter 5.

CHAPTER 4

MEAN DDD AND SPLINES

This chapter focuses on the mean DDD and the analysis regarding it. These analysis involves both simple plots with 95% confidence interval, cubic splines with two knots and without knots, and anova tests to compare the different splines. After introducing the plots of the mean DDD and making two splines for the whole population, the 3 following sections are made according to the 3 groups of sub populations listed in section 3.2.

Every calculated spline in this chapter uses weights equal to the number of patients using medication each quarter. Every plot contains both two vertical green lines and two horizontal lines. As mentioned earlier the two green lines correspond to the two points in time, where FDA and the Danish Health and Medicines Authority issued a warningguideline, respectively. The two horizontal lines corresponds to the two limits for medication-use; the recommended limit of 7 DDD and the limit consistent with death (21 DDD), respectively.

4.1 Plots of the Mean DDD

Figure 4.1 illustrates the mean number of DDDs per quarter from the first quarter of year 1995 to the fourth quarter of year 2012. It is clear that the average amount of antipsychotics used exceeds the recommended limit with several DDD. Furthermore the guidelines issued in 2007 has not helped, on the contrary the mean DDD has increased from 2007 to 2012.

Appendix A contains plots of the mean DDD for different sub-populations. The

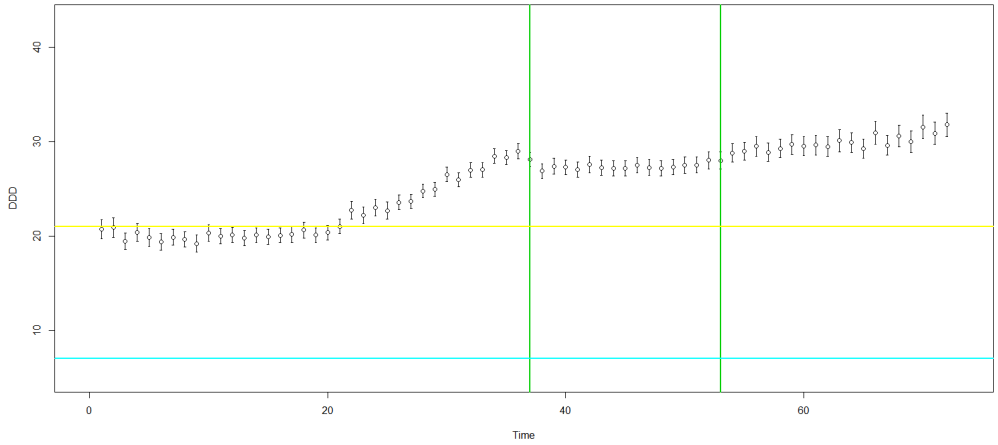

Figure 4.1: Mean DDD with 95% CI over time.

progress over time looks familiar for both genders; the mean number of DDD is increasing in the whole period with a spike around the 32-35th quarter. In general the first two plots look like the overall plot, except the underlying level is higher for those diagnosed in psychiatric care than for those diagnosed in somatic care.

As a supplement to the figures in appendix A table 4.1 was constructed. It contains information about the number of observations and the number of patients using antipsychotics in each of the 5 sub-populations.

Sub population	Patients*	Prevalence (%)	Observations
Whole Population	56,218	37.5	401,526
Females	34,655	37.1	267,125
Males	21,563	38.3	134,401
Diagnosis type:			
Psychiatric	26,373	49.0	215,029
Somatic	26,450	31.0	162,563
Prescription	3,395	31.8	23,934
Diagnosis time:			
Before guideline	45,963	42.6	355,804
After guideline	10,255	24.4	45,722

Table 4.1: Overview of the different sub-populations used to calculate means DDD. *using antipsychotics

It is distinctive that nearly half of the patients diagnosed in psychiatric care uses antipsychotics, while it is a third of those diagnosed in somatic care or by prescription. Even though the prevalence rates in table 4.1 are not corrected for age effects, they are consistent with the new study, which states that prevalence has decreased in the last years [Nørgaard et al., 2016].

4.2 Making the Splines

From the graphs in figure 4.1, we see that a cubic spline seems appropriate. Maybe with two knots in $x = 37$ and $x = 53$. The first knot represents that time in point, where the FDA in the USA issued a warning against this drug for patients with dementia. On the plot of the mean DDD we can also see a little bump around that time (20042005). The second knot is because we hope to see a change in the pattern around the time of the guideline from the Danish Health and Medicines Authorities (2007).

First we fit the spline with two knots and obtain the results printed in table 4.2.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.4403	0.4501	47.64	0.0000
spl1	-6.9460	0.9551	-7.27	0.0000
spl2	9.1651	0.5644	16.24	0.0000
spl3	4.7496	0.7180	6.62	0.0000
spl4	10.0322	0.6470	15.51	0.0000
spl5	9.4476	0.6560	14.40	0.0000

Table 4.2: Table of results obtained from fitting a cubic spline with two knots ($x = 37$ and $x = 53$) to the mean DDD.

Then we fit the same spline, but this time without knots. The results are listed in table 4.3. The spline with the knots has two more estimates, since it has two more degrees of freedom than the spline without knots.

The next graph, figure 4.2, shows the mean DDD and the belonging cubic splines. The spline with two knots is colored blue, while the red curve is the spline without knots. We can see that the blue curve follows the points more precisely than the red curve, indicating that the spline with two knots fits significantly better. It also looks like the use of antipsychotics is increasing a little after the guideline in year 2007.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	18.4945	0.6193	29.86	0.0000
spl01	4.3217	1.6796	2.57	0.0123
spl02	12.1478	0.9905	12.26	0.0000
spl03	11.8035	0.9001	13.11	0.0000

Table 4.3: Table of results obtained from fitting a cubic spline without knots to the mean DDD.

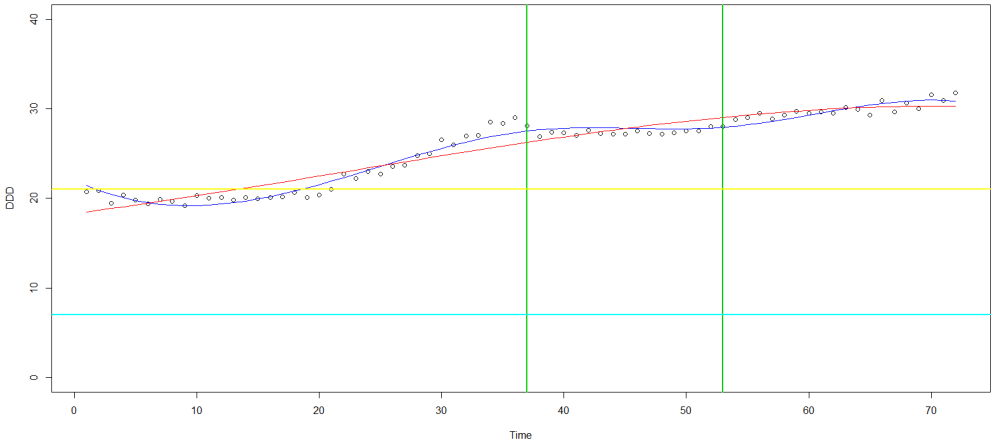

Figure 4.2: The mean DDD plotted with the belonging splines; with two knots (blue) and without knots (red).

We perform an anova to compare the two splines and decide, if the blue spline actually fits significantly better. The results from the anova are shown in table 4.4, where model 1 is the spline without knots and model 2 is the spline with two knots. Since Pr(>F) is close to zero, the spline with two knots fits significantly better than the spline without knots. This could be an indication, that the two points in time marking the FDA warning in USA and the guideline from The Danish Health and Medicines Authority each has some kind of influence on the use of antipsychotics in elderly patients with dementia. This is because when adding a knot to the spline, one indicates that this point in time could represent a change in the response variable.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	68	524530.96				
2	66	178646.05	2	345884.92	63.89	0.0000

Table 4.4: Table of results from the anova.

In the random intercept model in chapter 5 we shall use this spline with two knots for describing the time factor.

In the next 3 sections we will make splines and comparisons within the 3 groups of sub populations listed in section 3.2 in chapter 3.

4.3 According to Gender

In this section we look at the population divided into two groups; females and males. Table 4.5 contains the estimates for the fitted spline with two knots. We see that the differences in the splines are not significant, since the p-values are not close to zero. Therefore we use the drop1 test and the result is printed in table 4.6.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.0913	0.4297	49.08	0.0000
SM	1.0615	0.7450	1.42	0.1565
Spl1	-7.1589	0.9108	-7.86	0.0000
Spl2	8.7906	0.5373	16.36	0.0000
Spl3	4.5753	0.6873	6.66	0.0000
Spl4	10.2569	0.6226	16.48	0.0000
Spl5	9.0924	0.6328	14.37	0.0000
SM:Spl1	0.5908	1.5830	0.37	0.7096
SM:Spl2	1.3094	0.9372	1.40	0.1647
SM:Spl3	0.3988	1.1856	0.34	0.7371
SM:Spl4	-0.6976	1.0636	-0.66	0.5130
SM:Spl5	0.8873	1.0762	0.82	0.4112

Table 4.5: Table of the results obtained from fitting a cubic spline with two knots (in $x = 37$ and $x = 53$) to the mean DDD separated according to gender.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			217343.03	1078.00		
S:Spl	5	12173.26	229516.29	1075.84	1.48	0.2010

Table 4.6: The results from the drop1 test of the spline fitted to the mean DDD separated according to gender.

Since Pr(>F) is not close to zero, we conclude, that it is reasonable to assume, that the spline does not interact with gender. This means, we consider the two

splines for females and males, respectively, to be parallel, and the new fitted model is shown in table 4.7. We can see that every variable except one is significant, and now the estimate gender variable is also significant. This estimate states that males in general uses 1.5 DDD more than females. Of course this result should should be thought of with consideration, since we have not adjusted for age or anything else.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	20.9178	0.3570	58.60	0.0000
SM	1.5784	0.1369	11.53	0.0000
Spl1	-6.9539	0.7514	-9.25	0.0000
Spl2	9.2109	0.4440	20.74	0.0000
Spl3	4.7207	0.5649	8.36	0.0000
Spl4	9.9947	0.5090	19.63	0.0000
Spl5	9.3988	0.5161	18.21	0.0000

Table 4.7: Table of results obtained from fitting a cubic spline with two knots to the mean DDD separated according to gender.

The fitted splines from the model in table 4.7 are illustrated in figure 4.3 along with the mean DDD. The black dots and the black line are for the males, while the red dots and red line represents the females. The splines are parallel as the drop1 test showed. The development of the drug use does not seem to be decreasing after the year 2007 as hoped. On the contrary, it is increasing over time.

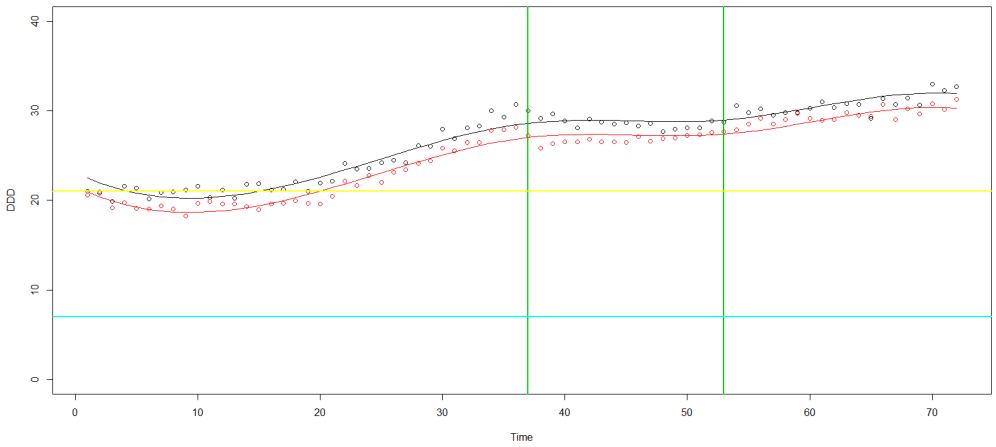

Figure 4.3: Mean DDD with fitted spline with two knots for females (red) and males (black) respectively.

The fitted spline without knots is shown in table 4.8. Now we use an anova to compare the two models. The result from this anova test is displayed in table 4.9, where model 1 is the spline without knots and model 2 is the spline with two knots.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.9624	0.4608	38.98	0.0000
SM	1.5672	0.2158	7.26	0.0000
Spl01	4.3866	1.2337	3.56	0.0005
Spl02	12.1559	0.7276	16.71	0.0000
Spl03	11.7642	0.6612	17.79	0.0000

Table 4.8: Table of results obtained from fitting a cubic spline without knots to the mean DDD separated according to gender.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	139	578519.91				
2	137	229516.29	2	349003.62	104.16	0.0000

Table 4.9: Results from the anova made to compare the spline without knots with the one with two knots.

Since Pr(>F) is close to zero, we conclude, that the model with the knots fits significantly better than the model without knots.

4.4 According to Diagnosis Type

In this section the study population is divided into three groups according to how the patient was diagnosed; in psychiatric care, in somatic care or by prescription. We start with fitting the usual spline with two knots, and the resulting estimates are written in table 4.10.

Then we perform the drop1 test, which is shown in table 4.11. Since Pr(>F) is very close to zero, it is clear, that there is some kind of interaction between the spline and diagnosis type. Hence we can not assume, that they are parallel.

We plot all three splines with the corresponding observations in figure 4.4. The data and model for patients diagnosed in psychiatric care have the color black, while those diagnosed in somatic care are red. The blue spline and dots represents the patients diagnosed by prescription.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.1855	0.4320	58.30	0.0000
SPR	-10.7853	5.3622	-2.01	0.0457
SS	-7.5448	0.6231	-12.11	0.0000
Spl1	-8.6421	0.9023	-9.58	0.0000
Spl2	6.4164	0.5319	12.06	0.0000
Spl3	3.6663	0.6881	5.33	0.0000
Spl4	8.3036	0.6241	13.31	0.0000
Spl5	8.1862	0.6404	12.78	0.0000
SPR:Spl1	22.1248	7.6288	2.90	0.0042
SS:Spl1	2.2705	1.3362	1.70	0.0909
SPR:Spl2	12.2718	4.8761	2.52	0.0127
SS:Spl2	4.2885	0.7990	5.37	0.0000
SPR:Spl3	2.9003	5.9425	0.49	0.6261
SS:Spl3	2.2665	1.0184	2.23	0.0272
SPR:Spl4	3.5643	5.3035	0.67	0.5024
SS:Spl4	4.5518	0.9273	4.91	0.0000
SPR:Spl5	2.0438	5.5261	0.37	0.7119
SS:Spl5	3.8776	0.9389	4.13	0.0001

Table 4.10: Table of results from the fitted spline with two knots for the mean DDD separated according to diagnosis type.

When investigating these graphs we see, that maybe two of the splines could be assumed to be parallel; the black and the red one. Therefore we perform yet another drop1 test involving only two of the splines. The results from the new fitted model is placed in table 4.12, while the result obtained from the drop1 test is in table 4.13.

Even though in figure 4.3 the black curve seems to almost be parallel with the red curve, the drop1 test concludes that is not the case. The interaction between the spline and the diagnosis type is significant. Hence the two splines cannot be assumed to be parallel.

For each of the three sub-populations we have fitted two splines; one with two knots and one without knots. The two splines have been plotted against each other and the mean DDD, and an anova test has been performed. Common to all three anovas is, that the spline with two knots fits significantly better than the model without knots. All tables from these calculations and the belonging plots are placed in appendix B.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			240853.31	1485.13		
S:Spl	10	467753.48	708606.79	1686.35	36.32	0.0000

Table 4.11: Table of results from the drop1 test when separated according to diagnosis type.

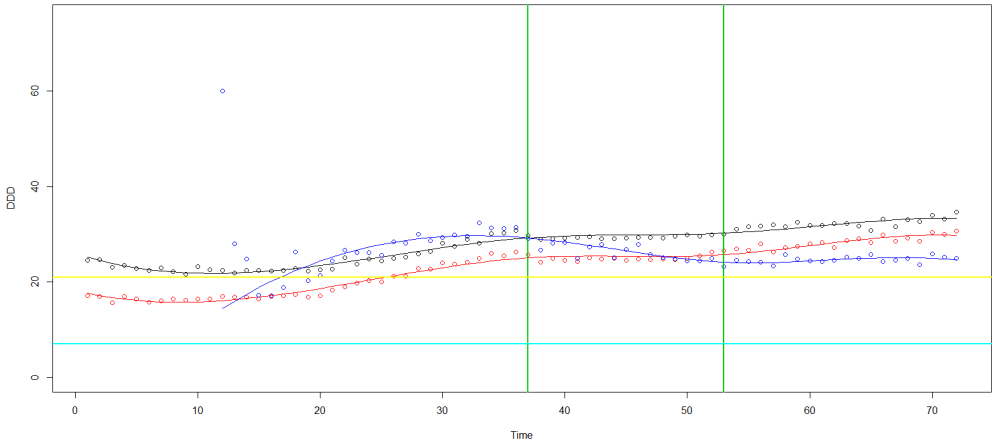

Figure 4.4: Cubic splines with two knots for the 3 diagnosis types: inpsyk:black, insom:red, prescpt:blue.

4.5 According to Time of Diagnosis

In this section we separate the patients diagnosed before January 1st 2008 from those diagnosed after. When dividing the population with regard to the time of a patients diagnosis, we of course obtain a group of patients, who use antipsychotics only after the time of the guideline in the year 2007. Therefore it is not relevant to model a spline with the usual two knots. Hence we fit only splines without knots, and the resulting model is printes in table 4.14.

Again we use the drop1 test to determine, if the two splines can be assumed parallel. The result is printed in table 4.15. Since Pr(>F) is very close to zero, some kind of interaction between time (the spline) and time of diagnosis exist.

We plot the splines and the corresponding observations in 4.5, where the group of patients diagnosed before the guideline are black, while the group diagnosed after are red.

This whole chapter has focused on aggregated data and has not adjusted for any other variables, but it is more appropriate to treat the observations as repeated

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.1855	0.4801	52.46	0.0000
S2S	-7.5448	0.6924	-10.90	0.0000
Spl21	-8.6421	1.0027	-8.62	0.0000
Spl22	6.4164	0.5911	10.86	0.0000
Spl23	3.6663	0.7646	4.79	0.0000
Spl24	8.3036	0.6935	11.97	0.0000
Spl25	8.1862	0.7116	11.50	0.0000
S2S:Spl21	2.2705	1.4849	1.53	0.1286
S2S:Spl22	4.2885	0.8879	4.83	0.0000
S2S:Spl23	2.2665	1.1317	2.00	0.0473
S2S:Spl24	4.5518	1.0305	4.42	0.0000
S2S:Spl25	3.8776	1.0434	3.72	0.0003

Table 4.12: Table of results from the fitted spline including only inpsyk and insom.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			209966.75	1073.02		
S2:Spl2	5	74650.35	284617.10	1106.83	9.39	0.0000

Table 4.13: Table of results for the second drop1 test when separated according to diagnosis type.

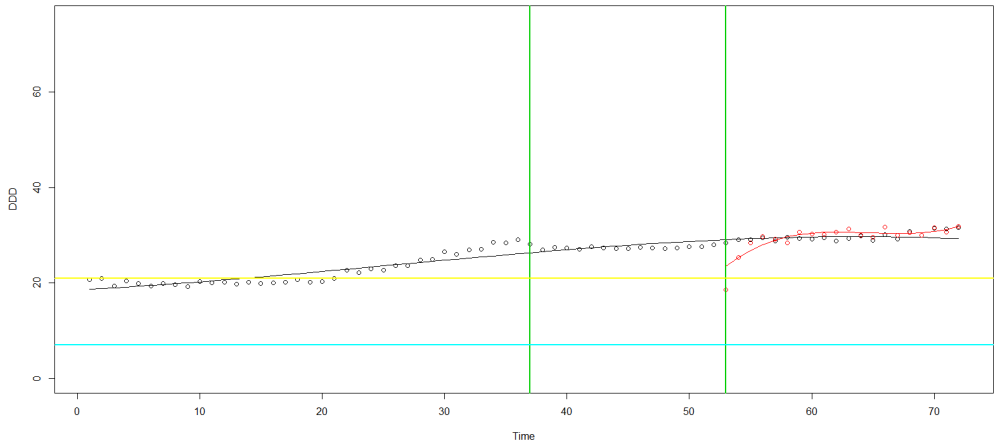

Figure 4.5: Cubic splines without knots for the 2 diagnosis times: before:black and after:red.

measurements and analyze them using the theory of longitudinal data, since that kind

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	23.5466	2.6362	8.93	0.0000
SBG	-4.8651	2.6962	-1.80	0.0748
Spl1	12.6551	5.9244	2.14	0.0356
Spl2	3.5351	2.6843	1.32	0.1914
Spl3	8.3227	3.0241	2.75	0.0073
SBG:Spl1	-9.3031	6.1342	-1.52	0.1331
SBG:Spl2	9.6968	2.8740	3.37	0.0011
SBG:Spl3	2.2583	3.1848	0.71	0.4802

Table 4.14: Table of results for the fitted spline without knots for the mean DDD when separated according to time of diagnosis.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			523302.89	811.44		
S:Spl	3	205893.50	729196.39	835.97	11.02	0.0000

Table 4.15: Table of results from the drop1 test when separated according til time of diagnosis.

of analysis allows the repeated observations to depend on each other as described in [Ipsen, 2015, p. 3]. That is exactly what we will do in the following chapter, chapter 5.

CHAPTER 5

THE RANDOM INTERCEPT MODEL

The focus of this chapter is analysis using a linear mixed effects model with random intercepts as presented in chapter 3; Model 1 described in equation 3.6, and Model 2 plus Model 3 from equation 3.7. First we calculate the estimates according to Model 1, and then we look at the qq-plots for the residuals. Since they do not look too good, we use Model 2 and calculate the estimates again. Variables with insignificant effect are excluded from the model, and then the data are fitted again, called Model 3. The qq-plots for Model 3 looks much better, and all variables now have significant effect. Lastly, we check the histograms and see that the assumptions about the model are met.

5.1 Model 1

The results from the estimation are shown in table 5.1. Since all the t-values are outside ± 3, all the estimates are considered very significant. The gender factor is negative, so women use more medication than men. However it is only 1 DDD, so the real effect is arguable. The same goes for the estimate for the diag_after variable; it is positive, which means, that patients diagnosed after the guideline in 2007 uses more medication than patients diagnosed before this point in time. It is the opposite effect, than we hoped for. Patients drug use should be decreasing in line with the concerns from the Danish Health and Medicines Authority. But since it is only 1.7 DDD, the effect is arguable, and can maybe be considered irrelevant.

The negative estimates for the factor describing diagnosis type (in psychiatric

care, in somatic care or by prescription) tells that patients diagnosed in psychiatric care uses more medication than patients diagnosed in somatic care. Those diagnosed in somatic care uses more medication than those diagnosed by prescription. This was expected, since the symptoms of those diagnosed by prescription are not so bad, that they had to go to the hospital. Their GP is able to handle them and their symptoms, it seems.

	Estimate	Std. Error	t value
(Intercept)	20.12	0.34	59.42
Time1	-4.72	0.56	-8.42
Time2	7.08	0.37	19.13
Time3	4.25	0.47	9.00
Time4	6.56	0.50	13.19
Time5	7.08	0.52	13.71
factor(kqn2)2	-1.04	0.22	-4.68
diag_after	1.60	0.38	4.26
factor(diag_how)2	-2.95	0.22	-13.46
factor(diag_how)3	-5.14	0.46	-11.07
I(diag_age - 80)	-0.66	0.01	-46.91
time_sin_diag	-0.31	0.03	-11.72

Table 5.1: Table of results for Model 1

The last two estimates are negative, which means that the older a patient is when diagnosed, the less medication he uses, and the longer it has been since diagnosis, the less medication does the patient use. The effects are measured in DDD per year. Therefore the estimate -0.66 DDD per year means that a patient diagnosed at age 70 uses 6.6 DDD more than a patient diagnosed at age 80 year. Since the recommended limit is 7 DDD, this must be considered an important difference.

Furthermore the estimate -0.31 DDD per year means that a patient who had been diagnosed 20 years ago uses 3.1 DDD more than a patient who have been diagnosed 10 years ago. Again this is an important difference. Both effects should cause the Danish Health and Medicines Authority to take action, especially against patients diagnosed relatively young.

The control of Model 1 consist of two qq-plots; one for the residuals of random intercepts and one for the residuals. They are shown in figure 5.1. The closer the points are to form a line, the better the model fit. These qq-plots does not look

as nice as they could or should. Therefore we take the logarithm of the response variable, and fit the model again. This model is called Model 2, and the estimation is described in the following section.

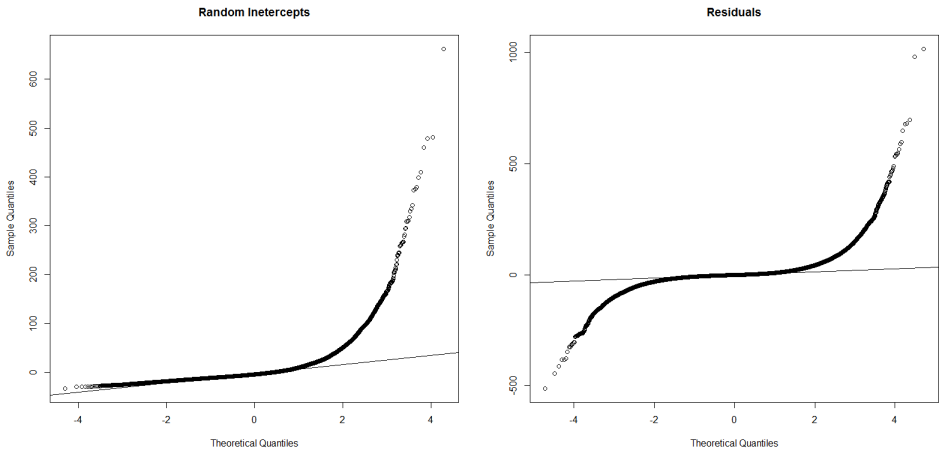

Figure 5.1: QQ-plots for model 1

5.2 Model 2

The results from the estimation are shown in table 5.2. Since it is a logarithmic model, the effects are measured on the logarithmic scale and are easily converted to percentage (per year). All the t-values except one are outside ± 3, meaning all the estimates except one are considered very significant. The estimate for the variable named diag_after is not significant, which means that the patients diagnosed after the guideline was issued in 2007 does not use less or more antipsychotic medication than those diagnosed before the guideline.

According to these results, men in general uses

$$\exp(-0.05) = 0.95 \tag{5.1}$$

5% less antipsychotics than women, while those diagnosed in somatic care use

$$\exp(-0.13) = 0.88 \tag{5.2}$$

12% less medication than those diagnosed in psychiatric care, and those diagnosed by prescription use

$$\exp(-0.15) = 0.86 \tag{5.3}$$

	Estimate	Std. Error	t value
(Intercept)	2.50	0.01	248.76
Time1	-0.16	0.02	-10.11
Time2	0.30	0.01	28.46
Time3	0.20	0.01	14.66
Time4	0.31	0.02	20.41
Time5	0.32	0.02	20.12
factor(kqn2)2	-0.05	0.01	-6.76
diag_after	-0.01	0.01	-0.81
factor(diag_how)2	-0.13	0.01	-18.27
factor(diag_how)3	-0.15	0.01	-9.95
I(diag_age - 80)	-0.02	0.00	-41.22
time_sin_diag	-0.01	0.00	-10.79

Table 5.2: Table of results for Model 2

14% less medication than those diagnosed in psychiatric care.

The results from Model 1 recur; the older a patient is at the time of the diagnosis and the longer it has been since time of diagnosis, the less medication does the patients use. The estimate -0.02 means that a patient diagnosed at age 80 uses ,

$$\exp(-0.2) = 0.82, \tag{5.4}$$

18% less than a patient diagnosed at age 70 year. Likewise the estimate -0.01 means that a patient who had been diagnosed 20 years ago uses

$$\exp(-0.1) = 0.90 \tag{5.5}$$

10% less than a patient who have been diagnosed 10 years ago.

5.3 Model 3

Now we fit the same model again, but without the variable diag_after, since it is not significant. The results are shown in table 5.3.

According to Model 3 all significant estimates from the previous model are the same. This means that men use 5% less antipsychotics than women, while those diagnosed in somatic care use 12% less medication than those diagnosed in psychiatric care, and those diagnosed by prescription use 14% less medication than those diagnosed in psychiatric care.

	Estimate	Std. Error	t value
(Intercept)	2.50	0.01	248.98
Time1	-0.16	0.02	-10.16
Time2	0.30	0.01	28.87
Time3	0.20	0.01	15.29
Time4	0.30	0.01	23.08
Time5	0.31	0.01	23.14
factor(kqn2)2	-0.05	0.01	-6.75
factor(diag_how)2	-0.13	0.01	-18.27
factor(diag_how)3	-0.15	0.01	-9.94
I(diag_age - 80)	-0.02	0.00	-41.24
time_sin_diag	-0.01	0.00	-12.32

Table 5.3: Table of results for Model 3

Again we see that the older a patient is at the time of the diagnosis and the longer it has been since time of diagnosis, the less medication does the patients use. The estimate -0.02 means that a patient diagnosed at age 80 uses, 18% less than a patient diagnosed at age 70 year. Likewise the estimate -0.01 means that a patient who had been diagnosed 20 years ago uses 10% less than a patient who have been diagnosed 10 years ago.

The qq-plots for the log-model are plotted in figure 5.2. It is clear that these residuals lies much closer to form a line than those from Model 1, meaning Model 3 fits better.

The corresponding histograms are plotted in figure 5.3. The residuals are nicely normally distributed with zero mean. The residuals for the random intercepts looks bimodal, which indicates an undiscovered binary variable dividing the patients in two groups; one with a mean just below zero, and another one with a mean just above zero. Maybe this has something to do with the fact, that some of the patients in the population might have another (and earlier) psychiatric diagnosis requiring antipsychotics, fx schizophrenia. This will be investigated further in chapter 6.

When we have adjusted for all other factors such as age, diagnosis type, age when diagnosed and so on, we end up with the time effect, which is plotted in figure 5.4. It is a reference spline for an 80 year old woman, who was just diagnosed in psychiatric care. The increase in DDD over time is visually large, and from the y-axis we can conclude, that the increase in percent is also significant; 35% from 1995 to 2012.

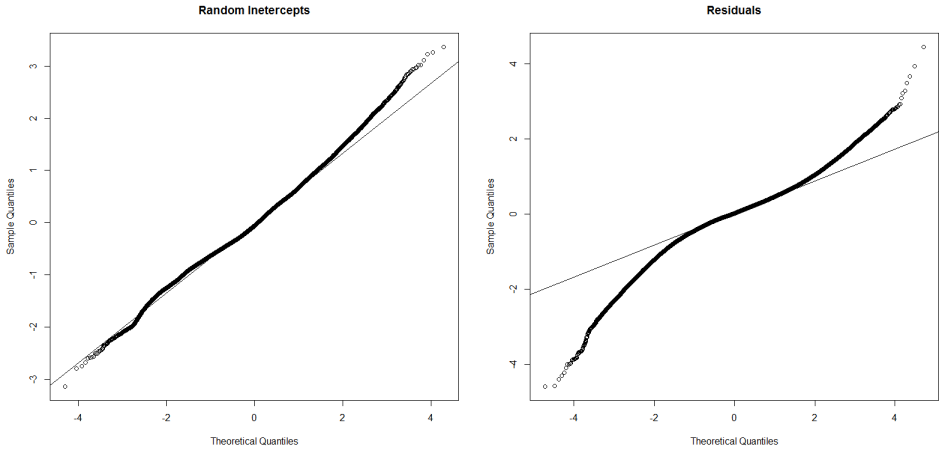

Figure 5.2: qq-plots for Model 2

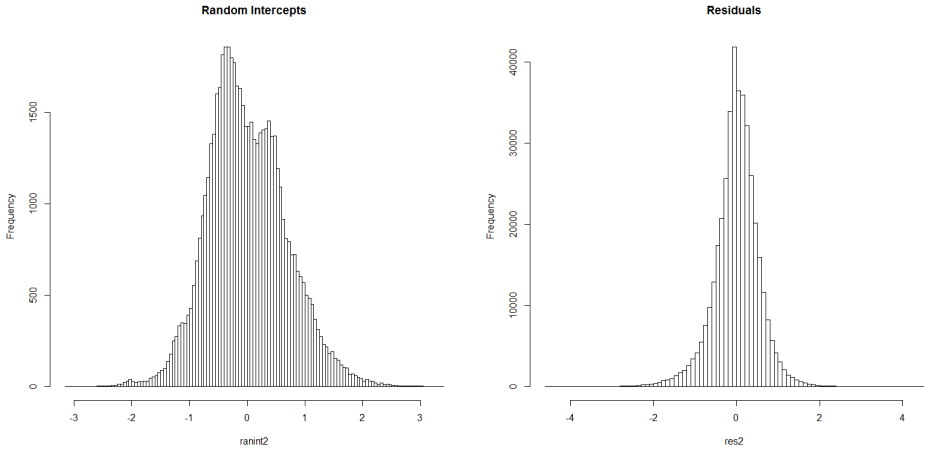

Figure 5.3: Histograms for Model 3

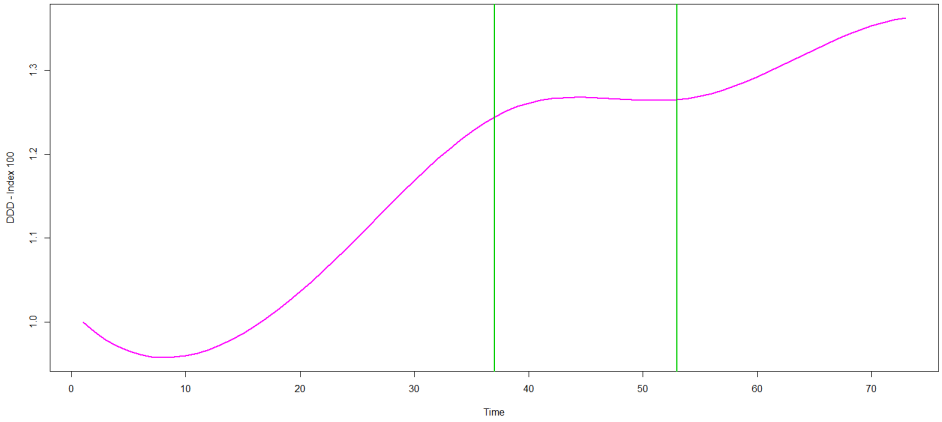

Figure 5.4: Change in DDD in percent over time when adjusted for all other variables.

CHAPTER 6

PREVIOUS USE OF ANTIPSYCHOTICS

As discussed earlier we might deal with a group of patients using antipsychotics for other reasons than symptoms of dementia. Therefore it would be optimal to avoid the patients with a previous diagnosis requiring these kind of drugs. But since it is not easy to determine exactly which diagnosis that is, we create a new binary variable coded 1 if the patient filed a prescription of antipsychotics before he got the dementia diagnosis, and 0 otherwise. Both the mean DDD and the linear mixed effects model with random intercept will be investigated with this new variable.

If the variable has a large effect it is logical to exclude all patient who filed a prescription for antipsychotic medication before their dementia diagnosis and perfrom the analysis on this restricted population.

6.1 Previous Druguse as a Variable

In this section a new binary variable, prev_druguse, is constructed, and the same linear mixed effects with random intercepts are modeled yet again. The new variable is set to 1 if the patient did fulfill a prescription of antipsychotics before they got the dementia diagnosis, and 0 if the patient did not receive any antipsychotics before the diagnosis.

As in chapter 4 we start with fitting the usual cubic splines with two knots, and the results obtained are printed in table 6.1. From the table we can see, that most of the differences in the splines are significant, meaning we can probably not consider the two splines to be parallel. Still a drop1 test is made, these results er shown in

table 6.2. Since Pr(>F) is close to zero, some kind of interaction exist between the spline and previous drug-use.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	19.3814	0.9887	19.60	0.0000
SU	2.3140	1.0598	2.18	0.0308
Spl1	-3.3726	1.7558	-1.92	0.0569
Spl2	12.8726	0.9106	14.14	0.0000
Spl3	13.2773	1.2716	10.44	0.0000
Spl4	20.1723	1.0674	18.90	0.0000
Spl5	21.2844	1.1476	18.55	0.0000
SU:Spl1	-3.8353	1.9602	-1.96	0.0525
SU:Spl2	-5.6509	1.0664	-5.30	0.0000
SU:Spl3	-14.2501	1.4444	-9.87	0.0000
SU:Spl4	-18.1589	1.2471	-14.56	0.0000
SU:Spl5	-20.7803	1.3127	-15.83	0.0000

Table 6.1: Table of results obtained from fitting a cubic spline with two knots to the mean DDD when divided according to previous drug-use.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			220143.16	1079.84		
S:Spl	5	2810256.83	3030399.99	1447.43	337.01	0.0000

Table 6.2: Table of results for the drop1 test when separated according to previous druguse.

The splines are plotted in figure 6.1. The black graph represents the patients with a previous use of antipsychotics, while the red graph is for those patients without a previous use of antipsychotics. It is clear that the two spline not can be considered as parallel.

6.1.1 Model 3

As it became clear in chapter 5 the variable diag_after has no effect, so this will be left out in the following analysis. Furthermore, based on the analysis in the previous chapter, we carry out only the log-analysis, that is Model 3.

The results obtained in Model 3 is printed in table 6.3. Since all t-values lie outside ± 3, the estimates are considered significant. The estimate for time since diagnosis

is -0.004953 and the standard error is 0.0007, but R rounds off the numbers, so in the table they appear as zeros. But since

$$\exp(-0.005) = 0.995,$$

the effect of time since diagnosis is 0.5% per year, which corresponds to 5% per 10 years. This effect is significant statistically speaking, and also clinically relevant.

	Estimate	Std. Error	t value
(Intercept)	2.44	0.01	237.05
Time1	-0.18	0.02	-11.53
Time2	0.27	0.01	25.71
Time3	0.15	0.01	11.66
Time4	0.25	0.01	18.73
Time5	0.25	0.01	18.68
factor(kqn2)2	-0.04	0.01	-5.55
factor(diag_how)2	-0.10	0.01	-14.08
factor(diag_how)3	-0.14	0.01	-9.07
I(diag_age - 80)	-0.02	0.00	-42.21
time_sin_diag	-0.00	0.00	-7.00
prev_druguse	0.19	0.01	25.75

Table 6.3: Table of results for Model 3

The estimates regarding age when diagnosed, how diagnosed and gender are all similar to the results in the previous chapter in table 5.3. Since

$$\exp(0.19) = 1.21,$$

the effect of previous druguse is 21%, which means that patients with a previous use of antipsychotics uses 21% more medication than other patients. This must be considered a large effect both statistically and clinically speaking.

Before moving on we check Model 3 by looking at the relevant qq-plots. They are printed in figure 6.2, and looks relatively nice.

In figure 6.3 the corresponding histograms are plotted. The residuals look normally distributed with zero mean, just as described in the assumptions for the model. The residuals for the random intercepts look bimodal again, which means that the population is still divided into two groups; one with a mean just below zero and another one with a mean just above zero. So even after we have adjusted for previous

drug-use, there still is a binary variable that we have not accounted for. Maybe it could be the presence of a spouse or partner.

When adjusted for all other factors; gender, diagnosis type, age when diagnosed, time since diagnosis, and previous druguse, we end up with the pure time effect, which is plotted in figure 6.4. It is a reference spline for an 80 year old woman, who was just diagnosed in psychiatric care, and who does not have a previous use of antipsychotics. The shape is very similar to the reference spline printed in the previous chapter in figure 5.4, but the level lies a little lower; the use of antipsychotics in 2012 is approximately 28% higher than in 1995.

6.2 Restricted Population

In this section all patients with a use of antipsychotics before their dementia diagnosis are excluded from the population, thus we are only considering patients with (hopefully) no other diagnosis than dementia as a reason to use antipsychotics.

A group of $32,463$ patients from the original study population are excluded thus we end up with a restricted study population including $117,406$ patients. Only $20,564$ of the excluded patients used medication after their dementia diagnosis as well, leaving $35,654$ patients using antipsychotics in this restricted population. Their characteristics are shown in table 6.4, and are similar to the original study population. The characteristic has not changed a whole lot, except for the grouping using vs. not using antipsychotics; we now have more non-users than before.

The barchart in figure 6.5 is illustrating how many of the patients with dementia are using antipsychotics also looks similar to the corresponding one for the original study population.

Figure 6.6 shows a plot of the mean DDD with corresponding 95% confidence intervals. As mentioned earlier there are $35,654$ patients using antipsychotics in this restricted population. Table 6.5 presents how those patients spread out in the different sub-populations along with the prevalence and number of observations for each sub-population.

We can see that the prevalence is much higher among the patients diagnosed before the guideline compared to those diagnosed after.

The mean DDD with 95% confidence intervals for the restricted study population is printed in figure 6.6.

In the following analysis the variable diag_after is included again. This is because we still hope to see that this variable has an effect on the use of antipsychtics on

	Total		Females		Males	
	n	%	n	%	n	%
Gender	117,406	100	72,416	61.7	44,990	38.3
Age when diagnosed:						
60 – 69 years	11,659	9.9	5,418	46.5	6,241	53.5
70 – 79 years	37,763	32.2	21,273	56.3	16,490	43.7
80 – 89 years	56,106	47.8	36,755	65.5	19,351	34.5
90+ years	11,878	10.1	8,970	75.5	2,908	24.5
Not using antipsychotics	81,752	69.6	51,038	62.4	30,714	37.6
Using antipsychotics	35.654	30.4	21,378	60.0	14,276	40.0
Diagnosis type:						
Psychiatric	37,834	32.2	24,119	63.7	13,715	36.3
Somatic	71,117	60.6	43,539	61.2	27,578	38.8
Prescription	8,455	7.2	4,758	56.3	3,697	43.7
Diagnosis time:						
Before	84,211	71.7	52,501	62.3	31,710	37.7
After	33,195	28.3	19,915	60.0	13,280	40.0

Table 6.4: Overview of the restricted study population with regards to gender, age when diagnosed, if they use medication, and how they are diagnosed.

Sub population	Patients*	Prevalence (%)	Observations
Whole Population	35,654	30.4	232,958
Females	21,378	29.5	151,436
Males	14,276	31.7	81,522
Diagnosis type:			
Psychiatric	14,740	39.0	111,482
Somatic	18,903	26.6	109,443
Prescription	2,011	23.8	12,033
Diagnosis time:			
Before guideline	30,110	35.8	212,529
After guideline	5,544	16.7	20,429

Table 6.5: Overview of the different sub-populations used to calculate mean DDD for the restricted population. *using antipsychotics

the restricted study population, even though its effect was insignificant for the whole population.

The results from this Model 2 are shown in table 6.6. According to the t-values two of the estimates are not significant; diag_after and time_sin_diag, which suggests that the issuing of the guideline in 2007 and time since diagnosis do not have any effect on the use of antipsychotics. Therefore we perform a final model without these two variables.

	Estimate	Std. Error	t value
(Intercept)	2.45	0.01	218.74
Time1	-0.21	0.02	-11.55
Time2	0.21	0.01	16.43
Time3	0.11	0.02	6.82
Time4	0.18	0.02	10.17
Time5	0.18	0.02	10.11
factor(kqn2)2	-0.02	0.01	-2.85
diag_after	-0.02	0.02	-1.49
factor(diag_how)2	-0.08	0.01	-8.98
factor(diag_how)3	-0.08	0.02	-3.99
I(diag_age - 80)	-0.01	0.00	-24.86
time_sin_diag	0.00	0.00	0.60

Table 6.6: Table of results from Model 2 for the restricted population.

6.2.1 Model 3

The model is still a log-model, and it is named Model 3. Table 6.7 shows the results from Model 3. Compared to the estimates in table 5.3 we see that the estimates in general are slightly smaller. However all estimates have the same operational sign.

Since

$$\exp(-0.03) = 0.97,$$

the estimate -0.03 means that men use 3% less medication than women.

Furthermore since

$$\exp(-0.08) = 0.92,$$

patients diagnosed in somatic care or by prescription use 8% less medication than patients diagnosed in psychiatric care.

	Estimate	Std. Error	t value
(Intercept)	2.45	0.01	219.51
Time1	-0.21	0.02	-11.55
Time2	0.21	0.01	17.43
Time3	0.11	0.01	7.70
Time4	0.18	0.01	12.79
Time5	0.18	0.01	13.15
factor(kqn2)2	-0.03	0.01	-3.02
factor(diag_how)2	-0.08	0.01	-8.97
factor(diag_how)3	-0.08	0.02	-4.17
I(diag_age - 80)	-0.01	0.00	-26.20

Table 6.7: Results from Model 3 on the restricted population.

The estimate for diag_age means that patients use 10% less medication for each 10 years older when diagnosed.

The qq-plots for Model 3 are plotted in figure 6.7. As usual they look relatively nice.

The corresponding histograms er plotted in figure 6.8. While the histogram for the residuals looks fine, the one for the random intercepts looks bimodal. As mentioned earlier this is likely because of an undiscovered binary variable, which separate the population in two.

When interpreting the values of the estimates in Model 3, we have to remember transformations for the log-normal distribution. The exponential of the mean returns the median - not the mean. The formulas for the median and the mean are:

$$\begin{aligned} \text{Median} &: \exp(\mu) \\ \text{Mean} &: \exp(\mu + \frac{\sigma^2}{2}). \end{aligned}$$

[Madsen and Thyregod, 2011, p.280].

This means that according to Model 3 in table **??**, the use of antipsychotics in elderly people with dementia has the following values:

$$\begin{aligned} \text{Median} &: \exp(2.45) = 11.59 \\ \text{Mean} &: \exp(2.45 + \frac{0.81}{2}) = 17.37, \end{aligned}$$

because the variance is equal to the variance of the random intercepts plus the variance of the residuals: $0.49 + 0.32 = 0.81$.

In relation to the limit of 7 DDD we have that:

$$\begin{aligned} \text{Median} &: \frac{7}{11.59} = 0.60 \\ \text{Mean} &: \frac{7}{17.37} = 0.40. \end{aligned}$$

This translate to the conclusion, that the median use of antipsychotics is 60% of index 100, and that the mean use of antipsychotics is 40%.

When we have adjusted for all other factors such as age, diagnosis type, age when diagnosed and so on, we end up with the time effect, which is plotted in figure 6.9. It is a reference spline for an 80 year old woman from the restricted population, who was just diagnosed in psychiatric care.

The shape is very similar to the previous reference splines, but we can see on the y-axis that the general level is lower. For the restricted population we have that the use of antipsychotics in 2012 lies 20% higher than in 1995. This means, that the mean amount of antipsychotics used is three times higher than recommended by the Danish Health and Medicines Authority.

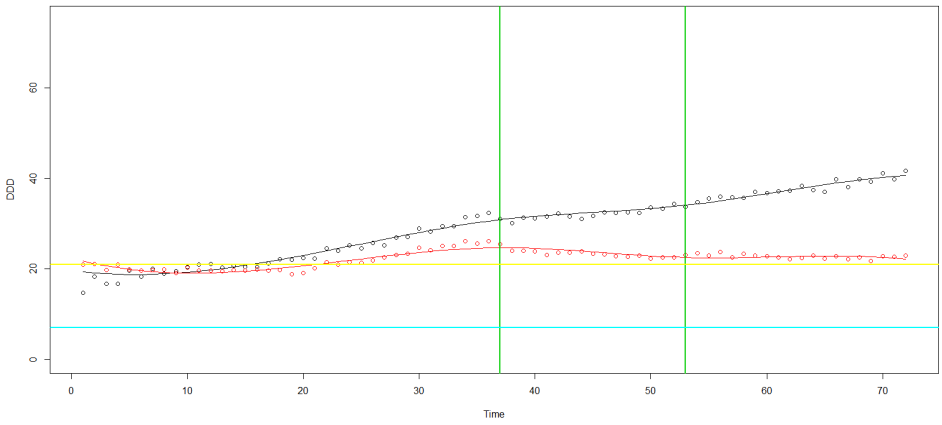

Figure 6.1: Splines for the mean DDD for patients with a previous druguse (black) and patients without a previous druguse (red).

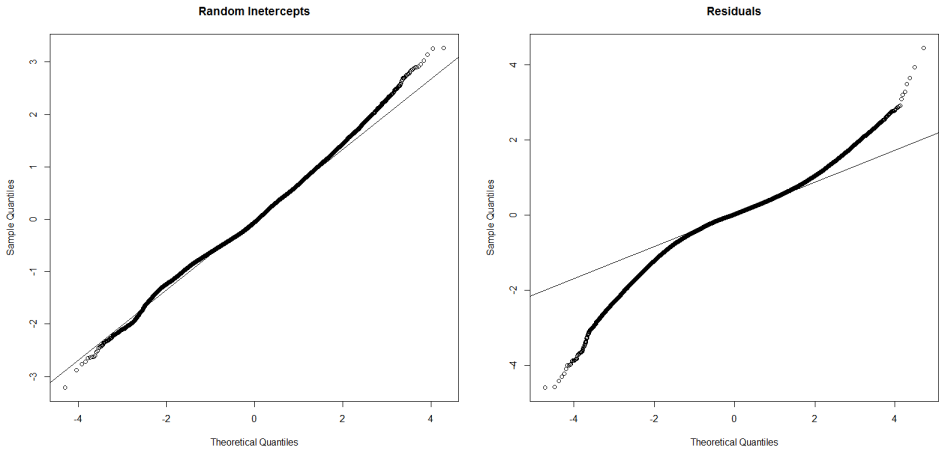

Figure 6.2: qq-plot for Model 3.

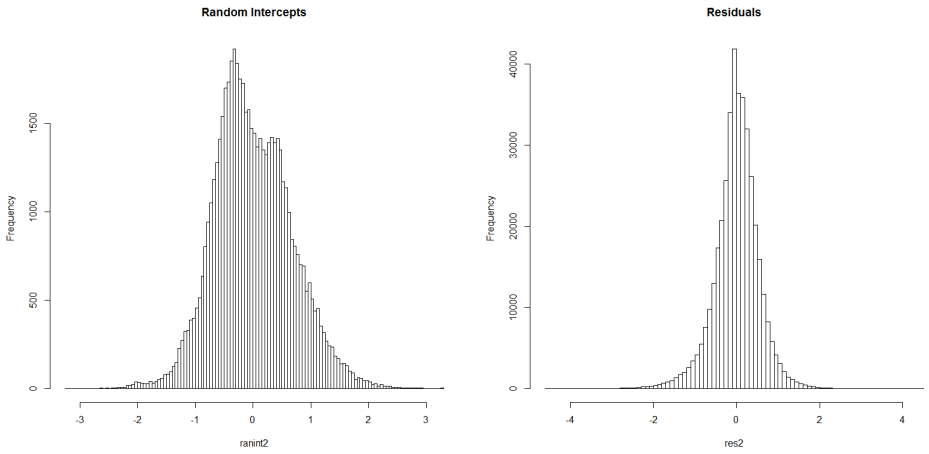

Figure 6.3: Histogram for Model 3.

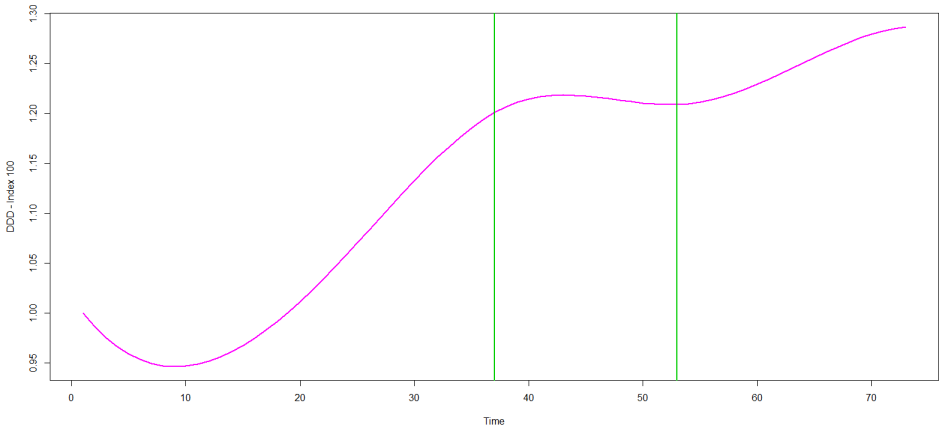

Figure 6.4: Change in DDD in percent over time.

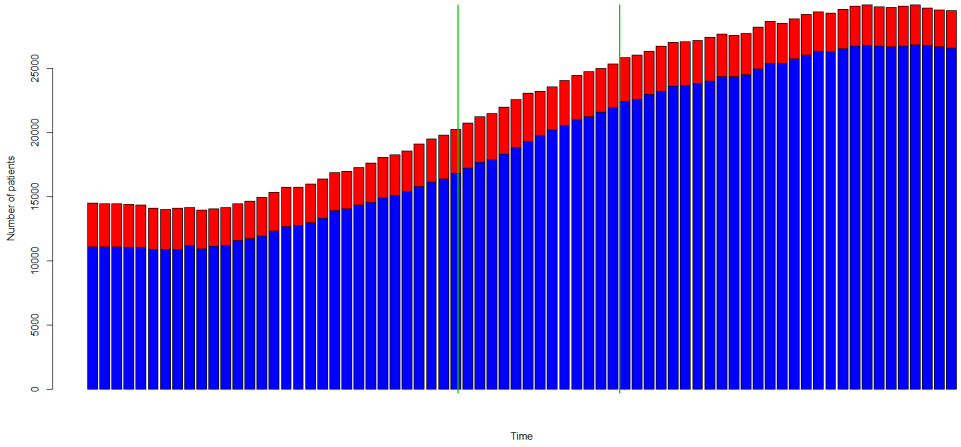

Figure 6.5: Barchart of the number of patients with dementia not using antipsychotics (blue) and using antipsychotics (red) over time from the first quarter of 1995 to the fourth quarter of 2012 for the restricted population.

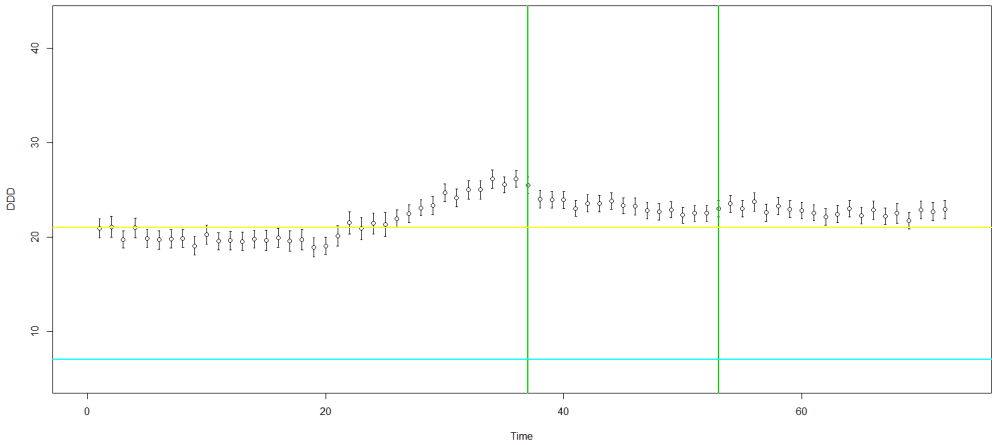

Figure 6.6: Mean DDD for the restricted population with 95% CI over time.

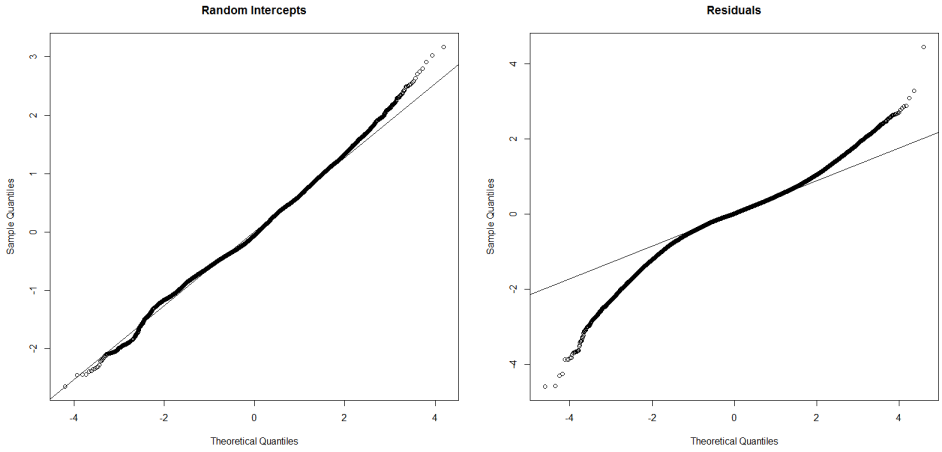

Figure 6.7: qq-plot for Model 3 for the restricted population.

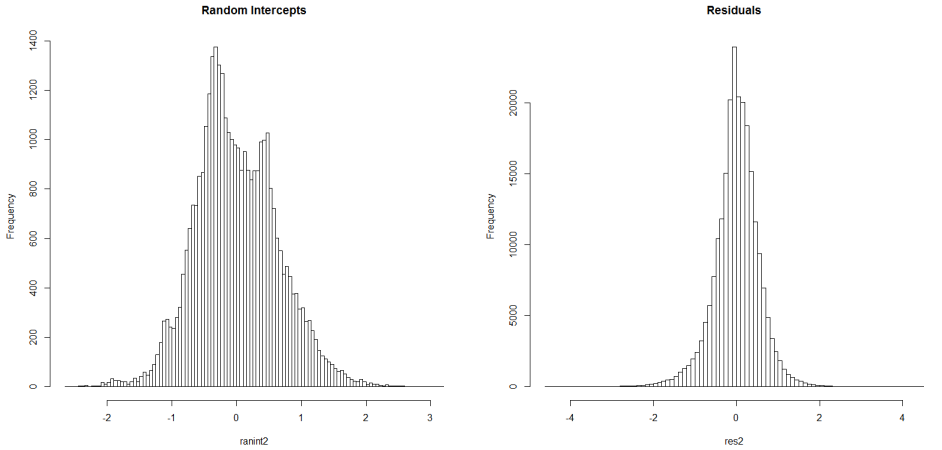

Figure 6.8: Histogram of the residuals from Model 3 on the restricted population.

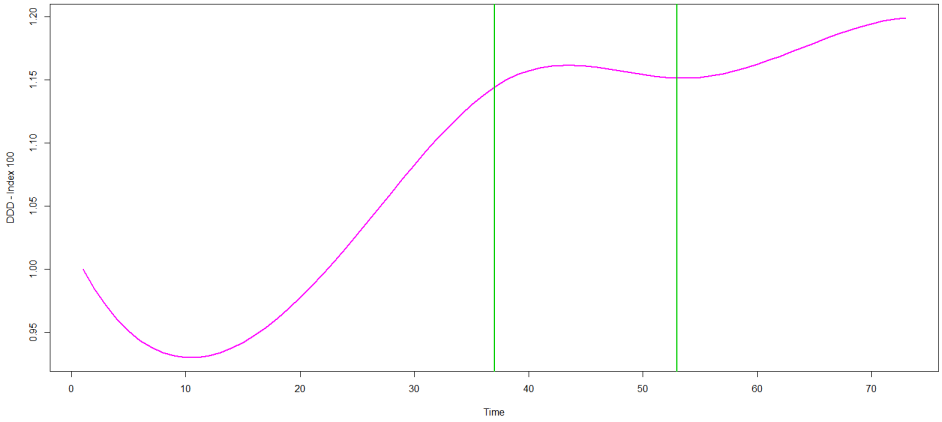

Figure 6.9: Percentwise change in the use of antipsychotics for the restricted population.

CHAPTER 7

CONCLUSION

This chapter includes a summary of the results obtained from the analysis performed in this thesis as well as a discussion of the weaknesses of the model. We have analyzed the use of antipsychotic medication in elderly patients with dementia using the theory of longitudinal data. The theory have proven useful in spite of the depressing results obtained.

We succeeded in fitting af linear mixed effects model with random intercepts, such allowing the repeated observations within a patient to be correlated, and each patient to have en individual baseline. We found significant estimates for the following variables: gender, diagnosis type, age when diagnosed, previous use of antipsychotics and time. The variable time since diagnosis was only significant for the whole population, hence insignificant for the restricted population. Only one variables proved to have no effect on medication-use regardless of study population; time of diagnosis (before or after guideline).

7.1 Results

Overall Denmark has a problem regarding elderly patients with dementia and their us of antipsychotic medication. Despite the authorities attempts to lower the amount of antipsychotics used by this especially vulnerable group of patients, it has almost only increased in the years from 1997 to 2012.

When looking at only those patients without a history of use of antipsychotics before the time of their dementia diagnosis, we can see that the mean amount of

medication used in 2012 is three time as high as the recommended limit, while the median amount used corresponds to twice the limit of 7 DDD. The corresponding relations when we do not adjust for previous drug-use are 3.6 and 2.3, respectively, which just confirms the observed fact, that it is important to take previous drug-use into account when modeling the use of antipsychotics.

We can conclude, that the use of antipsychotics has increased from index 115 to index 120 in the five years after the Danish Health and Medicines Authority issued a guideline including a clear maximum limit equivalent to index 40.

The variable for age when diagnosed had the opposite operational sign than expected. The model showed that patients uses 10% less medication for each 10 years older when diagnosed. Usually we think, that an older patient means a more severe degree of dementia symptoms, which would increase the use of antipsychotics. On the other hand one could assume, that the younger the patient, the more severe illness.

When it comes to type of diagnosis, it is natural that patients diagnosed in psychiatric care uses the most medication. This is probably the patients with such severe symptoms, and maybe other kind of psychiatric problems, that ends up in psychiatric care.

All in all the Danish Health and Medicines Authority ought to address the problem somehow, as their effort up until now has not changed the use of antipsychotics for the better, quite the contrary.

7.2 Discussion of the Model

One of the models weaknesses is that it does not include potential serial correlation. Next thing to do, if we have had more time, was to make and explorer a plot of the autocorrelations.

Furthermore, we have found inaccuracies suggesting an undiscovered binary variable. Further investigations should try to uncover this, maybe by including a variable describing whether or not the patient has a spouse or partner.

BIBLIOGRAPHY

Ballard, C. and Corbett, A. (2013). Agitation and aggression in people with Alzheimer's disease. *Current Opinion Psychiatry; 26: p. 252-259.*

Everitt, B. S. and Hothorn, T. (2010). *A Handbook of Statistical Analyses Using R.* Chapman & Hall, 2. edition.

Gulmann, N. C. (2006). Psykofarmaka til ældre med delir eller demens. *Rationel Farmakoterapi nr. 1.*

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning.* Springer, 2. edition.

Ipsen, S. (2015). Using the theory of longitudinal data to model the use of antipsychotics in elderly patients with dementia. http://github.com/SallyIpsen/ninja-turtles.

Larsen, D. (2016). Statistics Denmark. http://www.dst.dk/da/Statistik/emner/befolkning-og-befolkningsfremskrivning/folketal.

Madsen, H. and Thyregod, P. (2011). *Introduction to General and Generalized Linear Models.* CRC Press, 1. edition.

Nørgaard, A., Jensen-Dahm, C., Gasse, C., Hansen, H. V., and Waldemar, G. (2016). Time trend in Diagnosing Dementia in Secondary Care. *Journal of Alzheimer's Disease; 49: p. 211-220.*

NVD (2015). Forekomst og fordeling af demens. http://www.videnscenterfordemens.dk/viden-om-demens/demenssygdomme/forekomst-og-fordeling-af-demens/.

Phung, T. K. T., Andersen, B. B., Høgh, P., Kessing, L. V., Mortensen, P. B., and Waldemar, G. (2007a). Validity of Dementia Diagnoses in the Danish Hospital Registers. *Dementia and Geriatric Cognitive Disorders; 24: p. 220-228.*

Phung, T. K. T., Waltoft, B. L., Kessing, L. V., Mortensen, P. B., and Waldemar, G. (2007b). Time trend in Diagnosing Dementia in Secondary Care. *Dementia and Geriatric Cognitive Disorders; 29: p. 146-153.*

Socialministeriet (2010). Kortlægning af demensområdet i Danmark 2010. http://sim.dk/media/946402/17_kortl__gning-demensomr__det-dec2010.pdf.

Sundhedsstyrelsen (2005). Forbruget af anti-psykotiske lægemidler blandt ældre. http://www.sst.dk/~/media/2614FAA8CDC448559E686E07C80365D4.ashx.

Sundhedsstyrelsen (2006). *Klassifikation af Sygdomme Systematisk Del.* Munksgaard Danmark.

Sundhedsstyrelsen (2013). National klinisk retningslinje for udredning og behandling af demens. http://sundhedsstyrelsen.dk/~/media/C80BB3CA58574F7BB968B07F9AEDBB4B.ashx.

WHO (2003). Introduction to Drug Utilization Research. http://www.whocc.no/filearchive/publications/drug_utilization_research.pdf.

APPENDIX A

PLOTS OF THE MEAN DDD

This appendix contains the plots of the mean DDD with 95% confidence intervals.

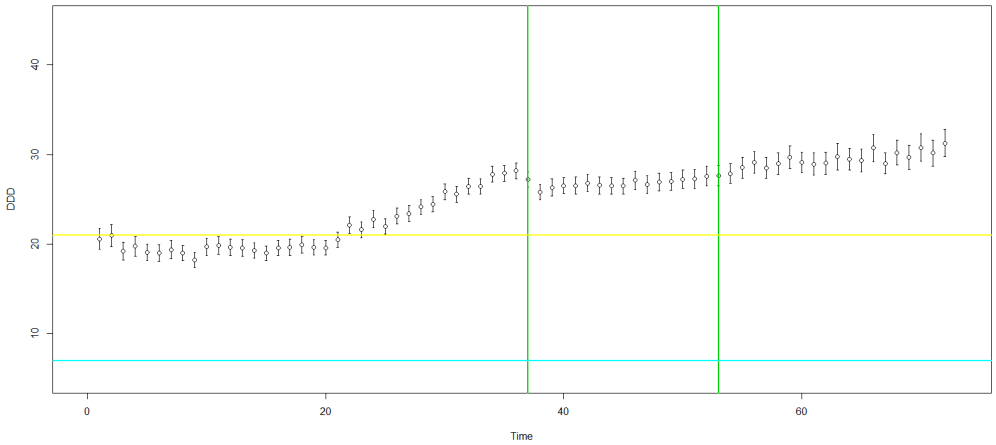

Figure A.1: Mean DDD for females with 95% CI over time.

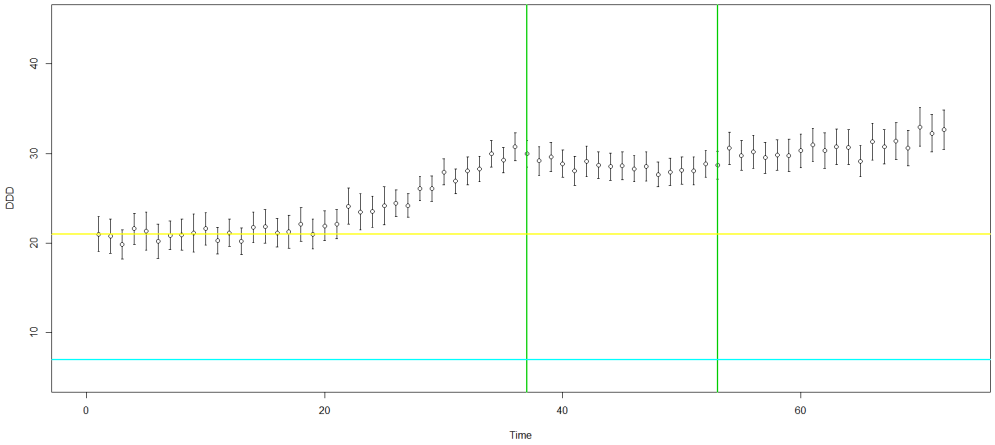

Figure A.2: Mean DDD for males with 95% CI over time.

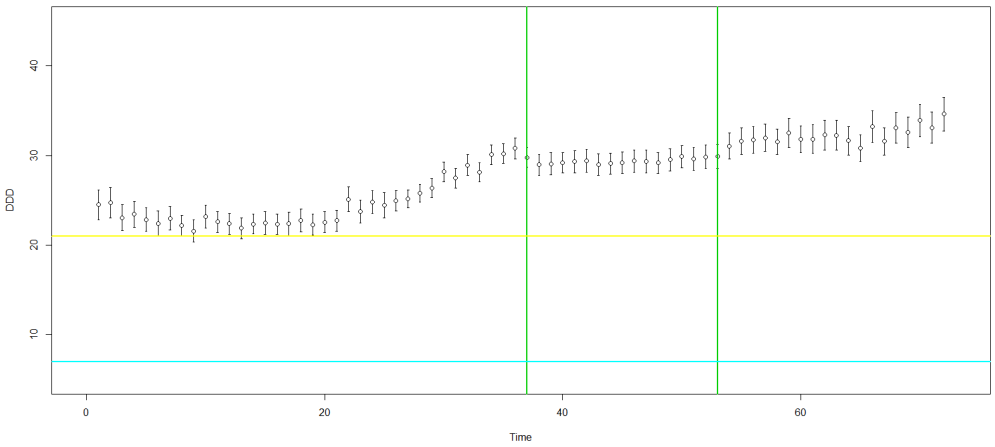

Figure A.3: Mean DDD for patients diagnosed in psychiatric care (with 95% CI) over time.

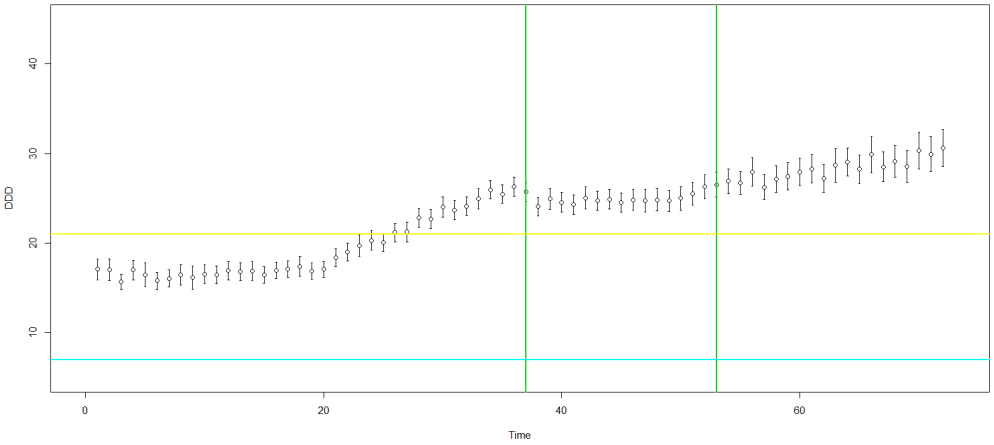

Figure A.4: Mean DDD for patients diagnosed in somatic care (with 95% CI) over time.

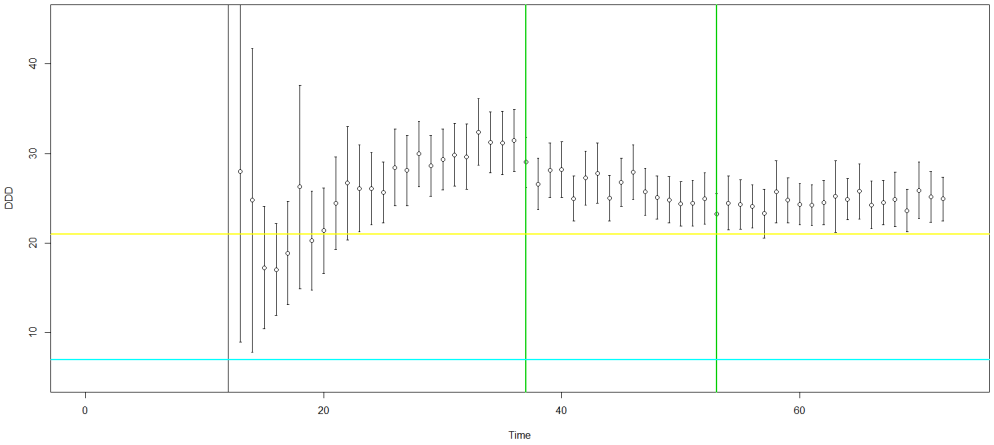

Figure A.5: Mean DDD for patients diagnosed by prescription (with 95% CI) over time.

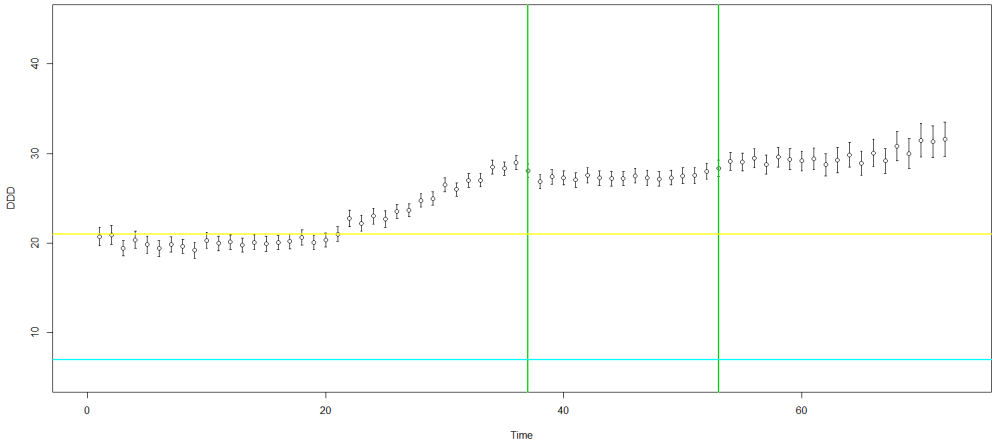

Figure A.6: Mean DDD for patients diagnosed before guideline (with 95% CI) over time.

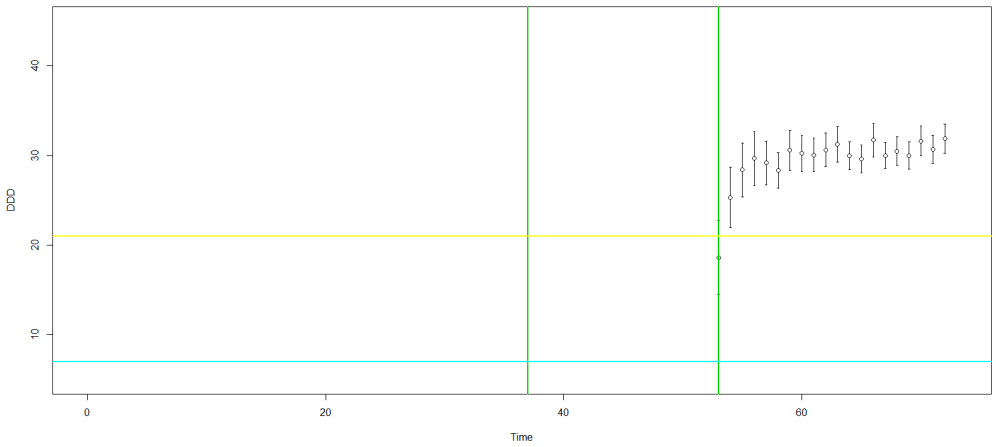

Figure A.7: Mean DDD for patients diagnosed after guideline (with 95% CI) over time.

APPENDIX B

SPLINES ACCORDING TO DIAGNOSIS TYPE

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.1855	0.5211	48.33	0.0000
spl1	-8.6421	1.0884	-7.94	0.0000
spl2	6.4164	0.6416	10.00	0.0000
spl3	3.6663	0.8300	4.42	0.0000
spl4	8.3036	0.7528	11.03	0.0000
spl5	8.1862	0.7725	10.60	0.0000

Table B.1: Fitted spline with two knots for the mean DDD for those diagnosed in psychiatric care.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.2973	0.6405	34.81	0.0000
spl01	-0.0143	1.7302	-0.01	0.9934
spl02	11.5271	1.0177	11.33	0.0000
spl03	10.1844	0.9436	10.79	0.0000

Table B.2: Fitted spline without knots for the mean DDD for those diagnosed in psychiatric care.

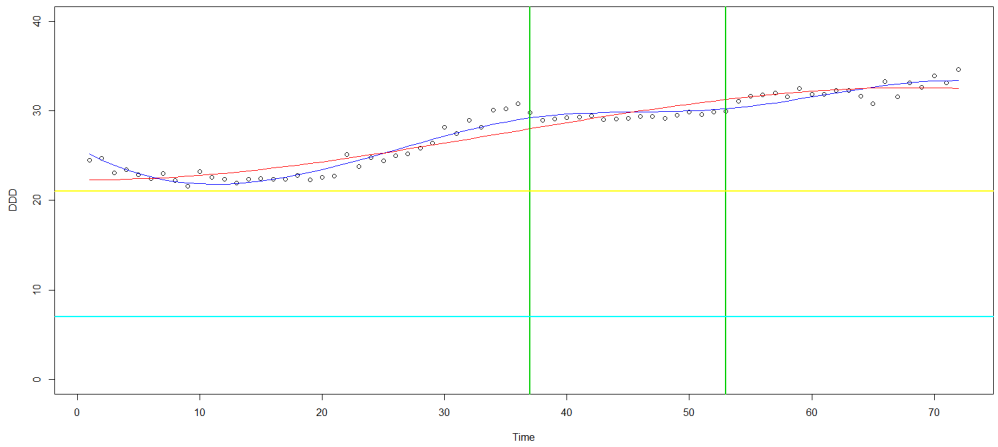

Figure B.1: Mean DDD and corresponding splines: with two knots (blue) and without knot (red) for patients diagnosed in psychiatric care.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	68	292579.30				
2	66	123690.78	2	168888.52	45.06	0.0000

Table B.3: Result of the anova comparing model 1: spline without knot and model 2: spline with two knots, for patients diagnosed in psychiatric care.

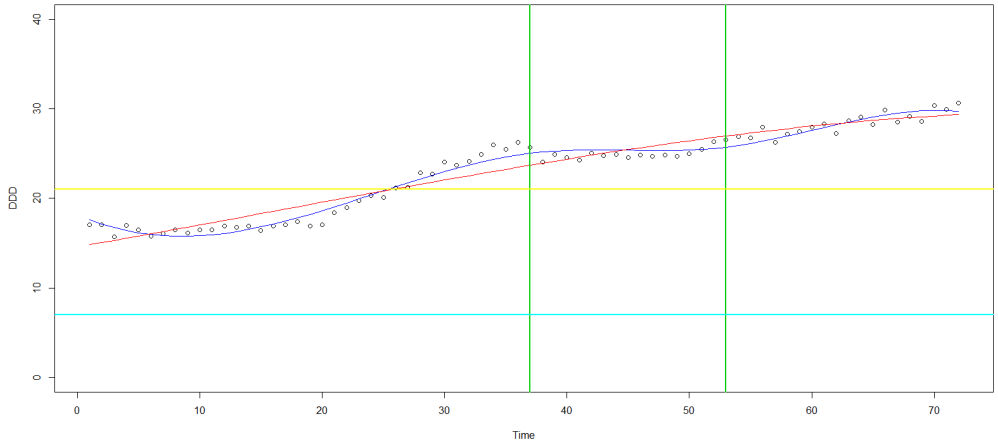

Figure B.2: Mean DDD and corresponding splines: with two knots (blue) and without knot (red) for patients diagnosed in somatic care.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.6407	0.4524	39.00	0.0000
spl1	-6.3716	0.9928	-6.42	0.0000
spl2	10.7050	0.6006	17.82	0.0000
spl3	5.9328	0.7563	7.84	0.0000
spl4	12.8554	0.6910	18.60	0.0000
spl5	12.0638	0.6918	17.44	0.0000

Table B.4: Fitted spline with two knots for the mean DDD for those diagnosed in somatic care.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	14.8357	0.6174	24.03	0.0000
spl01	5.6416	1.7272	3.27	0.0017
spl02	12.8744	1.0672	12.06	0.0000
spl03	14.5045	0.9287	15.62	0.0000

Table B.5: Fitted spline without knot for the mean DDD for those diagnosed in somatic care.

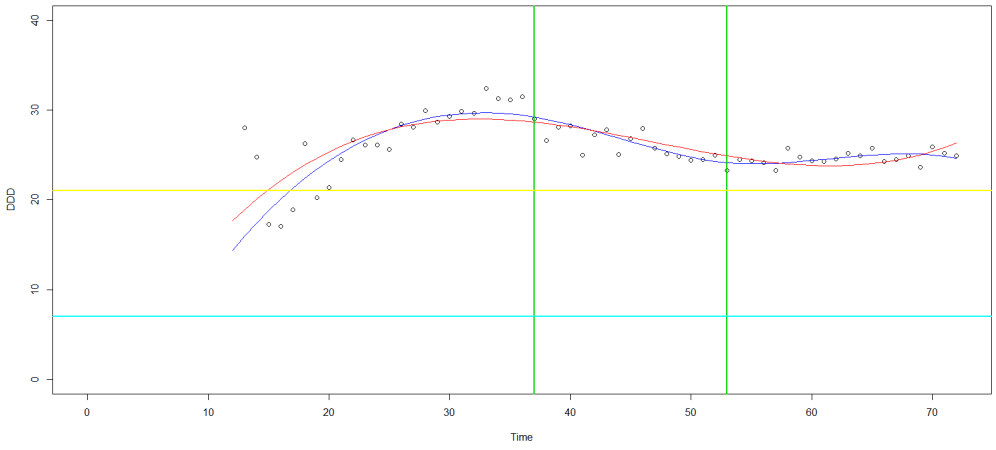

Figure B.3: Mean DDD and corresponding splines: with two knots (blue) and without knot (red) for patients diagnosed by prescription.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	68	244033.64				
2	66	86275.96	2	157757.67	60.34	0.0000

Table B.6: Result of the anova comparing model 1: spline without knot and model 2: spline with two knots, for patients diagnosed in somatic care.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	14.4003	3.5292	4.08	0.0001
spl_p1	13.4827	5.0020	2.70	0.0093
spl_p2	18.6882	3.2005	5.84	0.0000
spl_p3	6.5666	3.8975	1.68	0.0977
spl_p4	11.8680	3.4777	3.41	0.0012
spl_p5	10.2299	3.6243	2.82	0.0066

Table B.7: Fitted spline with two knots for the mean DDD for those diagnosed by prescription.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.6831	2.1830	8.10	0.0000
spl0_p1	25.7291	4.3255	5.95	0.0000
spl0_p2	-1.9047	1.7422	-1.09	0.2789
spl0_p3	8.6300	2.4119	3.58	0.0007

Table B.8: Fitted spline without knot for the mean DDD for those diagnosed by prescription.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	57	42475.82				
2	55	30886.56	2	11589.26	10.32	0.0002

Table B.9: Result of the anova comparing model 1: spline without knot and model 2: spline with two knots, for patients diagnosed by prescription.