



Universiteit Utrecht



Thesis for the Master of Science degree

# KNOWLEDGE DISCOVERY WITHIN THE DUTCH LONG-TERM CARE SECTOR

Robert Anthony Vroon

<i>1<sup>st</sup> supervisor:</i>	dr. Marco Spruit	<i>Utrecht University</i>
<i>2<sup>nd</sup> supervisor</i>	dr. Ronald Batenburg	<i>Utrecht University</i>
<i>Daily supervisor:</i>	Laura Tolsma	<i>Technology to Serve</i>



Name: Robert Anthony Vroon  
Email address: r.a.vroon@students.uu.nl  
Student no.: 3440516

Master programme: Master Business Informatics ( 2010-2012 )  
Department of Information & Computing Sciences  
Utrecht University  
Utrecht, The Netherlands



## PREFACE

This thesis is on “*Knowledge discovery within the Dutch long-term care sector*”. The research was conducted at Technology to Serve B.V., in cooperation with one care institution. Due to confidentiality reasons this care institution will remain anonymous.

In this preface I would like to thank all the people that supported, motivated and assisted me during my graduation period. First of all, I would like to thank my supervisors dr. Marco Spruit (Utrecht University) and dr. Ronald Batenburg (Utrecht University), for their support and advice during the whole research. I would also like to thank my supervisor at Technology to Serve, Laura Tolsma for the possibility to perform this research at Technology to Serve. Also for her efforts to arrange contact with care institutions and stakeholders.

Besides the supervisors, I would like to thank all the interviewees that have provided me with valuable insight and understanding in the Dutch long-term care sector. These interviewees enabled me to perform this research. In addition, I would like to thank the care institution that provided me with their operational data that enabled the data analysis part of this study.

Last, but certainly not least, I would like to thank my parents, family and friends for supporting me during my whole study.

Robert Anthony Vroon

Monday, August 20, 2012



# TABLE OF CONTENTS

PREFACE.....	5
TABLE OF CONTENTS .....	7
LIST OF FIGURES .....	9
LIST OF TABLES.....	10
GLOSSARY .....	11
1. INTRODUCTION.....	12
1.1. PROBLEM STATEMENT .....	12
1.2. RESEARCH QUESTIONS .....	13
1.3. RELEVANCE .....	14
1.3.1. Scientific.....	14
1.3.2. Social.....	15
2. THEORETICAL BACKGROUND.....	16
2.1. DUTCH LONG-TERM CARE SECTOR .....	16
2.1.1. Financing schemes.....	17
2.1.2. Stakeholders .....	18
2.1.3. Care process .....	20
2.2. KNOWLEDGE DISCOVERY PROCESSES .....	21
2.2.1. Knowledge Discovery in Databases (KDD).....	21
2.2.2. CRoss-Industry Standard Process for Data Mining (CRISP-DM).....	22
2.2.3. Sample, Explore, Modify, Model, Assess (SEMMA) .....	23
2.2.4. Conclusion.....	24
2.3. KNOWLEDGE DISCOVERY IN HEALTHCARE.....	25
2.4. DATA MINING.....	26
2.4.1. Data mining goals.....	26
2.4.2. Model functions.....	26
3. RESEARCH APPROACH .....	28
3.1. RESEARCH METHOD .....	28
3.2. ACTIVITIES.....	29
3.2.1. Business / Environment understanding .....	29
3.2.2. Data understanding .....	30
3.2.3. Data preparation .....	31
3.2.4. Modelling .....	31
3.2.5. Evaluation.....	31
3.2.6. Deployment .....	32
4. FIELD STUDY: KD IN LONG-TERM CARE SECTOR .....	33
4.1. BUSINESS UNDERSTANDING.....	33
4.1.1. Information needs.....	34
4.1.2. Data mining goals.....	41
4.2. DATA UNDERSTANDING .....	46
4.2.1. Explore & Describe data.....	46
4.2.2. Verify data quality .....	51
4.3. DATA PREPARATION & MODELLING .....	52
4.4. EVALUATION & VALIDATION .....	78
4.4.1. Evaluate results.....	78
4.4.2. Review process .....	78
4.4.3. Validation .....	79
5. CONCLUSION .....	81

6. LIMITATIONS .....	83
7. BIBLIOGRAPHY .....	84
APPENDIX I – INTERVIEW I.....	87
APPENDIX II – INTERVIEW II .....	88
APPENDIX III – INTERVIEW III .....	89
APPENDIX IV – INTERVIEW IV .....	90
APPENDIX V – INTERVIEW V .....	91
APPENDIX VI – INTERVIEW VI .....	92
APPENDIX VII – INTERVIEW VII .....	93
APPENDIX VIII – INTERVIEW VIII.....	94
APPENDIX IX – INTERVIEW IX .....	95
APPENDIX X – INTERVIEW X.....	96
APPENDIX XI – INTERVIEW XI .....	97
APPENDIX XII – INTERVIEW XII .....	99
APPENDIX XIII – INTERVIEW XIII.....	101
APPENDIX XIV – INTERVIEW XIV .....	103
APPENDIX XV – INTERVIEW XV .....	104
APPENDIX XVI – INTERVIEW XVI .....	105
APPENDIX XVII – INTERVIEW XVII.....	106
APPENDIX XVIII – INTERVIEW XVIII.....	108
APPENDIX XIX – INFORMATION NEEDS.....	110
APPENDIX XX – INFORMATION NEEDS BY SIZE .....	112
APPENDIX XXI – DATA MINING GOALS VS. INFORMATION NEEDS.....	114
APPENDIX XXII – ZZP PRODUCTS .....	116



## LIST OF FIGURES

Figure 1: Research method .....	28
Figure 2: Research activities.....	29
Figure 3: Treatment goals.....	48
Figure 4: Number of incidents per month.....	52
Figure 5: Number of incidents per hour.....	53
Figure 6: Number of incidents per hour per location.....	53
Figure 7: Number of incidents per day of week.....	54
Figure 8: Number of clients in care .....	58
Figure 9: Total number of clients per location.....	59
Figure 10: Number of intramural clients per location.....	60
Figure 11: Number of extramural clients per location .....	60
Figure 12: Forecast for the number of extramural clients (total) .....	61
Figure 13: Forecast for the number of extramural clients (blue location).....	62
Figure 14: Forecast for the number of extramural clients (black location).....	62
Figure 15: Needed FTE for care plan tasks .....	64
Figure 16: Average duration of stay (per location).....	65
Figure 17: Average duration a client is customer .....	65
Figure 18: Average duration of stay forecast.....	66
Figure 19: Average duration total forecast .....	66
Figure 20: Age distribution per status.....	67
Figure 21: Current ZZP-mix (delivered).....	69
Figure 22: ZZP-mix over time (delivered).....	70
Figure 23: Number of ZZP1 clients (delivered).....	71
Figure 24: Number of ZZP2 clients (delivered).....	71
Figure 25: Number of ZZP3 clients (delivered).....	72
Figure 26: Number of ZZP6 clients (delivered).....	72
Figure 27: Occupancy rate per location .....	74
Figure 28: Occupancy rate forecast (location 3).....	74
Figure 29: Waitinglist per month per location .....	75
Figure 30: Waiting list per product.....	76
Figure 31: Average duration of ZZPs .....	76
Figure 32: Data mining goals matrix .....	81

## LIST OF TABLES

Table 1: Interviewees.....	33
Table 2: Quality of Care information needs (by type of experts) .....	34
Table 3: Financial information needs (by type of experts) .....	37
Table 4: Information needs top 10 .....	39
Table 5: Care institution locations .....	46
Table 6: Number of clients per gender and status .....	47
Table 7: Number of incidents per type .....	49
Table 8: Number of incidents per incident location.....	54
Table 9: Mental damage per incident type.....	55
Table 10: Physical damage per incident type.....	55
Table 11: Risk assessment .....	56
Table 12: Association rules increased risk.....	57
Table 13: Number of incidents with respect to risk assessment outcome.....	58
Table 14: Coercive measures / restraints .....	67
Table 15: Age distribution per status .....	67
Table 16: Number of different values .....	68

## GLOSSARY

<b>Electronic Client Record (ECR)</b>	Tailor made software for the long-term care sector which enables care institutions to collect, consult and maintain valuable information about clients and to connect with stakeholders to exchange data.
<b>Electronic Patient Record (EPR)</b>	Tailor made software for mostly hospitals, which enables hospitals to collect, consult and maintain valuable information about their patients and to connect with stakeholders to exchange data.
<b>Exceptional Medical Expenses Act (AWBZ)</b>	The risk of exceptional medical expenses, which cannot be insured on an individual bases, is insured under the Exceptional Medical Expenses Act (Algemene Wet Bijzondere Ziektekosten; AWBZ). Everyone who works or resides in the Netherlands is insured and entitled for AWBZ care reimbursement (Ministerie van Sociale Zaken en Werkgelegenheid, 2010).
<b>Social Support Act (WMO)</b>	WMO is meant for citizens who want and are able to stay in their own house, but need some kind of support. Examples of support that could be financed by the WMO are: help with the household, adjustments in house, catering, transport services, etc. (Rijksoverheid, 2012).
<b>Personal Budget (PGB)</b>	PGB is meant for citizens who want to purchase the desired care by themselves. PGB is paid to the citizen and is a percentage of the money that is normally paid to the care institutions (Rijksoverheid, 2012).
<b>Nursing, Care &amp; Home care (VV&amp;T)</b>	VV&T or VVT is a common Dutch abbreviation used for all nursing homes, care institutions & home care organizations.
<b>Mental Healthcare (GGZ)</b>	GGZ is the Dutch abbreviation for the mental healthcare sector.
<b>Disability Care (GZ)</b>	GZ is the Dutch abbreviation for the disability care sector.
<b>Care Intensity Package (ZZP)</b>	ZZP is the Dutch abbreviation for a care intensity package, which describes the care that a client need. A wide range of ZZP's exist to provide clients with the right amount and type of care.
<b>Knowledge Discovery (KD)</b>	<i>“Knowledge discovery is a process that seeks new knowledge about an application domain”</i> (Cios, Pedrycz, & Swiniarski, 2007) (Kurgan & Musilek, 2006)
<b>Knowledge Discovery Process (KDP)</b>	A KDP is the process that enables to seek new knowledge about an application domain. CRISP-DM, KDD and SEMMA are 3 examples of KDP's.
<b>Data Mining (DM)</b>	<i>“the application of specific algorithms for extracting patterns from data”</i> (Fayyad, Piatetsky-Shapiro, & Smyth, 1996)

# 1. INTRODUCTION

*“To ensure that for persons with a long-term or chronic disorder of a physical, intellectual or psychological nature, care of good quality is available and that the cost level of this care is acceptable to society.” (Mot, Aouragh, Groot, & Mannaerts, 2010).*

In 2008, the general policy goal for long-term care, as stated above, was formulated. Long-term care is care for people with a long-term or chronic disorder. This chronic disorder could be of physical, intellectual or psychological nature. The two most important goals for long-term care could be extracted from this general policy goal and are: care of good quality at an acceptable cost level. These goals apply to all care institutions. In order to support these goals, care institutions should have insight in the financial and quality state of their internal organization.

Nowadays, a growing number of long-term care institutions, in the Netherlands, use ECR software in order to collect all the necessary information about the core process of a care institution. All the information in an ECR is client related, this means that the client is the central entity. Information in an ECR is very elaborate. Personal details, medical information, financial information, presence and absence, indication, production, care plan, pharmaceutical use, agenda, documents, treatment plans, etc. is all registered in an ECR. All this registered data is mostly used on an individual basis by care personnel. Care personnel use the information when they are with a client, in order to help and support the client as well as possible. Also staff functions use the data collected with ECR software, for example to declare the realized production. The declaration of the realized production is necessary for care institutions to receive their monthly budget.

ECR software is different than EPR software that is used in the cure sector. This distinction between an ECR and an EPR is very important for this research, in order to avoid confusion. The purpose of the software is quite the same, but there are some major differences in functionality at this moment. One could argue that one system could be used in both hospitals (EPR) and care institutions (ECR), but such a system is not in place at this moment. Both ECR and EPR software is tailor made to the sector in which the software is used. ECR software enables care institutions to declare production by the standards (AZR 3.0 standard) used in the care sector.

## 1.1. PROBLEM STATEMENT

Long-term care within the Netherlands has become one of the biggest expenses at this moment for the Dutch government. The expenditures of only the AWBZ grew from €14 billion in 2000 to €23 billion in 2010 (Ministerie van Volksgezondheid, Welzijn en Sport, 2011). According to the population forecast of the CBS in 2010 the amount of elderly in The Netherlands will increase from 2.4 million to 4.6 million in 2040 (Duin & Garssen, 2010). This means that a growing amount of people will need some form of care, which is also addressed by Mot, Aouragh, Groot and Mannaerts (2010).

Not only the number of people increases, also the life expectancy increases (Duin & Garssen, 2010). This means that more people will need some form of care and they need it for a longer period of time. Result of this is that the expenditures of care will increase, this is a continuation of the current upward trend. With this on-going trend of ageing, it is necessary to search for solutions to make the long-term care affordable and sustainable in The Netherlands (Mot, Aouragh, Groot, & Mannaerts, 2010). This is a major task for the ministry, but this will affect the budgets of care institutions. It is therefore important for care institutions to know exactly how its internal business operates and how the core process could be improved.

The number of attributes that is registered for one client, is constantly increasing which results in even more data. One cause for this is that ECR is relatively new in the care sector and far from mature. Another cause is that the control pressure is increasing. This means that care institutions need more data in order to justify itself, which as a result creates a large database with a lot of valuable data. Care personnel on the other hand are contracted to deliver care to the clients and are negative towards the registration of all kinds of data.

At this moment, long-term care institutions generate the required reports for their stakeholders, in order to receive budget and to justify itself. The data, that is collected during their core process, could be used for numerous other analysis, for example to improve their core process. There is no research available that describes the information needs of care institutions to gain insight in their internal organization. Therefore this research focuses on the information needs of care institutions and the data that is present in an ECR.

The Ministry of Health, Welfare and Sport (2011) has also addressed the importance of management information on quality: *“In order to supply care of good quality it is essential that the care institution is managed well and that the institution has permanent information about the quality of the institution”*. The Ministry of Health, Welfare and Sport is aiming at quality improvement within care institutions. One of the current projects is a new care client right law (Wet Cliëntenrechten Zorg; WCZ). This law encapsulates also that structural deficiencies, in the quality of care, can lead to the order to cease the provision of care, as the ultimate remedy.

As already said before, the use of an ECR is important for a care institution. In order to receive budget it is crucial, for a long-term care institution, that the information in this ECR is up-to-date and correct. Receiving budget is not the only important task of ECR software. ECR software is also used to deliver the best possible care for and to the clients, manage and have insight in their internal business, and to comply with the laws and regulations in the Netherlands. The data is not fully exploited at this moment. Care personnel hold responsibility to register all the required data during the core process. This data, however, is not fully exploited to improve the core process. As a result, the registration of data is seen as bureaucracy and unnecessary. Care institutions are not aware of the fact that they possess valuable data. Gathering data should not be the goal of care institutions, using the gathered data to create interesting reports and to find interesting trends, patterns and correlations should be the goal of gathering data for care institutions. Using the gathered data is directly the main problem in the current information-centric world (Kurgan & Musilek, 2006). If the unstructured data can be made explicit, the information arising therefrom could be used to improve the processes (Feelders, Daniels, & Holsheimer, 2000).

## 1.2. RESEARCH QUESTIONS

From the problem description the following main research question is composed:

*“ How can knowledge discovery support Dutch long-term care institutions to manage the internal organization? ”*

Knowledge discovery is a process that could help to discover new knowledge from unstructured data. The exact description could be found in the Related literature chapter. The main research question aims at the use of knowledge discovery to support Dutch long-term care institutions to manage their internal organization. Focus of the research is on the information needs of Dutch long-term care institutions and the data that is gathered during the core process. The aim of this research question is to use the collected data more efficient and effective. More efficient and effective use could result in more willingness of care personnel to register the required data.

This main question is divided in sub questions. These sub questions are:

- *What are the information needs of Dutch long-term care institutions?*

It is important to know what the information needs of Dutch long-term care institutions are. The information needs are the basis for further analysis. Without knowledge of the information needs it is impossible to construct the right models and to do proper analysis. Unstructured in-depth interviews are used to discover the information needs of Dutch long-term care institutions. In addition, scientific literature will be used to support the outcome of the interviews.

- *What data are currently available for Dutch long-term care organizations?*

Before the analysis of the data, it is important to know what data is currently collected by Dutch long-term care organizations and how this data is being registered. Data of one long-term care institution is used to answer this sub research question. The long-term care institution consists of more than 5 locations and support almost 1000 clients.

- *To what extent can knowledge discovery satisfy the information needs of Dutch long-term care organizations, at this moment?*

Based on the information needs and the gathered data it is possible to judge whether knowledge discovery could satisfy the information needs of Dutch long-term care organizations. For this sub research question the same data will be used as the data used in the previous sub question.

- *How can the use of the gathered data be increased?*

The assumption on forehand is that there will be an information gap. This means that the gathered data could not fully satisfy the information needs of Dutch long-term care institutions. Increasing the use of the gathered data is important to decrease bureaucracy and registration of unnecessary data. As a result, care personnel can spend more time on the core process.

### 1.3. RELEVANCE

#### 1.3.1. SCIENTIFIC

This research contributes in the field of knowledge discovery and data mining. Knowledge discovery and data mining literature is available for the cure sector, but not for the care sector. One of the goals of this research is to apply the CRISP-DM method (Chapman, et al., 2000) to the care sector. As researched in this study the CRISP-DM method is the 'de-facto standard', but the care sector is not standard. The environment of a Dutch long-term care institution is very complex. This complexity is caused by the laws and regulations that apply to the care sector, because the care is mainly financed by public money. Also stakeholders of care institutions, for example care insurers, demand all kinds of information.

The other goal is to discover the information needs of long-term care institutions. Long-term care institutions in the Netherlands need to hand in all kinds of information, in order to receive budget and to comply with the laws and

regulations. The use of data for managing their internal business is not exploited yet. This research tries to visualize the information needs and the gathered data in order to give insight in the information gaps of long-term care institutions. Bridging this gap would ensure that management of long-term care institutions have the proper information.

### 1.3.2. SOCIAL

From a social perspective, this research is relevant because of its insight into the information needs of long-term institutions. This research reveals the information needs of long-term care institutions and their stakeholders. As a result all long-term institutions could benefit from this research in order to manage their internal business. Not only the management of long-term care institutions could benefit from this research, also care personnel could benefit. If the information needs are clear, it is also clear what kind of data should be registered. The amount of data that should be registered could then be minimized.

Various initiatives have started to gain insight in the quality of the care. *Zichtbare Zorg (ZiZo)* is one of the examples, which creates selection information for clients, purchasing information for care insurers, monitoring information for the inspectorates, and improve information for the institution (*Zichtbare Zorg, 2011*). But, these initiatives and the information is not explicitly for managing the internal business, the focus is more on the selection information for clients and monitoring information for the inspectorate. In order to manage the internal business, more information is necessary about the internal process of the care organization. This research aims at identifying the information needs in a clear and reliable way. This aim will be fulfilled by interviewing various care institutions and their stakeholders.

## 2. THEORETICAL BACKGROUND

### 2.1. DUTCH LONG-TERM CARE SECTOR

The complete Dutch healthcare sector can be divided into several subsectors, which are: cure, care, prevention, protection and promotion. Subsector care, the focus of this research, is distinctive because the treatment is not aimed at the healing of a client, but clients require support, care or nursing for a long period of time. The clients can either be admitted to a care institution (intramural) or receive the support that is needed at their own home (extramural) (van Rijn, 2011).

The care sector consists of three subsectors, which are Nursing, Care & Home care, Mental Healthcare and Disability care. These three subsectors are split up as follows (Nederlandse Zorgautoriteit, 2012):

- Nursing, Care & Home care (Verpleging, Verzorging & Thuiszorg; VVT);
  - Nursing & Care (Verpleging en Verzorging; VV)
  - Home care (Thuiszorg)
- Mental Healthcare (Geestelijke Gezondheidszorg; GGZ);
  - Continued residence (B category)
  - Residence with support (C category)
- Disability Care (Gehandicaptenzorg; GZ).
  - Mentally deficient (Verstandelijk gehandicapt; VG)
  - Light mentally deficient (Licht verstandelijk gehandicapt; LVG)
  - Highly disturbed behaviour light mentally deficient (Sterk gedragsgestoord licht verstandelijk gehandicapt; SGLVG)
  - Physically handicapped (Lichamelijk gehandicapt; LG)
  - Sensory disability (Zintuiglijk gehandicapt; VG)
    - Auditory and communicative (Zintuiglijk gehandicapt auditief en communicatief; ZGA)
    - Visual (Zintuiglijk gehandicapt visueel; ZGV)

Another distinction is that of residential and ambulant care. When the client reside in a care institution or hospital this means that it is residential care. Examples of residential care are the mental healthcare institutions, disability care institutions, care institutions and nursing homes. Residential care is therefore a type of intramural care. Ambulant care is care for which a client needs to come to the care giver. Examples of ambulant care are treatments provided by doctors without the need to pass the night (van Rijn, 2011).

In the Netherlands there is also a distinction possible that relates to the care chain. This care chain is the chain in which the client resides fictitious, during its treatment. The care chain can be divided into three different places: primary care, secondary care and tertiary care. Primary care is the care that is publicly available and direct accessible for every individual. Some examples are: first aid, general practitioners and emergency care. Secondary care is not direct accessible. For this care it is mandatory to have an indication. This type of care is in most cases outpatient care, so the client has to come to the care giver. Medical specialists is one of the examples. Tertiary care is mostly residential care for which also an indication is mandatory. The client resides in a care institution or hospital in order to conceive the desired care (van Rijn, 2011). This research aims at the secondary and tertiary care.



### 2.1.1. FINANCING SCHEMES

Long-term care is one of the biggest costs of the Ministry of Health, Welfare and Sport, with its 38% of the total health care budget (Schäfer, et al., 2010) (Ministerie van Volksgezondheid, Welzijn en Sport, 2011). Different financing schemes are in place, which are elaborated in this sector.

- **Exceptional Expenses Act (AWBZ)**

The risk of exceptional medical expenses, which cannot be insured on an individual bases, is insured under the Exceptional Medical Expenses Act (Algemene Wet Bijzondere Ziektekosten; AWBZ). Everyone who works or resides in the Netherlands is insured and entitled for AWBZ care reimbursement (Ministerie van Sociale Zaken en Werkgelegenheid, 2010). The AWBZ is the biggest financing scheme for the long-term care and is in place since 1968 (Ministerie van Volksgezondheid, Welzijn en Sport, 2011). Care that is insured by the AWBZ consists of: personal care, nursing, support in daily life, treatment, and residence.

Currently, the Ministry of Health, Welfare and Sport is working one of the biggest reforms since the inception of the AWBZ (Ministerie van Volksgezondheid, Welzijn en Sport, 2011). These reforms have complications for care institutions and their clients. One of the reforms is to decrease the amount of laws and regulations, which has a major impact for care institutions. Other reforms are increase the quality of care by investing 852 million euro each year for extra personnel and education, the establishment of a Quality Institute, improve the monitoring of the Health Care Inspectorate (Inspectie voor de Gezondheidszorg; IGZ) and strengthen the client rights by establishing the new WCZ law (Ministerie van Volksgezondheid, Welzijn en Sport, 2011) (Schäfer, et al., 2010).

- **Social Support Act (WMO)**

Another financing schema is the Social Support Act (Wet Maatschappelijke Ondersteuning; WMO). This care is budgeted by the Ministry of Health, Welfare and Sport at 1,44 billion euro (Ministerie van Volksgezondheid, Welzijn en Sport, 2011) and started on January 1, 2007 (Cadsand, Cazemier, Koster, Roose, & Vos, 2011). Municipalities are responsible for proper implementation of this act (Schäfer, et al., 2010). WMO is meant for citizens who want and are able to stay in their own house, but need some kind of support. Examples of the support that could be financed by the WMO are: help with the household, adjustments in house, catering, transport services, etc. People who need support in their daily life could request budget by the municipalities. AWBZ financed care is legally enforceable, in contrast with WMO financed care which is a provision (Rijksoverheid, 2012).

- **Personal Budget (PGB)**

Another possibility to finance the needed care is to request a personal budget (Persoonsgebonden Budget; PGB). In this situation the client wants to purchase the desired care by itself. A percentage of the AWBZ or WMO budget is then transferred to the client for which the client can purchase the desired care. After the desired care has been delivered the client receives an invoice, which has to be paid (Rijksoverheid, 2012).

- **Private finance**

Clients can also choose for private financed care. This is nothing more than care that is being paid by the client itself, without getting a budget from the Ministry of Health, Welfare and Sport or other parties. The client receives an invoice, from the care provider, that has to be paid.

### 2.1.2. STAKEHOLDERS

The Dutch long-term care sector is very complex and long-term care institutions have a lot of stakeholders. Most care institutions are set up as a foundation or association. The Dutch law makes it mandatory to have a board of directors (Raad van Bestuur; RvB). Most care institutions have also a board of supervisors (Raad van Toezicht; RvT), that appoint and dismiss the directors and supervises their decisions. On top of these boards there are other governmental organizations and institutions that supervise the care institutions. This stems from the fact that the long-term care sector is financed by money from the community. This money must be used in an appropriate and reliable way. The two most important supervisors are the Dutch Health Authority (Nederlandse Zorgautoriteit; NZa) and the IGZ. Besides these two large supervisors, there are a lot more stakeholders in the domain.

- **Ministry of Health, Welfare and Sport**

Dutch Ministry of Health, Welfare & Sport is an administrative body, ultimately responsible for the healthcare sector. Rules and regulations are initiated and created by the ministry in order to support the major goal of high quality care which is affordable and accessible (Schäfer, et al., 2010). The ministry can be seen as the policy maker in the health care sector and not as a stakeholder that demands information. The information used by the ministry is not more than the information the IGZ and care offices uses. Some additional information used by the ministry is being gathered from the Central Bureau of Statistics (CBS). This is being stated by the two interviewees interviewed during this study.

- **Dutch Health Authority (NZa)**

The NZa is an independent Dutch supervisory body, funded by the Ministry of Health, Welfare and Sport, that monitors the competition within the care sector and determines the maximum tariffs. Another task of NZa is determining what care should be delivered in order to obtain the tariff, and controlling whether the rules are followed (Mot, Aouragh, Groot, & Mannaerts, 2010) (Nederlandse Zorgautoriteit, 2011). The mission of the NZa is: *“The NZa creates and monitors properly functioning healthcare markets. The interests of the consumers are central in the performance of these tasks. Efficiency, both in the short and long term, market transparency, freedom of choice, access to healthcare and quality are guaranteed. This gives the consumer the best value for his or her healthcare euros.”* (Nederlandse Zorgautoriteit, 2011).

- **Health Care Inspectorate (IGZ)**

IGZ is another supervisory body and is also funded by the Ministry of Health, Welfare and Sport. Their tasks are explained, by IGZ itself, as: *“promoting public health through effective enforcement of the quality of health services, prevention measures and medical products.”* (Inspectie voor de Gezondheidszorg, 2011). It is an independent institute that advises the responsible ministers in order to ensure that only ‘responsible’ care is provided. One of the main tasks of IGZ in the long-term care sector is to ensure that the quality of care

institutions is at a certain level (Schäfer, et al., 2010). Risk assessment is used in order to achieve this. Inspection takes place when a certain care institution is at an above normal risk level.

- **Municipalities**

Municipalities are responsible for the implementing WMO care. Clients that require social support could contact the municipality to request the desired type of care. Municipalities have considerable freedom regarding the establishment of the WMO execution. This results in the many variations of WMO implementations (Schäfer, et al., 2010).

- **Centre for Needs Assessment (CIZ)**

The CIZ, founded in 2005, assesses whether people are entitled to receive AWBZ financed care. In order to ensure that the assessment is the same for every citizen of the Netherlands, the assessment is based on objective criteria. The criteria are a translation of the guidelines of the Ministry of Health, Welfare and Sport (Centrum Indicatiestelling Zorg, 2012).

- **Care offices**

Care offices have the mandate of care insurers to purchase AWBZ care on a regional basis for all the client in that region. Care offices purchase care for clients by care institutions, care institutions deliver care to the clients and clients pay for care. In 2013 the role of care offices will be taken over by care insurers. Care insurers will then purchase care for their own clients by care institutions.

At this moment care institutions need to send AZR messages to the care offices in order to get budget. In these AZR messages information about clients and production is transmitted. This enables the care offices to purchase the right amount of care for the right price.

- **Care insurers**

Care insurers will purchase from 2013 onwards the AWBZ care for their own clients. This gives them a lot of grip in the sector. If a care institutions delivers care of low quality the care insurer could choose to purchase the care for its clients by another care institution. Care insurers get money for purchasing care from the Ministry of Health, Welfare and Sport. AWBZ is financed by the clients through taxes. This means that the client does not pay insurance to the care insurer for AWBZ financed care. Clients pay insurance for the Care Insurance Law (Zorgverzekeringswet; ZVW) to the care insurer, and this care insurer also arranges the AWBZ for the client (if needed). The ZVW law makes it mandatory for all citizens in the Netherlands and all people that live outside the Netherlands but do get salary from the Netherlands to be insured.

- **Central Administration Office (CAK)**

The CAK is a body that operates on behalf of the care insurers. The CAK is the company that administrates and registers the own contribution of all AWBZ and WMO financed care clients (Het CAK, 2012). Every clients that receive care financed by the AWBZ and WMO are obliged to pay an own contribution. This own contribution is determined by the CAK.

- **CIBG**

CIBG is an agency of the Ministry of Health, Welfare & Sport, founded in 1995 as a project and it is independent since 2003. Care institutions that want to deliver AWBZ financed care should be provided with an admission of the minister or secretary of state of the Ministry of Health, Welfare & Sport. CIBG is the agency that manages these admissions. Also the admissions for care givers are managed by the CIBG (CIBG, 2012).

- **Care clients**

The clients are one of the most important stakeholders of care institutions. Clients have a demand for care which should be delivered by the care institution. The care delivered by the care institution should be of high quality at an acceptable cost level.

### 2.1.3. CARE PROCESS

When a person wants to receive care financed by AWBZ, WMO or PGB, an indication should be requested by the CIZ. For the AWBZ and WMO financed care applies that the person can purchase care within the received indication and the purchased care is financed by the care insurers via the care offices. As already stated before, PGB is a budget which is a percentage of the indicated AWBZ or WMO care that is directly transferred to the client. The client should use this budget to purchase the desired care.

For AWBZ financed care the client receives an Care Intensity Package (*Zorgzwaartepakket*; ZZZP), since 2009. A ZZZP is the indication that described the amount and type of care that is financed by the AWBZ. There is a large number of ZZZP's available in the three different subsectors of long-term care (VVT, GGZ, GZ) (Rijksoverheid, 2011). For these ZZZP's the distinction has been made between VVT, GGZ and GZ, but these subsectors are also split up as presented at the beginning of this section. All the ZZZP's are elaborated in Appendix XXII – ZZZP products. This appendix gives insight in the ZZZP's per sector, and a short description per ZZZP. One interesting ZZZP is the ZZZP X, old indications that could not be migrated to a proper ZZZP, that is used as an undefined ZZZP. No clients will get an ZZZP X indication from the CIZ, only old indications from before the current structure are migrated to a ZZZP X indication.

The indication needs to be requested by the CIZ. Clients that need low intensity care will receive a low ZZZP indication, clients that need high intensity care will receive a high ZZZP indication. The care institution declares all care delivered to the client and will receive the money from the care insurers via the care offices. If a client desires more care than is indicated by the CIZ, care institutions are allowed to deliver more care. However, this extra care is not financed by the AWBZ. This means that in practice a small number of care institutions will deliver more care than is indicated, because that extra care is not financed. This extra care could also be invoiced to the client, but this should be communicated with the client. Every client is responsible for the right indication, which means that the client should request a new indication when it desires more care than indicated. All clients that receive AWBZ care are obliged to pay an own contribution. This own contribution is determined by the CAK for each client.

Also for WMO care the client needs an indication. But this indication should be requested by the municipality, where the person is registered. The municipality determines which support is needed for the client to stay in its own house. The indication describes what care the client should receive. Also for WMO care the client is obliged to pay an own contribution. CAK is also in place to determine the own contribution.

## 2.2. KNOWLEDGE DISCOVERY PROCESSES

“*Knowledge discovery is a process that seeks new knowledge about an application domain*” (Cios, Pedrycz, & Swiniarski, 2007) (Kurgan & Musilek, 2006). This definition is used for knowledge discovery in this study, because it covers the whole process, from understanding the problem to the implementation of actions (Feelders, Daniels, & Holsheimer, 2000). In this study KDD is seen as a Knowledge Discovery Process (KDP), and data mining is one step in this KDP. In the past, several knowledge discovery processes have been researched. This study itself is a KDP, applied to the Dutch long-term care sector. The CRISP-DM method (Chapman, et al., 2000) is chosen to be the research method and thus the KDP for this study. A description of the researched KDP’s can be found in the Theoretical background section.

The first phase in many KDP’s is to create an understanding of the application domain (Azevedo & Santos, 2008) (Chapman, et al., 2000) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). It is essential to understand the business, its information needs, and the data before new knowledge could be derived. Without this understanding, it can be dangerous to apply data mining. The interpretation of the created models is very different for people who do not understand the business, and its environment and could therefore be dangerous. This means that new patterns could be discovered which are impossible, in practice (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

The KDP’s described in this section are: Knowledge Discovery in Databases (KDD) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), CRoss-Industry Standard Process for Data Mining (CRISP-DM) (Chapman, et al., 2000) and the SEMMA method (Azevedo & Santos, 2008). These knowledge discovery processes are seen as the most used in the field and in research (Harding, Shahbaz, Kusiak, & Srinivas, 2006) (Onwubolu, 2009) (Azevedo & Santos, 2008). Other KDP, such as 5 A’s, Human-Centered, Two Crows, Annand & Buchner, Cabena et al., Cios et al., RAMSYS and DMIE (Berendt, 2012) (Marban, Mariscal, & Segovia, 2009), have not been researched because the use of these KDP’s is limited.

### 2.2.1. KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Fayyad, Piatetsky-Shapiro and Smyth (1996) define KDD as: “*the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*”. This method consists of the following steps: Selection, Pre-processing, Transformation, Data Mining, Interpretation/Evaluation (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Before the process, as described below, can start it is essential to understand the application domain. This includes the goal of this process.

- **Selection**

Understanding the application domain is the first thing to do. The description of KDD does not explicitly describe the task to understand the application domain. Creating a target data set, or focusing on a subset of variables or data samples, on which discovery is to be performed is also part of this step.

- **Pre-processing**

The target data will be cleaned and pre-processed in order to obtain consistent data. This includes removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time-sequence information and known changes.

- **Transformation**

Transformation of the data using dimensionality reduction or transformation methods will be performed in this step. Useful features to represent the data depending on the goal of the task should be discovered.

- **Data Mining**

This step focus on the search for patterns of interest in a particular representational form, depending on the DM objective. Matching the goals to data mining methods, exploratory analysis and hypothesis selection are also part of this step and will be performed before the actual search for patterns.

- **Interpretation / Evaluation**

Interpretation and evaluation of the mined patterns will take place during this step. This includes visualization of the data if necessary. The use of the discovered knowledge is the final objective (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) (Azevedo & Santos, 2008).

KDD consists of 9 steps which are mapped on the above stated steps. It is an iterative process and the data is the basis . Loops may exist between any two steps in the process. Before the process can start, it is essential to understand the application domain. The authors described this, but did not include it the first step of their KDD model. A description of what should be understood of the application domain is therefore not available (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) (Azevedo & Santos, 2008).

#### 2.2.2. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

CRISP-DM is an industrial process developed with the support of IBM (Chapman, et al., 2000) (Cios, Pedrycz, & Swiniarski, 2007). The method is created with the aim to be neutral with regard to industries and tools (Berendt, 2012). The method is iterative and consists of six main phases, that are: Business understanding, Data understanding, Data preparation, Modelling, Evaluation, Deployment.

- **Business understanding**

Understanding the application domain is the initial step of this method. The understanding includes the objectives and requirements from a business perspective. This knowledge will then be translated into data mining goals. The project plan consists of a preliminary description how the goals will be achieved.

- **Data understanding**

The data understanding will determine which data is available or should be collected. This step will also be used to get insight in the data and to determine the quality of the data. If the data is more elaborate than the business knows, the process could return to the Business understanding step in order to create more elaborate goals.

- **Data preparation**

This step will then prepare the data for proper modelling. The raw data is being transformed to a final dataset that is used in the modelling step. This includes solving bad data and missing data problems. Also the creation of new variables using multiple existing variables could be included in this step.

- **Modelling**

Creating different models with the use of different data mining algorithms is the main activity of this step. Parameters are calibrated to optimal values in order to create optimal models. In order to make use of some data mining algorithms requires data to be in a certain format, that makes it necessary to return to the data preparation step in those situations.

- **Evaluation**

Models and the process of arriving at those models will be thoroughly evaluated during this step, before deployment of the model. Objective of this phase is to check if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision should be made whether the data mining results should be used or not.

- **Deployment**

How the new knowledge will be implemented will determined in the last phase, the deployment phase. Depending on the requirements the deployment phase could be simple or complex. The customer should now up front what actions need to be carried out in order to actually make use of the created models (Chapman, et al., 2000).

The CRISP-DM method is clearly described and is seen as the “*de-facto standard*” (Marban, Mariscal, & Segovia, 2009). Every phase, activity and even the output is described by the authors. This makes the method easy and fast to use and implement (Chapman, et al., 2000). Unlike the KDD method, the CRISP-DM included the Business Understanding step in which an understanding of the application domain could be generated. Also the last step, Deployment, is a step that is not included in the KDD method.

### 2.2.3. SAMPLE, EXPLORE, MODIFY, MODEL, ASSESS (SEMMA)

SEMMA is the abbreviation for Sample, Explore, Modify, Model, Asses, which are the steps in this method. The SEMMA method is developed as a set of functional tools for SAS’s Enterprise miner. The method does not include a step in which an understanding of the application domain could be generated, which is an important step for knowledge discovery projects (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). It is, with the CRISP-DM method, one of the most used methods in the field of knowledge discovery (Azevedo & Santos, 2008) (Harding, Shahbaz, Kusiak, & Srinivas, 2006) (Onwubolu, 2009).

- **Sample**

The method starts with the Sample stage, in which a portion of the large dataset is extracted. The extracted portion should be big enough to contain significant information. Another important aspect is to create a portion that is small enough to be able to manipulate quickly.

- **Explore**

The Explore stage consists of a first exploration of the data. This is done by searching for unanticipated trends and anomalies. Via this search understanding and ideas are generated which could be used during the rest of the process or project.

- **Modify**

Before the Model stage can start it is important to modify the data. The collected data is mostly not ready for modelling, therefore variables should be selected, transformed or maybe created. New variables could be created from multiple existing variables. The Modify stage is important for proper modelling.

- **Model**

During this stage the data is used to create models. The software used is allowed to search for a combination of data that reliably predicts an outcome.

- **Assess**

This stage consists on assessing the data by evaluating the usefulness and reliability of the finding from the DM process and estimate how well it performs (Azevedo & Santos, 2008).

One of the shortcomings of SEMMA is that it starts with the Sample step. In this step a sample of the population will be extracted. In the method there is no step in which the understanding of the business could be generated. In order to create proper models, it is important to understand the application domain (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

#### 2.2.4. CONCLUSION

The business and data understanding steps of the CRISP-DM method are the most important steps for this research. CRISP-DM is most the elaborate in terms of these two steps compared to SEMMA and KDD (Azevedo & Santos, 2008) (Berendt, 2012). All the steps from the KDD method are integrated in the CRISP-DM method. This makes the CRISP-DM method a more extensive method.

Besides that, the CRISP-DM method is clearly described, industrial of intent, and one of the most used methods in the field (Giraud-Carrier & Povel, 2001) (Harding, Shahbaz, Kusiak, & Srinivas, 2006) (Onwubolu, 2009). Marban, Mariscal and Segovia (2009) state that the CRISP-DM method is the “*de facto standard*” in the field for developing DM & KD projects. The Dutch long-term care sector is all but standard, which makes it interesting to use the method for this study.



### 2.3. KNOWLEDGE DISCOVERY IN HEALTHCARE

The last 15 years knowledge discovery is evolved in the healthcare domain from predicting epidemics (Prather, et al., 1997) to a broad application of data mining techniques (Koh & Tan, 2011). The data that is being captured by healthcare organizations is enormous, which makes it a treasure for data analysts (Lucas, 2004). Koh & Tan (2011) state that the use of knowledge discovery is increasingly popular, if not essential for healthcare organizations. Lucas (2004) identified the trend of increasing application of Bayesian statistical methods. These methods makes it possible to incorporate prior knowledge during the analysis of the data. A couple of applications of data mining in practice are (Koh & Tan, 2011):

- Diagnose new patients;
- Identify and understand high-cost patients;
- Evaluate the effectiveness of medical treatments;
- Identify and track chronic disease states and high-risk patients;
- Hospital infection control system (early-warning system in the event of epidemics);
- Determine the preferences, usage patterns, and current and future needs of individuals to improve their level of satisfaction;
- Predict other products that a healthcare customer is likely to purchase;
- Detect unusual or abnormal patterns of claims by physicians, laboratories, clinics and others.

Electronic Medical Records (EMR) or EPR are being primarily used for patient administration. But, the data gathered in EMR or EPR could be used for data analysis to determine relevant patterns and trends (Norèn, Hopstadius, Bate, Star, & Edwards, 2010). Because the primary use of EMR and EPR software was to administrate patient information, the database architecture is not optimal for data analysis. This leads to a large amount of problems. A couple of these problems are defined by Cios & Moore (2002), Goodwin, VanDyne, Lin & Talbert (2003), and Lindell & Pinkas (2002):

- Heterogeneity, quality and completeness of medical data
- Ethical, privacy, legal, and social issues
- Statistical philosophy
- Special status of medicine
- Lack of standardized language

## 2.4. DATA MINING

Data Mining (DM) is one step in a KDP and is being defined, as already stated in the problem statement section, by Fayyad, Piatetsky-Shapiro, & Smyth (1996), as: “*the application of specific algorithms for extracting patterns from data*”. This definition is the exact application of data mining in this research. Knowledge discovery and DM are used as synonyms by many researchers (Kurgan & Musilek, 2006). Data mining is used in this research as one step in a knowledge discovery process (Cios, Pedrycz, & Swiniarski, 2007). Therefore the definition of DM is the exact description of how DM is being applied in this research. The other steps that should be taken in order to do proper analysis is described in the different knowledge discovery processes. The data mining step from KDD, the modelling step from CRISP-DM and the model step from SEMMA are the DM steps (Azevedo & Santos, 2008). In those steps specific algorithms will be applied in order to extract patterns from data. DM originates from statistics, data analysts and management information systems communities (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

### 2.4.1. DATA MINING GOALS

There are two major goals for using DM, which are: prediction and description (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Feelders, Daniels & Holsheimer (2000) identified a third major goal for DM, which is intervention.

- **Description**

With description the goal is to discover interesting patterns, trends or relationships in the available data. For example, management of care institutions want to know if there are relations between care products. The gathered data could then be used to discover which products are purchased together (Feelders, Daniels, & Holsheimer, 2000) (Giraud-Carrier & Povel, 2001).

- **Prediction**

The available data is used to discover patterns, trends or relations in order to use these discovered information for data outside the database. When new treatment goals are formed the system could predict whether the treatment goal will be obtained or not, based on available data. With this knowledge care personnel could form more realistic treatment goals (Feelders, Daniels, & Holsheimer, 2000) (Giraud-Carrier & Povel, 2001).

- **Intervention**

Intervention is explained by Feelders, Daniels & Holsheimer (2000), as models that could enable active intervention in a system. For example, if care personnel registers an incident for a client. The system could then intervene and advice the care giver on the action that should be performed.

### 2.4.2. MODEL FUNCTIONS

A wide variety of DM model functions emerged from research in the past. Each functions contains a wide variety of algorithms and technologies. Some of the functions, most relevant for this research, are described in this subsection.

- **Summarization**

Summarization is the basic visualization of the data, or a subset of the data. One example is to visualize the data in a table with the mean and the standard deviations (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). These techniques are widely used in this study, because of the explorative character of the performed case study.

- **Classification**

Classification is used to classify a data item. The goal of classification is to predict the value of an attribute based on values of other attributes. Historical data is needed to learn the classification rules. For instance, email messages will be labelled as spam or not by a spam filter, depending on characteristics of the email. Whether an email is spam or not is based on historical data. Users label email messages as spam when the email is not desired, this information could be used to classify new emails automatically. (Chen, Han, & Yu, 1996) (Freitas, 2003).

- **Regression**

Regression is used to find relationships between a dependent variable and one or more independent variables. For example the price of a house is dependent of a couple of independent variables such as for example: cubic meters, location, square metres of garden and the view. Regression analysis enables to propose a price based on the independent variables using the historical data (Chen, Han, & Yu, 1996).

- **Clustering**

Using clustering functions it is possible to group a set of objects. The objects in one group are more similar to each other than the objects in the other groups. Clustering could be used to group for example all clients into specific focus groups. Different variables of clients can be used to group the clients. At this moment psychogeriatric clients reside on the same department. Clustering could group the clients based on other variables (Chen, Han, & Yu, 1996) (Freitas, 2003).

- **Association rule learning**

Association rule learning is used to discover interesting relationships between variables within a large database or dataset. A well-known example is that of the diapers and beer in the supermarket. A big percentage of people who buy diapers, buy also beer. Supermarkets use association rules with all their products in order to create the most efficient planogram. (Chen, Han, & Yu, 1996) (Freitas, 2003) Different algorithms exist to identify or learn the association rules, one example is the Apriori algorithm. This algorithm learns association rules in multiple steps and contains two parameters, a support and a confidence parameter. The support parameter indicates a percentage of transactions that meet the rule. The confidence parameter indicates the minimal level of confidence a rule should meet. Based on these parameters the algorithm learns the association rules in the multiple steps (Agrawal, Srikant, & others, 1994).

- **Time series forecasting**

Time series forecasting is used to forecast future values based on historical data. An example could be the prediction of the weather for the upcoming week. Many forecasting algorithms exist, as de Gooijer and Hyndman (2006) described. With the forecast() function in R it is possible to create forecasts based on time series, the forecast algorithm selects the best forecast algorithm based on the in-sample fit.

### 3. RESEARCH APPROACH

The approach for this study is elaborated in this section. For this study the CRISP-DM method is chosen to be the leading method. The reason for choosing the CRISP-DM method is grounded in the Theoretical Background section. All the phases and steps of the CRISP-DM method have been applied to this study.

#### 3.1. RESEARCH METHOD

The CRISP-DM method consists of 6 phases, which are: Business / Environment Understanding, Data Understanding, Data preparation, Modelling, Evaluation, Deployment. There is a possibility to return to Business / Environment understanding during the Data understanding and Evaluation phase. This research is based on one single loop of the whole process. These phases are also used during this study. The phases are made explicit in this subsection and is visualized in the following figure:

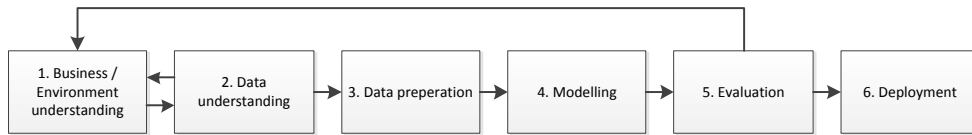


FIGURE 1: RESEARCH METHOD

#### 1. Business / Environment understanding

Unstructured in-depth interviews are used to create an understanding of the long-term sector. Unstructured in-depth interviews are used to ensure the richness of detail through clarification of questions and answers. Yin (2009) stated that open interviews are the best way to discover explorative information. Information needs are the result of the interviews, which are translated to data mining goals during this phase. The use of information needs is different than the business goals as prescribed in the CRISP-DM method, this is done because the information needs are more elaborate than the business goals. The data mining goals are input for the next phase: data understanding.

#### 2. Data understanding

To allow proper modelling, it is important to understand the gathered data. A single case study is performed, this means that the data that has been gathered by one care institution is used to do the modelling. During this step the data will be described and explored, also the data quality will be described. Due to security and confidentiality issues the database will not be described in detail. Some basic statistics about the database will be described.

#### 3. Data preparation

In order to use the gathered data of the care organization, the data must be prepared first. Missing data or wrong data must be revealed and solved in order to do proper modelling and evaluation. For some data mining goals it will be necessary to construct, integrate and format the data to enable proper modelling. This is also part of this phase.

#### 4. Modelling

Different models will be used during this phase to discover patterns and trends in the data, based on the data mining goals created during the business understanding phase. The models created in this step will be evaluated in the next step.

#### 5. Evaluation

In this research phase the models are being evaluated. Experts from the long-term care organizations will be used to evaluate the models. These experts will evaluate to what extent the models, created in the previous step, cover the information needs. If necessary, the process returns to the business/environment understanding step in order to clarify the goal of the model.

#### 6. Deployment

This last stage consists of the conclusion and result of this study. The data mining goals are based on the information needs of the interviewed care institutions and stakeholders, data collected by one care institution is used to discover the availability of data, modelling and evaluation is used to check whether the models support the management of Dutch long-term care institutions. The conclusion consists of a clear description of the discovered added value of knowledge discovery and the shortcomings in the gathered data. Solutions are presented in order to create a more effective and efficient use of gathered data.

### 3.2. ACTIVITIES

The CRISP-DM method (Chapman, et al., 2000) consists of 6 phases in which different activities should be performed. In the following subsections the activities are described and linked to this study. All the activities and steps are visualized in the following figure:

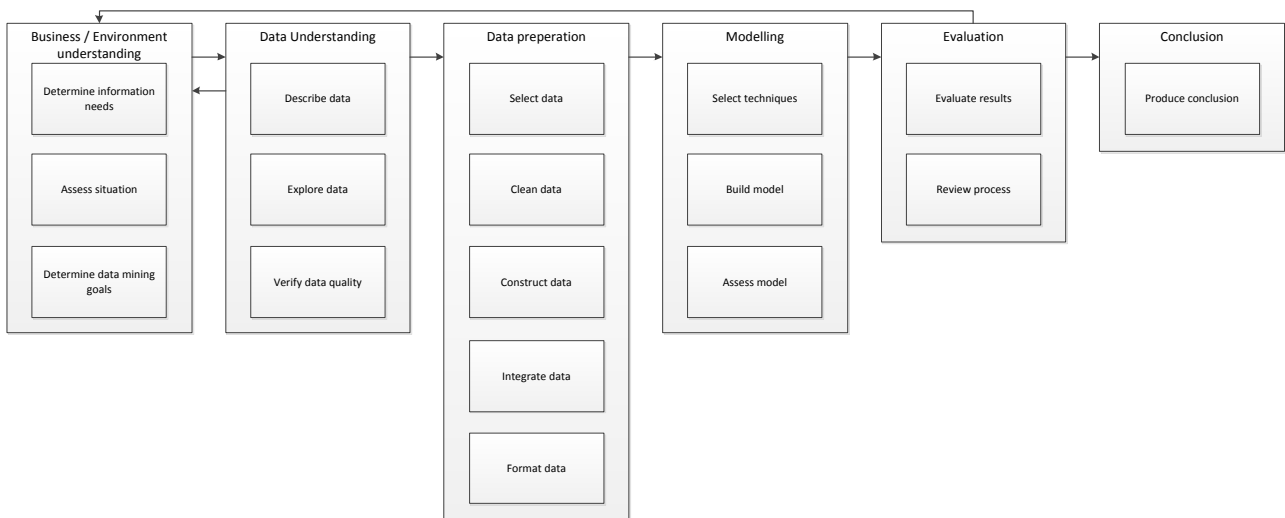


FIGURE 2: RESEARCH ACTIVITIES

#### 3.2.1. BUSINESS / ENVIRONMENT UNDERSTANDING

**Discover information needs** Multiple long-term care organizations will be interviewed, as well as other important stakeholders in this sector. In total 8 experts from the directors / board of directors level are interviewed, 7 experts from management level are interviewed, and 7

experts from stakeholders point-of-view have been interviewed. This makes in total 18 in-depth interviews with 22 experts. The outcome of this step consist of the information needs of both the care institutions and their environment. These information needs will be translated to data mining goals.

**Assess situation**

The resources used for this study are: an ECR dataset that contains a broad set of information about the care process, MySQL Workbench to select the required data from the dataset, R Studio in order to apply data mining techniques to the selected data and Microsoft Excel to create standard overviews of the selected data.

**Determine data mining goals**

Determining the data mining goals of long-term care organizations is one of the most important parts of this research. The information needs will be translated during this activity, that results in the data mining goals of long-term care organizations and their environment.

**Produce project plan**

This activity is omitted, because this document is the project plan.

3.2.2. DATA UNDERSTANDING

**Collect initial data**

This activity is omitted, because the data is already collected. Data of one care institution is used for the remainder of this study. The care institution has currently around 850 clients (intramural + extramural) and is working with an ECD, called ResidentWeb, since 2008. In order to use the data, personal information (names, BSN numbers, addresses, etc.) about the clients had to be deleted. This research is commissioned by Technology To Serve, which is the developer of ResidentWeb and the care institution is closely involved in the (on-going) development of the software. Using the data has been approved by a member of the board of directors. A confidentiality statement had to be signed before using the data for this research. It is also required to display the data anonymously. The database that is used for this research contains information until April 17, 2012.

**Describe data**

In order to do proper modelling it is crucial to know what data has been gathered. During this step the data, that will be used in the remaining steps, is described. This does not mean that the complete database will be described. The database is very large and complex, with over 300 tables. Security and confidentiality are two other arguments why the database will not be described.

**Explore data**

During this step the data is being explored with the data mining goals in mind. Some basic numbers are used to discover interesting preliminary results or sub-populations. With the output of this step the data mining goals could be fine-tuned.

**Verify data quality**

In order to use the data it is important to know what the quality of the data is. In this step the data quality will be determined. Is there missing data? Or incorrect data? What to do with it? These are some of the questions that will be answered in this step.

### 3.2.3. DATA PREPARATION

- Select data** After clarification of the data mining goals, the data that can be used for modelling must be selected during this step. Only the data that is relevant for the data mining goal will be selected.
- Clean data** The collected data must be of proper quality. During this step the data quality will be raised to the proper level by selecting subsets, insertion of default values or other techniques to raise the data quality.
- Construct data** This step includes the construction of the data that will be used during the modelling phase of this research. New attributes will be constructed from one or more existing attributes, if necessary.
- Integrate data** Data from different tables will be integrated in one table in order to create new records or values. This makes the data, for some data mining goals, more suitable for modelling and analysis.
- Format data** In this step the data will be formatted to make the data, for some data mining goals, more suitable for analysis, the meaning of the data will remain the same.

### 3.2.4. MODELLING

- Select techniques** During this step the appropriate modelling techniques will be selected for each data mining goal. For some goals it may be necessary to select more than one modelling technique.
- Generate test design** This step is omitted, test designs are not built during this research due to time constraints. The selected data is verified before building the model. For the used data mining techniques in this study it was unnecessary to generate test designs. The goal of this study is to increase the usability of the gathered data, not to create the most complex models.
- Build model** The modelling technique will be used during this step to build a model on the prepared data.
- Assess model** After the model is created it will be assessed using the domain knowledge that has been part of this research. The model will also be assessed using the test design. This assessment is mainly focussed on the accuracy and generality of the model.

### 3.2.5. EVALUATION

- Evaluate results** During this step the experts will assess the created model in the previous step. They will look if the model satisfies their information needs (data mining goals). If the model does not satisfy their information needs, they will give feedback on the model.
- Review process** If all the models, that could be created, are created it is time to review the whole process. During the review we will look at the models and see if the correct data is

used, or if we overlooked something. Also the not satisfies information needs will be reviewed during this step. What data is missing to do proper analysis? What data should be collected in the future?

**Determine next steps**

This activity is omitted, because the next step is defined up front.

3.2.6. DEPLOYMENT

**Produce conclusion**

The conclusion of this study describes the information needs, data mining goals, models that were created for each information need or data mining goal, data quality and the objectives for the future in order to satisfy all the information needs and to create a more effective and efficient use of the gathered data.



## 4. FIELD STUDY: KD IN LONG-TERM CARE SECTOR

### 4.1. BUSINESS UNDERSTANDING

Before new knowledge will be derived from the gathered data, it is important to create an understanding of both the business and the data. Unstructured in-depth interviews are used to create an understanding of the business, in this case care institutions. A total number of 15 experts of care institutions, 8 experts from the board of directors / directors level and 7 experts from the management level, have been interviewed. The distinction between directors and management has been made due to the different viewpoints. Experts from the board of directors / directors level are more focussed on the financial part of their care institution, whereas management has the responsibility to maintain the quality of care. After the interviews it became clear that this difference in focus does not provide an obvious difference in the desired information.

Another 7 experts at different stakeholder positions have been interviewed, in order to gain an understanding from the stakeholders perspective. Different stakeholders are responsible for assessing care institutions. In order to assess a care institution it is important to have proper information about the care institutions. That is the reason for researching the information needs from the stakeholders perspective. In the table below the type of interviewees can be found with the number of experts interviewed.

TABLE 1: INTERVIEWEES

Type of interviewee	Number of experts	Number of interviews	Valuation
Board of directors / Director	8	8	10
Management	7	5	6
Stakeholders	7	5	3
Total	22	18	-

All the information needs of the interviewees are listed in the tables below. The information needs that have emerged from the interviews, are split and added to the supporting goal (quality of care & financial state). However, a true division is difficult to make because the rationale for the use of the information may vary for different individuals. For example, one director wants to use the “staffing with respect to the operations” information to increase the number of operations and create more efficiency in staffing. The information is then used for financial reasons, whereas another director wants to use the same information to control the deployment of qualified personnel. In this case the information is used to increase the quality of care.

The information needs are sorted based on their importance. The following formula is used to create a fair sorting (the example is based on the information need Customer experience):

$$Score = \sum_{Expert\ level} \frac{Times\ mentioned}{Number\ of\ interviews} \times Valuation = \left(\frac{8}{8} \times 10\right) + \left(\frac{4}{5} \times 6\right) + \left(\frac{3}{5} \times 3\right) = 16,6$$

The valuation enables information needs to be more important because they were mentioned by experts on directors level and not by experts on stakeholders positions. This study is aimed at the long-term care institutions, therefore information needs mentioned by experts within care institutions are valued higher than those of stakeholders. The distinction between experts on directors and managers level is created in order to make it possible that information needs are more important because they were mentioned by experts on a higher level.

4.1.1. INFORMATION NEEDS

**Quality of care**

The information needs, for managing the quality of care, that have emerged from the interviews are listed in the table below. The numbers in the table represent the number of interviews in which the information need has been discussed. Grey cells represent information needs that are mentioned in 50% of the interviews or more.

TABLE 2: QUALITY OF CARE INFORMATION NEEDS (BY TYPE OF EXPERTS)

		(Board of) directors	Management	Stakeholders	Score
Quality of care	Customer experience	8	4	3	16,6
	Number of incidents occurred	6	4	2	13,5
	Types of incidents occurred	6	4	2	13,5
	Causes of the occurred incidents	6	4	2	13,5
	Number of complaints from the clients	7	2	2	12,4
	Types of complaints from the clients	7	2	2	12,4
	Causes of complaints from the clients	7	2	2	12,4
	Patterns in occurred incidents	5	4	2	12,3
	Incidents with respect to risk assessment	4	4	2	11,0
	Number of clients	4	4	0	9,8
	Incidents with respect to care-related measures	3	4	2	9,8
	Incidents with respect to pharmaceutical use	3	4	2	9,8
	Number of clients at an increased risk	2	5	2	9,7
	Types of risk the clients run	2	5	2	9,7
	Client-related time & non client-related time	4	3	0	8,6
	Progress of care-related measurements	1	5	2	8,5
	Treatment goals (obtained & not-obtained)	2	4	1	7,9
	Progress of clients	4	2	0	7,4
	Care plan information	1	4	2	7,3
	Number of coercive measures / restraints	2	3	1	6,7
	Types of coercive measures / restraints	2	3	1	6,7
	Reason for coercive measures / restraints	2	3	1	6,7
	Age distribution of clients	2	3	0	6,1
	Average duration of stay	2	3	0	6,1
	Number of clients per demand for care	1	3	0	4,9
	Pharmaceutical use by the clients	0	3	2	4,8
	Client milestones (ageing & marriage milestones)	2	1	0	3,7
	Percentage male/female	1	2	0	3,7
	Activities of Daily Living tracking	0	2	1	3,0
	Reports of confidential counsellor	1	1	0	2,5
Registration delay (# days between application and first intake)	1	0	0	1,3	
Placement delay (# days between first intake and placement)	1	0	0	1,3	
Reason for admission	0	1	0	1,2	

The distinction could also be made, based on the size of the care institutions. Interviewed care institutions are categorized based on their size in number of clients. Four interviewed care institutions are categorized as small (less than 500 clients). Five interviewed care institutions are categorized as medium (between 500 and 1000 clients) and four care institutions are categorized as large (more than 1000 clients). This distinction has been made, and could be found in 'Appendix XX – Information needs by size'. This table is interesting, because it visualizes that small care institutions are interested in a limited amount of information. This is most probably because small care institutions have more control due to the small scale. Another reason could be that because of the small scale the staff has more contact with the same clients. The differences are small, making it difficult to identify real differences. Conclusion could also not be extracted from the available data.

Customer experience is information that is mentioned in every interview. Different types of customer experience surveys do exist. Most care institutions make use of the CQ index, which is the mandatory survey that have to be performed once in two years by an independent organization.

Incident information is important information as it is mentioned in many interviews, both on management and directors level. Every care institution is obliged to describe the types and numbers of incidents in their annual report. This makes the insight in the number and type of incidents an important information need for care institutions. It is not the direct goal to minimize the number of incidents, it is the goal to improve the follow-up and to reduce the impact of incidents. Interrelationships between types of incidents could indicate a connection between multiple incidents which could lead to improvement of care. Also care related measures and pharmaceutical use could indicate the cause of an incident. Patterns in occurred incidents is also discussed in many of the interviews. Patterns in occurred incidents could indicate that clients have more falling incidents in the corridor or have more aggression incidents in the weekend. Insight in this information could increase the quality of care by making the corridor safer or by deploying more personnel.

Complaints from the clients are considered very important on a directors level. Clients are entitled to complain about the care institution, when the client is dissatisfied. Care institutions are required to facilitate a complaints committee. This complaints committee must be independent. Most care institutions also have a complaints mediator. A complaints mediator could facilitate the client with information about the complaints scheme. The complaints mediator could also mediate by complaints in order to avoid formal complaints. Formal complaints are handled by the complaints committee. Because of the limited number of complaints (1 or 2 per year in normal situations), the application of data mining algorithms is considered unnecessary. Another reason is that complaints should be handled on an individual basis. Information about complaints may instead be used as a variable in other analyses, but because of the limited number of complaints this application is considered to be limited.

Information about the risk assessment is mentioned in many interviews. Especially management of care institutions is interested in this information. Risk assessments are carried out in order to improve the quality of care. Questions must be answered in order to assess if the client runs an increased risk at a certain problem. If the client runs an increased risk, based on the answers, care personnel could take proper precautions. Focus of the risk assessment is on falling, depressions, skin lesion, incontinence, use of medication, malnutrition, problem behaviour, etc.

Care institutions are also interested in some key figures about their clients, including the number of clients, age distribution, percentage male/female, client milestones and the average duration of stay. This is basic information that could easily be extracted from the data that is registered by care institutions.

The progress of clients is an important information need to manage the care institution. It is not mentioned by many experts, but when the progress of client increases it is important for care institutions to extend for example their waiting list. Average duration of stay, registration delay and placement delay have also to do with the progress of clients. Experts on the directors level are interested in this information in order to make strategic decisions. Currently, there is a trend that elderly stay longer at their own home. This means that the average age and the intensity of care will increase within care institutions. The average duration of stay will therefore decrease, resulting in a faster flow of clients.

Progress of care related measurements is tossed in many interviews, mostly interviews with managers, as being information that could help to prevent incidents. Care personnel measures occasionally the clients weight, length, heart rate, blood sucker, temperature, etc. An increase in the weight of the client could be observed immediately and care personnel could take the proper precautions. Also the yearly measurement of care-related indicators is part of the care related measures information. Care-related measures could also help to find relations in the data. Clients with a high heart rate may for example have more fall incidents than other clients.

A care plan contains information about the care a client receives, including the treatment goals. This information is important to deliver the best possible care for the client. Every half year a new care plan is created for every client. Treatment goals are evaluated at a date that has been planned during the care plan discussion. Management is interested in information about the number or percentage of obtained and not-obtained treatment goals. This information is not that important for experts on director level, because the responsibility to deliver care lies on a lower level in the care institution.

Client-related time versus non client-related time is information that is mentioned in both interviews with directors as in interviews with management. Client-related time is time that is directly used to deliver care to the client. This means that personnel is busy with the client. Non client-related time is the time to register all the information, education and training, meetings, etc. In order to maximize the client-related time, it is important to get insight in this information. The information could also help the care institutions to create more efficiency in their core process. The use of this information is, therefore, two-fold.

Activities of Daily Living (ADL) tracking is mentioned by two management experts. The reason that they are interested in this information is to track the progression or regression of a client. This could help management to increase the quality of care by measuring the abilities of a client over a period of time. For GGZ, GZ and rehabilitation in VVT this information could create better insight in the core process of the care institutions.

Coercive measures and restraints information is important for management of care institutions, because IGZ tries to minimize the use of coercive measures and restraints. Clients, representatives thereof, need to approve every restraint. Management is interested in this information in order to minimize the use of restraints. When restraints must be taken, the effectiveness of the restraints could be maximized. Insight in coercive measures is important for the same reason.

## Financial state

Besides the information needs that support the goal to deliver care of good quality, care institutions have also the goal to manage the financial state. In order to have insight in the financial state of the care institution it is important to know the production information, ZZP-mix information, occupancy information and staffing information. Focus of the interviews was on the data collected with ECR software. Information such as cost prices and other complex financial analysis have been held outside this study. In the table below, the information needs that have emerged from the interviews with the experts are elaborated.

TABLE 3: FINANCIAL INFORMATION NEEDS (BY TYPE OF EXPERTS)

		(Board of) directors	Management	Stakeholders	Score
<b>Financial state</b>	Staffing with respect to ZZP-mix	7	4	2	<b>14,8</b>
	ZZP-mix (per business unit)	7	4	0	<b>13,6</b>
	ZZP-mix prognosis	7	4	0	<b>13,6</b>
	Staffing with respect to operations	6	4	2	<b>13,5</b>
	Operations per ZZP	7	3	1	<b>13,0</b>
	Production information (planned, realized, declared)	7	3	1	<b>13,0</b>
	Care within & outside ZZP indication (planned, realized)	7	3	0	<b>12,4</b>
	Unplanned care	7	3	0	<b>12,4</b>
	Operations per function	6	3	1	<b>11,7</b>
	Staffing (per business unit)	6	3	0	<b>11,1</b>
	Current occupancy	5	3	0	<b>9,9</b>
	Occupancy prognosis	5	3	0	<b>9,9</b>
	Number of clients on waiting list	4	2	1	<b>8,0</b>
	Average duration on waiting list	4	2	1	<b>8,0</b>
	Short-term absenteeism personnel	3	2	1	<b>6,8</b>
	Long-term absenteeism personnel	3	2	1	<b>6,8</b>
	Frequency absenteeism personnel	3	2	1	<b>6,8</b>
	ZZP-mix on waiting list	4	1	0	<b>6,2</b>
	Average duration of ZZP	2	1	0	<b>3,7</b>
	Number of mutations	2	0	1	<b>3,1</b>
	Types of mutations	2	0	1	<b>3,1</b>
	Causes of mutations	2	0	0	<b>2,5</b>
	Bed capacity	0	1	0	<b>1,2</b>

The interviewed experts agree on the information needs for the financial state of the care institution. In the table above the most important information needs are listed on top. Also for these information needs the distinction is made based on the size of a care institution. The same rules apply to this distinction and can also be found in 'Appendix XX – Information needs by size'. Based on the size of a care institution, no clear difference could be extracted.

The information about a care institutions ZZP-mix could be identified as the most important information. Staffing with respect to ZZP-mix, current and forecasted ZZP-mix, operations per ZZP and information about the care within and outside the ZZP indication are the most mentioned information needs. Not only is this information important for care institutions, also stakeholders are very interested in this information.

Production information is another very important information need of care institutions. The production is declared monthly to the care offices and is the financial basis of every care institution. Differences in the planned, realized and declared production helps the directors to control the internal business. Every client has a care plan. This care plan contains the production that the client desires, that is the planned production. When care givers give the desired care to the client, the care is realized. Finally, the care institution should get financed for the realized care. Based on the information about the realized care, a message is send to the care office which contains a complete overview of the realized care. The care office assesses is the declaration is correct and will pay the care institutions. For care institutions it is important to receive money for all the realized care. In practice this is not always the case, therefore it is important to know the differences. Also unplanned care is identified as important information. Clients that receives every week a certain amount of time unplanned care should get a new care plan. Or if the desired care is more than is indicated, the client should request a higher indication.

Not all operations may be carried out by all care givers. This means that qualified personnel should be planned in order to carry out the planned operations. Information about the staffing with respect to the operations is important to comply with the laws. But, this information could also be used to increase the efficiency by creating an optimum. This means that the minimal qualified personnel is used to perform the planned operations.

Occupancy information is used to control the intramural care. An occupancy of 100% is the target and the experts are interested in their past, current and forecasted occupancy. This information could help the care institutions to detect weak spots in their organization on which they can take action.

A waiting list is an insurance for future work. Because it is impossible to forecast how old people will get, it is important to have a backup. It sounds very inhuman, but for care institutions it is important to pursue the 100% occupancy to maintain their position. Information about the waiting list consists of number of clients, ZZP-mix and duration on the waiting list, per business unit.

Absenteeism of personnel is information that is not managed from an ECR. This information is nevertheless interesting, not only for care institutions but also for the IGZ. IGZ is interested in this information because a high percentage of absenteeism or a high frequency of absenteeism could be an indicator for unrest in the institution. This could be an indicator for the IGZ to visit the care institution. For care institutions the same reason applies, this information must also be inserted in the annual report.

All clients that receive AWBZ care possess an indication that contains the intensity of care. Three experts are interested in information about the average duration of a client within a certain care intensity. This information enables them to create a strategic planning. Over time the intensity of care shall increase for most clients. How fast this process continues is unknown at this moment.

Information about mutations consists of mutations in the type of care, intensity of care, number of deceased and number of new clients. Mutations are communicated with the care offices by sending an AZR message. This AZR message contains information about the mutation. For example if a client dies, the care should be ended. The care institution will

end the care by filling in the date of death. By filling in this date, all care activities will be stopped in the system. This mutations must be communicated to the care office by sending an AZR message. This AZR message contains information about the client, care and the cause of the mutation. The care office will reply on this message to confirm the mutation.

The bed capacity is mentioned by one expert. This information is a key figure that could be extracted very easily from the ECR database. A basic functionality of ECR software is the client administration. This functionality enables care institutions to connect a client to a room. Every room has a certain number of beds. Without the proper knowledge about the number of beds per room and the number of rooms in the location it is impossible to maintain a proper client administration.

### Conclusion

The interviewed experts of Dutch long-term care institutions do agree upon the general information needs. Basic information such as customer experience, staffing, ZZP-mix, incidents, and production. These information needs are mentioned in many interviews. This basic information is used to manage care institutions and is requested on a monthly or even yearly basis. The top 10 information needs that have emerged from the interviews is listed in the following table, where Q represents the Quality of care information needs and F represents the Financial state information needs.

TABLE 4: INFORMATION NEEDS TOP 10

		(Board of) directors	Management	Stakeholders	Score
Information needs – Top 10	Q Customer experience	8	4	3	<b>16,6</b>
	F Staffing with respect to ZZP-mix	7	4	2	<b>14,8</b>
	F ZZP-mix (per business unit)	7	4	0	<b>13,6</b>
	F ZZP-mix prognosis	7	4	0	<b>13,6</b>
	F Staffing with respect to operations	6	4	2	<b>13,5</b>
	Q Number of incidents occurred	6	4	2	<b>13,5</b>
	Q Types of incidents occurred	6	4	2	<b>13,5</b>
	Q Causes of the occurred incidents	6	4	2	<b>13,5</b>
	F Operations per ZZP	7	3	1	<b>13,0</b>
	F Production information (planned, realized, declared)	7	3	1	<b>13,0</b>

Customer experience is the most mentioned information needs of both experts and stakeholders. This information is used to increase the quality of care and improve the quality of life. Customer experience is (mostly) measured by an external company.

Staffing with respect to the ZZP-mix is the most important indicator for directors in order to control the expenditure and revenue. Staffing information gives insight in the expenditures, whereas the ZZP-mix is important information for the directors to control the revenue. Forecasting the future ZZP-mix is also mentioned by many experts. The consequences of changes in laws and regulations can than directly be discovered by care institutions.

Information about incidents is used to increase the quality of care. Especially the causes of incidents could lead to improvements that increase the quality of care. All the qualitative information gathered from the interviews is made explicit and visible, resulting in the quantitative information presented in the tables.

But, based on the qualitative interviews it is also possible to make some added conclusions. Most directors of care institutions want to put the responsibilities as low as possible to reduce overhead. This results in the fact that most directors are primarily interested in financial information. And of course they also want some insight into the number of incidents, complaints and the user experience.

This leads to the observation that the financial related information of care institutions is evaluated more elaborate by the board of directors, than the care related information. This is most likely due to the competition that has been created by the Ministry of Health, Welfare and Sport in the healthcare sector. Care institutions are being paid by public money from the AWBZ and WMO law, which makes the room to make profit very narrow. Each client is obliged to have an indication from the CIZ to be entitled to receive a certain amount of care. The amount of care is linked to the ZZP type. For each ZZP type a fixed fee is in place. Care institutions should work more efficient in order to make more profit. A more efficient and effective care process can be achieved through better insight in the data.

During the interviews, the experts indicated that the prediction of different scenarios could be of high added value. This means that they would like to have the ability to predict different scenarios themselves. For example, residence for low ZZP's will not be financed by the AWBZ in the near future. This means that a care institution wants to predict what this means for their current operations. With this information it is possible to respond at forehand to the changes in the field.

Some interviewees also indicated that the application of technology should not go too far. This refers to the personal aspect of delivering care. Care personnel is educated and trained to deliver the best possible care. By improving the application of technology it could lead to a less personal approach. Because the delivery of care should have a personal approach, the application of technology should not get out of hand.



#### 4.1.2. DATA MINING GOALS

The information needs, that emerged from the interviews, are translated into data mining goals. Each data mining is based on one or multiple information needs, which can be found in 'Appendix XXI – Data mining goals vs. information needs'. The data mining goals will be used to discover the potential use of the collected data of care institutions. Data of one care institution will be used to understand the collected data and to test the use of the data for data mining applications.

The data mining goals are represented in the following subsectors, which are the translation of the information needs that emerged from the qualitative in-depth interviews. The extent to which these data mining goals can be achieved, will be discovered by applying data mining techniques to the available data. No data mining goals are created for the information needs that are not related to the information that is captured in an ECR.

#### **Quality of care**

For the quality of care information needs applies that 5 information needs could not be translated to data mining goals for this study, due to the nature of the data that is needed for those information needs. The following information needs could not be translated to data mining goals.

- Customer experience  
*This information is gathered anonymous by a third party. This prevents to use the data in this study. Unlike the following information needs, this information is quantitative of nature and is suitable for data mining application. In the future ECR software could implement functionality to investigate the customer experience.*
  
- Number / Types / Causes of complaints from the clients  
*Complaints information is not registered in ECR software, which prevents from using this information for analysis. Also the number of complaints (1 or 2 per year) makes this qualitative information not suitable for data mining application.*
  
- Reports of confidential counsellor  
*The reports of the confidential counsellor are confidential, as the name already indicates. Reports of the confidential counsellor are also qualitative of nature and not suitable for the application of data mining algorithms.*

The other quality of care information needs are translated into the following data mining goals:

- 1) Identify the patterns in incidents  
*Patterns in incidents include the recognition of focus areas. For example, if a large number of fall incidents take place in the corridor, the care institution could improve the safety in the corridor. Another example is the correlation between different incident types. This could help to improve the quality of care.*
  
- 2) Identify the patterns in risk assessment  
*Risk assessment creates insight in the risks a client runs. Insight in the relationship between different risks could help care givers to support the clients with the best possible care. Patterns based on client information (like age, gender and desired care) or business unit information could also help care institutions to increase*

*the quality of care. One business unit may have more clients with an increased risk at decubitus than other business units.*

3) Identify relationships between risk assessment and incidents

*Care institutions perform risk assessments once in a while. By asking question the assessment identifies whether the client runs a higher risk at, for example, depressions, falls or problem behaviour. This data is used in order to prevent or at least reduce the consequences. With the identified relationships between the incidents and the risk assessment, care institutions could gain insight in the effectiveness of their care process.*

4) Identify the patterns in the number of clients

5) Predict the number of clients

*These two data mining goals relate to the total number of clients. A distinction could be made for different business units or type of care (intramural or extramural). The number of intramural clients will show a stable picture, whereas the number of beds is also stable. Extramural clients could show a rise or decline in the number of clients. Forecasting the number of clients could help care institutions to attract new personnel or build extra residential.*

6) Identify the patterns in care-related measurements

*During the stay in care institutions, clients will be measured on different measurements. These measurements could indicate points that need attention. This could increase the quality of care because care personnel could take immediate action in order to prevent incidents. Care related measurements consists of weight, blood pressure, heart rate, blood sucker, etc.*

7) Identify the relationship between care-related measures and incidents

*Care-related measures could be reasons for some incidents. Clients with a high blood pressure could, for example, have more decubitus incidents than clients with a normal blood pressure. Insight in this relation could increase the quality of care by taking the proper precautions.*

8) Identify the relationship between pharmaceutical use and incidents

*Pharmaceutical use is very important in care institutions. Almost every client uses pharmaceuticals which could be reasons for incidents. Clients with medication X could have more fall incidents than clients with medication Y. Another interesting use could be to identify the impact of new medication. New medication could lead to more incidents. Insight in this information could increase the quality of care by applying this knowledge direct to the core process.*

9) Identify the patterns in client related and non-client related time

*Client related time is the time that is used to deliver care to the clients, non-client related time could be seen as overhead. Some prefer to use direct client related time and indirect client related time. Non-client related time includes training, multidisciplinary consultations, etc. Insight in this information could help management to increase the efficiency of the internal business and to increase the quality of care, by increasing the percentage of client related time.*

- 10) Identify the patterns in obtained and not-obtained treatment goals  
*Treatment goals are set up to achieve something together with the client. Especially in GGZ and GZ institutions it is important to increase the abilities of clients. Patterns in obtained and not-obtained treatment goals could gain insight in the feasibility of the treatment goals, which could eventually lead to an increase in the quality of care and quality of life of the clients.*
- 11) Identify the patterns in the average duration of stay
- 12) Predict the average duration of stay  
*Care institution with different locations or departments can compare these locations or departments based on the average duration of stay. In some short stay departments the average duration of stay will be lower in comparison with other departments. In order to manage the internal business it could be important to have insight in the average duration of stay and its fluctuations. In line with the identification of the average duration of stay, it will add value to predict the future duration of stay per business unit in order to make strategic decisions.*
- 13) Identify the patterns in coercive measures / restraints  
*The number of coercive measures and restraints should be minimized. In order to do so it is important to have insight in the number of coercive measures and restraints and the reason behind it. In some departments take clearly more coercive measures or restraints than other departments, this will be a trigger for management in order to evaluate it and, if necessary, to take actions in order to correct this difference.*
- 14) Identify the age distribution of clients  
*Information about the age distribution of clients could help management to make strategic decisions. Especially the difference in age distributions between different locations or departments and the difference in age distributions between clients in the different statuses. It is important to know if the age distribution of clients on the waiting list is higher than that of the current clients.*
- 15) Identify the patterns in ADL information  
*Patterns in ADL information could indicate the progression or regression of clients. For example, if a client is able to climb stairs and a week later the same client is unable to do that, this means a regression for that client. If all clients deteriorate or progress, this could have effect on the deployment of care personnel. Insight in the patterns of such ADL information is therefore important to increase the quality of care and to create a more efficient care process.*

### **Financial state**

Also some financial state information needs could not be translated to data mining goals, due to the fact that the information is not registered in ECR software. The following information needs could not be translated:

- Staffing with respect to ZZP-mix
- Staffing with respect to operations

- Staffing (per business unit)
- Short-term / Long-term / frequency absenteeism personnel

*These information needs are not elaborated in this study, because this study focusses on information that is gathered in ECR software. Staffing information is not gathered in such software, which means that such analysis are outside the scope of this research.*

The following data mining goals emerged from the financial state information needs.

- 16) Identify & predict the ZZP-mix

*Every client that receive AWBZ financed care, possesses a ZZP indication. This indication describes the amount of care that a client needs and for which the care institution is financed. If a care institution delivers more care than is described within a ZZP, the costs for that care will not be paid by the care office. In order to make decisions about the deployment of personnel and to make a strategic plan for new investments it is important to have insight in the current ZZP-mix and a forecast for the future ZZP-mix. In order to estimate the impact of changes in rules and regulations it could be important to have insight in the ZZP-mix.*

- 17) Identify the relationships between operations and ZZP

*As stated by the previous data mining goal, it is important for care institutions to have insight in the ZZP-mix. All operations performed should fit within the amount of time prescribed in the ZZP indication. The number of and type of performed operations is assumed to be different for each ZZP indication. Identification of these relationships could care institution help to gain insight in the operations a client needs.*

- 18) Identify the patterns in production

- 19) Identify the patterns in financed and non-financed care

- 20) Identify the patterns in unplanned care

*Production information is the important information in order to have insight in the revenue. Production is planned with the client and is included in a care plan. The products that are delivered to the clients are registered as delivered production. Finally the production is declared at the care offices. Differences in the amounts of these different production information should be analysed. Identifying patterns adds value for care institutions to minimize differences and to increase the percentage of declared production in relation to planned and delivered production. This makes the relation between the above three mentioned data mining goals. The goal is to find the patterns or the trends in the production information and to minimize the amount of non-financed care and unplanned care.*

- 21) Predict the unplanned care

*In order to deploy the right amount and qualified personnel it is important to take the unplanned care into account. A prediction of unplanned care could be based on historical data that is present in the database of ECR software. Management of care institutions could use this information in order to create the most efficient staffing.*

22) Predict the occupancy rate

*All the rooms and beds within a care institution should be provided with clients. A room that is empty costs money. Information about the occupancy rate could help management to increase the effectiveness of the rooms and beds. Also insight in the differences between locations and departments could help management to increase the occupancy rate. Very complex analysis are possible for this goal because of the information about waiting list and the abilities to make distinctions between locations and departments.*

23) Identify the patterns in the waiting list information

*Waiting list is before called as the insurance for care institutions. It is crucial for care institutions to have clients on the waiting list in order to maximize the occupancy and also revenues. Patterns and trends within the waiting list information in relation to the progress of clients can add large value for management. Fluctuations in the number of deceased must be absorbed by new clients.*

24) Identify the average duration of a ZZP

*This data mining goal is an extension of the progress of clients. For strategic decisions it can be important to know how long an average client will stay within a certain ZZP indication. Clients who need more or heavier care, require a different kind of care. This has consequences for the core process of care institutions. Knowledge about the average duration of a certain ZZP indication could help care institutions to control their internal business.*

25) Identify the patterns in mutations

*All the changes inflicted with the client are called mutations. All these mutations are reported to the care office using an AZR message. This also includes the changes in a client's indication. A high number of mutations means a lot of variability. Management could use this information to respond quickly to these variations.*

## 4.2. DATA UNDERSTANDING

### 4.2.1. EXPLORE & DESCRIBE DATA

The dataset that is used for the data analysis part of this study contains a large amount of data. All this data is captured in a MySQL database. MySQL Workbench will be used to query the data, and R Studio and Microsoft Excel 2010 will be used to create the analysis. R Studio is a free and open source integrated development environment for R (R Studio, 2012). R is a language, that includes a wide variety of algorithms and graphics for statistical computing (R Project, 2012). R Studio is used to create the complex analysis, whereas Microsoft Excel 2010 is used to create some basic graphics.

The data from the VVT care institution, used for this case study, is gathered with ECR software from the end of 2006. In February 2008 the care institution migrated to the current ECR software, ResidentWeb. Normally, the data mining goals are determined before the data is gathered. For this research the data has already been gathered. The extent to which data is available to satisfy the data mining goals, is researched in this study. The care institution consists of 5 locations which will be used and named in this study as follows:

TABLE 5: CARE INSTITUTION LOCATIONS

Location	Colour	Number of
1	Blue	142
2	Red	52
3	Yellow	150
4	Black	81
5	Green	34

An important observation is the change in laws and regulations. Because of these changes there will be differences in the availability of data. Some data mining goals could be based on more historical data than others, because of these differences. Changes in the laws about care have a direct impact on the value of historical data, and thus on the forecasts made based on this historical data. Because care is a public provision, the laws and regulations are prepared by the Dutch government.

Also changes within care institutions itself could immediately affect the value of historical data. Care institutions could for example change their care method. Different care methods exist, SAMPC method, Neuman Systems Model, 4 domains of nursing method, etc. Data collected using the Neuman Systems Model is different than the data collected using the SAMPC method. This makes it difficult, or even impossible, to create valuable analysis based on the information gathered for a particular care method.

A third cause of change is the lack or introduction of new functionality in ECR software. ECR software is relatively new in the care sector. New functionalities are implemented and improvements of already available functionality could have consequences on the data that is available.

Returning to the purpose of this study, it is not the goal to achieve all the described data mining goals, the purpose is to increase the efficiency of the collected data by identifying weak spots in the available data. The enormous amount of data collected by care personnel could be used for management to gain understanding of their internal business and to, if necessary, adjust their strategic plans.

In this section a first overview of the available data will be presented. Some basic information is visualized and presented. Because of confidentiality and privacy issues, a complete description of the database is not presented. With the first exploration of the dataset, some of the information needs that were mentioned during the interviews are satisfied already.

### Number of clients

In this sector, a first exploration of the data will be described. The first exploration consists of the number of clients, for whom the data is collected, in the database. The dataset contains information about 3704 clients in total. Clients could be recognized by their status, that is given in the ECD. The definition for the statuses are:

- **New client**  
When a new client is registered in the ECD, the client gets the status “New client”. Clients with this status do not receive care and are not listed on the waiting list.
- **Waiting list**  
When the care institution has no place for a new client, the client will be placed on the waiting list. Clients with this status do not receive care.
- **In care**  
The client has the status “In care” when the care institution delivers care to the client.
- **Temporarily out of care**  
This status is used for clients that are temporarily out of care. One can think of hospitalization or vacation.
- **Out of care**  
The “Out of care” status is used for clients that do not need care anymore, excluding the clients which are deceased.
- **Deceased**  
Deceased clients get the status “Deceased”.

In the following table, the number of clients per gender and status are elaborated for all the clients in the dataset and their current status. Also the percentages, shown as the percentage of the total number of clients per status, are displayed in the table. As expected, the percentage female in this care institution is much higher than the percentage of male.

TABLE 6: NUMBER OF CLIENTS PER GENDER AND STATUS

Type	Male	Female	Not available	Number of entries
New client	18 (29,03%)	42 (67,74%)	2 (3,22%)	62 (100%)
Waiting list	52 (27,51%)	137 (72,49%)	- (0,00%)	189 (100%)
In care	200 (23,87%)	638 (76,13%)	- (0,00%)	838 (100%)
Temporarily out of care	5 (27,78%)	13 (72,22%)	- (0,00%)	18 (100%)
Out of care	492 (27,56%)	1287 (72,10%)	6 (0,34%)	1785 (100%)
Deceased	260 (32,02%)	552 (67,98%)	- (0,00%)	812 (100%)
	<b>1027 (27,73%)</b>	<b>2669 (72,06%)</b>	<b>8 (0,21%)</b>	<b>3704 (100%)</b>

### Average duration of stay / waiting

One table in the dataset registers in which room the clients reside, and contains 1823 rows. With this gathered data it is possible to create an overview of the average duration of stay in a particular room. A room is linked to a location, it is

therefore not possible to generate the average duration of stay per department. Due to the limited number of rows in this table, it is also possible to use another table in which the clients are linked to a department. This table contains 5204 rows with data. The clients receive care at a department of the care institution. Based on this information it is possible to gain insight in the term a client is customer of the care institution. This means that also the extramural clients are then included, so it will then no longer be the average duration of stay.

The waiting duration per department is also an interesting figure, as it is mentioned by, for example, 4 experts from the (board of) directors level. One table, with 1315 rows, in the dataset contains information about the waiting list, containing the client, department, product, start date, end date and the priority. This data could be used to gain insight in the average waiting duration per location, department and product.

### Treatment goals

Every client could indicate their own treatment goals during their stay in a care institution. These goals are part of a care plan and are evaluated periodically. The complete dataset consists of 20725 records with treatment goals which are mapped onto a domain. Currently, most care institutions follow the 4 domain method. The care institution used for this case study also uses the 4 domain method. This method consists of 4 domains in which treatment goals could be created. The 4 domains are: Health (Gezondheid), Participation (Participatie), Wellbeing (Welbevinden) & Housing / Living conditions (Woon-/leefomstandigheden). In Figure 3, the treatment goals are mapped on the 4 domains. The blue bars are the goals that are not achieved, the red ones are the goals that have been achieved.

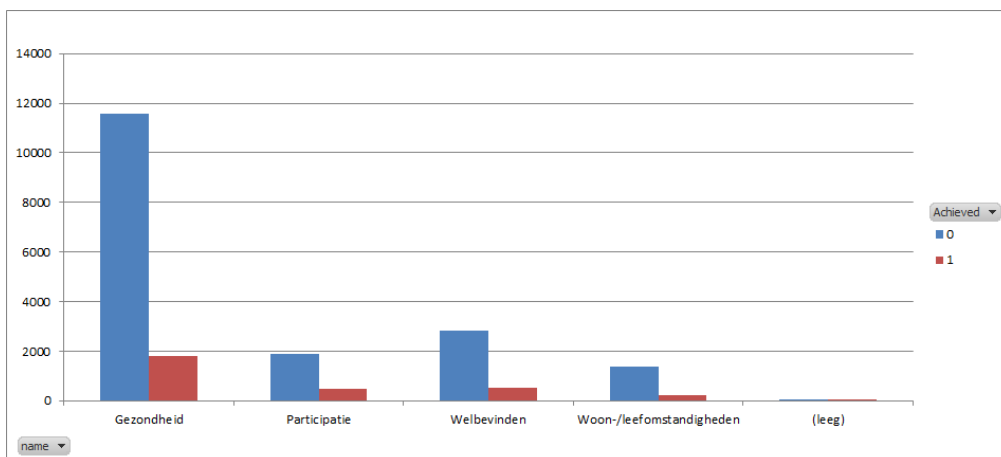


FIGURE 3: TREATMENT GOALS

Also for this visualization the raw data is used, without selections or transformations. From the above figure it is clear that most treatment goals are related to the health of the clients. Another interesting insight from this figure and data is that the majority of treatment goals are not achieved. In total 85,41% of the treatment goals are not achieved. This percentage of achieved treatment goals is for every status almost the same. That means that the percentage of achieved treatment goals for clients with the status 'Deceased' is almost the same as the percentage of achieved treatment goals for clients with the status 'In care'.

As indicated before, the dataset contains 20725 different treatment goals with 7229 different descriptions. The large number of different descriptions makes it very hard, if not impossible, to create complex analysis of the goals. This makes it impossible to gain insight into goals that are likely to be achieved or likely to be not achieved. Information



about the feasibility of a goal could be of added value for care institutions and care personnel. With the current collected data it is very hard or even impossible to determine the feasibility of treatment goals.

### **ZZP information**

Information about the ZZP types of clients is very important information for care institutions. The type of ZZP relates to the amount of care a client needs and to the fees that the care institution receives. In the dataset the clients are linked to a particular ZZP, with a start and end date. This makes it possible to create the current ZZP-mix, but also the ZZP-mix over time. This information could be used to predict the future ZZP-mix, to gain insight in the future ZZP-mix based on the historical data. In total 1831 rows with information about a client's delivered ZZP could create insight in the progression of the various ZZP's within the care institution.

### **Risk assessment**

Every care institution could perform a standardized risk assessment among their clients. The dataset used for this study also contains information about the risk assessment. Care institutions assess the clients risk multiple times over time. The used dataset contains various tables in which the risk assessments are stored, for each risk assessment one table. This makes the following risk assessments and number of entries:

- Depression (2129 rows)
- Falling (889 rows)
- Incontinence (877 rows)
- Medication (806 rows)
- Problem behaviour (0 rows)
- Weight extramural (0 rows)
- Weight intramural (567 rows)

For the risk assessment about weight there exist 2 different questionnaires, one for the intramural clients and one for the extramural clients.

### **Incident information**

Besides the number of clients, it is interesting to know how many incidents are registered in the used dataset. It is mandatory to select an incident type, when a new incident is registered. The ECD proposes 13 different kinds of incident type and the option to register another type of incident. In the following table, the first column represents the different types of incidents where the other columns represent the locations of the care institution. The cells represents the number of incidents registered in the dataset. The dataset contains 6126 registered incidents. Directly noticeable is that the number of medication and fall incidents are the most common incidents. Further analysis must demonstrate if interrelationships are present.

TABLE 7: NUMBER OF INCIDENTS PER TYPE

<b>Type of incident</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Total</b>
(intake of) harmful substances	-	-	-	-	-	-
Aggression / Harassment	19	19	48	1	-	87
Burn / Scorch	-	1	2	-	-	3
Behaviour	-	-	-	-	1	1
Medication	650	175	263	180	15	1283
Unsafe situation	11	6	13	-	-	30

Prick incident	-	-	1	-	-	1
Bumps / Pinch / Clash	12	5	7	2	2	28
Suicide	-	-	-	-	-	-
Fall	1554	1061	1457	266	104	4442
Loss / Theft	-	1	-	-	-	1
Loss / Theft (reported to the police)	-	-	-	-	-	-
Missing resident	3	6	40	1	-	50
Other	85	47	59	4	5	200
	<b>2334</b>	<b>1321</b>	<b>1890</b>	<b>454</b>	<b>127</b>	<b>6126</b>

### Care-related measures

Care-related measures could show trends which lead to incidents or increased risk. This dataset contains information about the care-related measures. One table contains 27174 rows with valuable care-related measures, including the clients weight, blood sucker, blood pressure, heart rate, length and temperature. Also the date and time of the measurement is being captured in the table.

### Pharmaceutical use information

Information about the use of pharmaceuticals is very important for care institutions. All clients should take the right medicines. The dataset used for this case study does not contain this information, because this information is managed by the external practitioners of the clients.

### Coercive measures / restraints

Clients with an article 60 statement (BOPZ status) are involuntary clients. These clients are a danger for themselves or their environment. For these clients a care institution could use coercive measures and restrains. All these coercive measures and restrains should be registered. The dataset contains information about 25 coercive measures / restrains over the past 2 years. The reason for applying the coercive measure / restraint could be registered using a free text field. It becomes hard to create valuable insights, because of this lack of standardization.

### Production information

The dataset contains data from a long period, this includes the production information. Production information is registered in one table with information about the client, product, duration, date of production, department and type of financing. The dataset consists of 1868143 rows with production information from January 1, 2006 until May 1, 2012. That means that the table contains production information in the future. This is most likely caused by the automatic production generation of the software. For each product a pattern could be linked, this pattern is used to register the production. This prevents the care givers from registering every task performed for the client, and also increases the data quality. Whenever the real production differs from the automatically generates production, the care institution changes the registered production.

#### 4.2.2. VERIFY DATA QUALITY

Verifying the data quality of the dataset is difficult. All the data from the dataset is production data, which means that the care institution employees generated the data and use the data on a daily basis. Because the data is widely used within the care institution, the quality of the data should be in order. Because the data is gathered for clients, it is hard to identify direct quality issues. Of course, invalid dates could not be used for analysis, and measures (for example weight and length) could be checked with common sense. All the analysis done in the next chapter are based on selections from the complete dataset. Most selections are based on time, which means that only entries that have been registered between two dates are selected. With these selections the invalid values with regard to the date values are not selected and will be left out of the selections.

Other data quality problems consist of lack of standardization and thus lack of homogeneity in the data. A lot of data is collected by filling free text fields, which results in heterogenic data. Every care giver interprets the clients situation different, which results in different registration. It is difficult to extract new knowledge from heterogenic data.

Also the complexity of the data increases the difficulty to gain new knowledge. The long-term care sector must adhere to a large number of laws and regulations. These laws and regulations change over time which decreases the usability of the historical data. For example, historical financial data could not be used as a reference if a new Dutch government decides to change the way how long-term care is financed. The influence of the Dutch government is caused by the way of financing the long-term care. Long-term care is financed by public money, which makes the Dutch government responsible for the long-term care.

Another difficulty with the gathered data is missing data. Because the data is gathered during the core process of a care institution, there is no knowledge about the amount of missing data. For example, incidents that occurred but have not been registered could not be used during the data analysis.

Due to the described difficulties it is impossible to create a clear view on the data quality. It is possible to gain insight in the data quality of the gathered data, but that does not gain insight in the overall data quality because of the unknown amount of missing data. In the next chapter, the data quality issues are elaborated per data mining goal, if present.

### 4.3. DATA PREPARATION & MODELLING

In this section all the data mining goals are presented with the available analysis. Not all the data mining goals could be carried out, due to unavailability of the data, lack of standardization or just too complex data. The dataset contains a lot of data that could be used for data mining as it is described in the previous sector. Again, the purpose of this study is not to achieve all the described data mining goals, the purpose is to increase the efficiency of the collected data by identifying weak spots in the available data.

#### (1) Identify the patterns in incidents

Every incident should be registered in the ECR software. The dataset contains 6126 incidents, which includes the client, department, date and time, type of incident, cause, location, physical damage and mental damage. This collection of data is very valuable, and could be used for various analysis. First of all, the incidents for which a client, department, date and time, type of incident and location are registered are selected, which result in a collection of 5692 incidents. Figure 4 visualize the number of incidents over per month, over the years. The first incident is registered in April 2008, the first month with more than 30 incidents was July 2008. This month will be taken as the start date for further analysis. The dataset contains information until April 17, 2012, which is the cause of the low number of incidents in April 2012. Figure 4 shows an upward trend, this does not mean that more and more incidents happen. Better registration is most likely the cause of this trend.

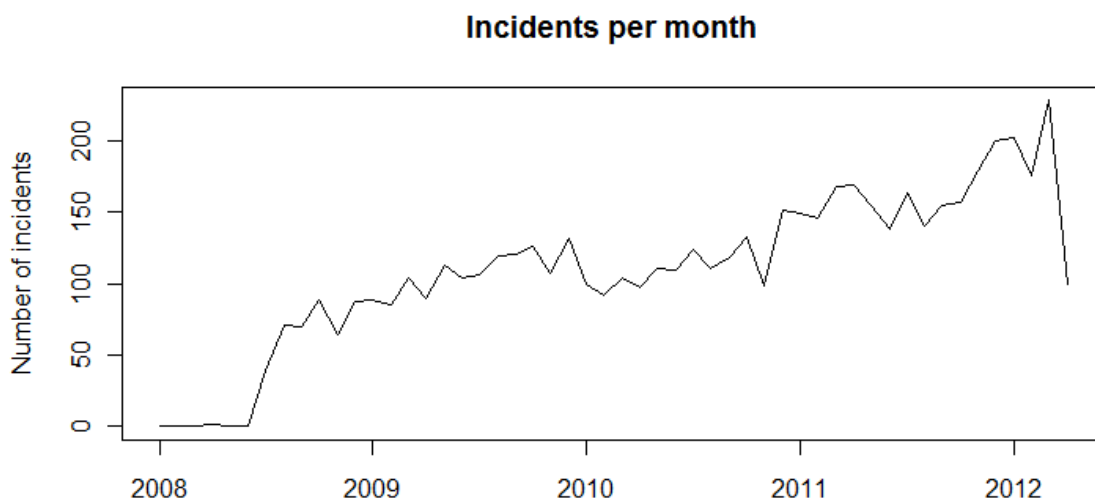


FIGURE 4: NUMBER OF INCIDENTS PER MONTH

A first interesting figure is the number of incidents with respect to the time of the incident. This creates insight in the critical moments during the day, on which many incidents occur. For this analysis, all the incidents per hour are counted using a query. This query counts the number of incidents between, for example, 0:00 and 0:59.

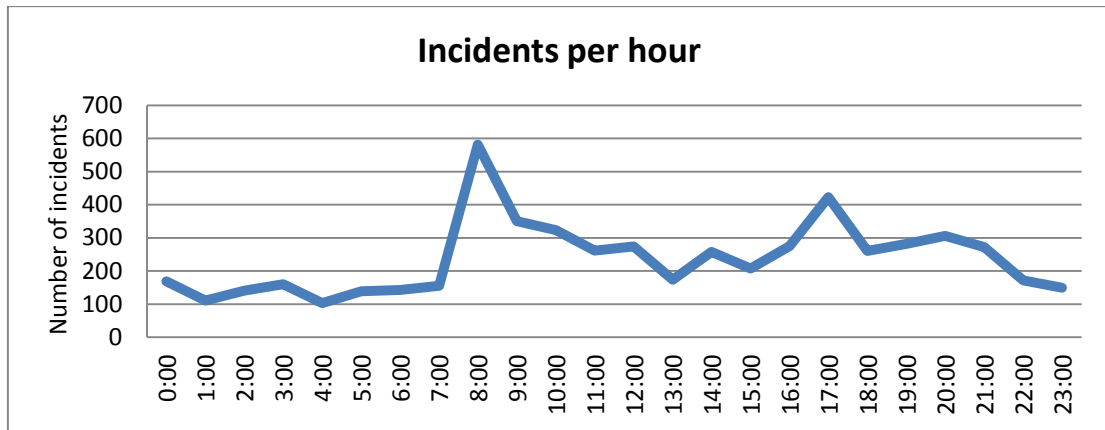


FIGURE 5: NUMBER OF INCIDENTS PER HOUR

Figure 5 visualizes the number of incidents per moment on the day. Most incidents occur during the day, between 8 am and 9 pm. The peaks between 8 and 9 am and between 5 and 6 pm are probably caused by the transfers of the clients (out of bed and diner).

Insight in the differences in number of incidents, per hour, per location is also interesting, and is visualized in Figure 6. The colours represent the locations, as defined in the previous chapter. All locations show the same peaks between 8 and 9 am and between 5 and 6 pm, similar to the overall incidents (Figure 5).

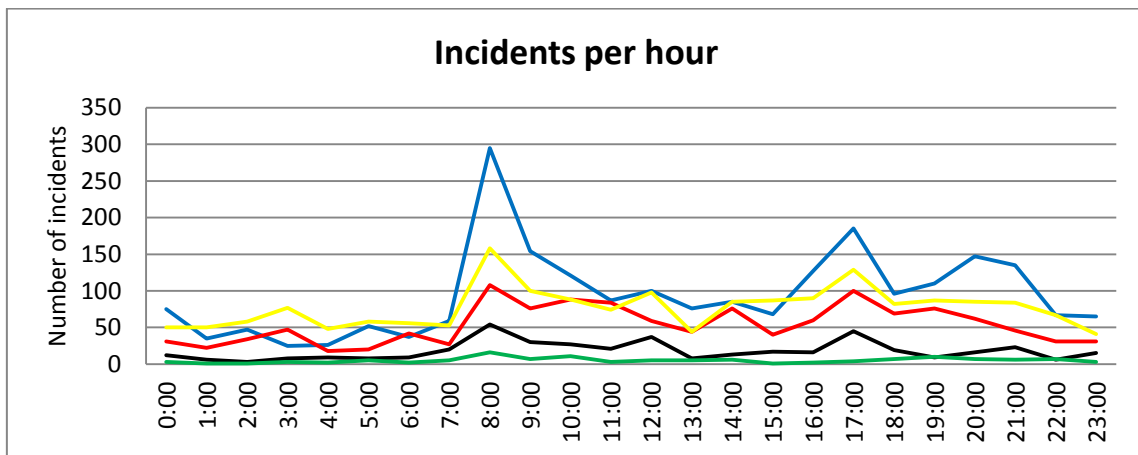


FIGURE 6: NUMBER OF INCIDENTS PER HOUR PER LOCATION

Another interesting analysis, is how the incidents are divided over the days of the week. For this analysis it is important to group all the incidents for the day of the week. A query is used to select this information from the dataset. Figure 7 visualizes this information, which gains the insight that the number of incidents is almost the same for every day. Because the number of incidents is evenly spread over the week there is no reason for the care institution to deploy more personnel on, for example, a Saturday.

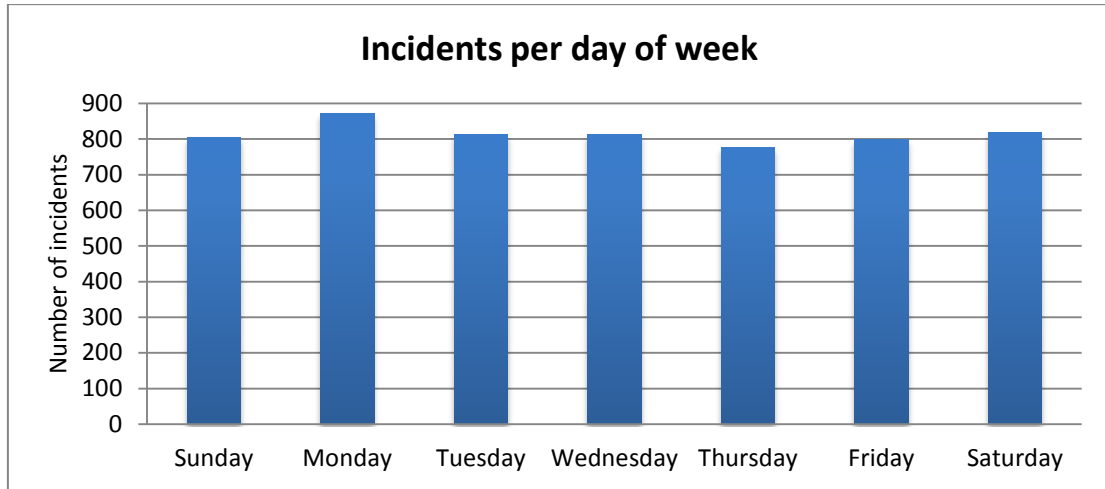


FIGURE 7: NUMBER OF INCIDENTS PER DAY OF WEEK

Also the location of the incidents is registered, which could be used to detect problem areas in the care institution. If most incidents take place in the corridor, it could trigger management to research this fact and to increase the safety in the corridor. All the 5692 incidents are selected and formatted based on the incident location and the care institution location, in the following table.

TABLE 8: NUMBER OF INCIDENTS PER INCIDENT LOCATION

	1	2	3	4	5	Total
Activities room	11	0	1	10	0	<b>22 (0,4%)</b>
Other department	8	6	1	1	0	<b>16 (0,3%)</b>
Bathroom	126 (6%)	77 (6%)	152 (9%)	43 (10%)	17 (14%)	<b>415 (7,3%)</b>
Outside the building	24	12	52	11	1	<b>100 (1,8%)</b>
Restaurant / Dining room	23	14	15	1	2	<b>56 (1,0%)</b>
Somewhere else in the building	36	24	19	22	2	<b>103 (1,8%)</b>
Corridor	101	72	92	9	3	<b>277 (4,9%)</b>
Shared living room	3	5	1	0	0	<b>9 (0,2%)</b>
Kitchen	225 (10%)	154 (12%)	110 (6%)	46 (11%)	10 (8%)	<b>545 (9,6%)</b>
Bedroom	524 (24%)	278 (22%)	575 (33%)	38 (9%)	31 (26%)	<b>1446 (25,4%)</b>
Toilet	62	45	54	5	8	<b>174 (3,1%)</b>
Stairs	2	1	0	0	0	<b>3 (0,00%)</b>
Living room	895 (41%)	500 (40%)	587 (34%)	231 (55%)	40 (33%)	<b>2253 (39,6%)</b>
Other	120	63	78	5	7	<b>273 (4,8%)</b>
	<b>2160</b>	<b>1251</b>	<b>1738</b>	<b>422</b>	<b>121</b>	<b>5692 (100%)</b>

In table 8, the top 3 locations for incidents are shaded, which visualizes the most common locations for incidents. For all the care institution locations means that the most incident take place in the living room. The other locations where incidents commonly happen are: bedroom, kitchen and the bathroom. For these (problem) areas the percentages per location are described, which makes it possible to compare the locations with each other. What the consequences are of the incidents could be extracted by analysing the physical and mental damage to the client, after an incident. Which is elaborated in the following tables.

TABLE 9: MENTAL DAMAGE PER INCIDENT TYPE

	Anxiety	No	Not yet noticeable	Unknown	Disquiet	Drowsiness	Other	Total
Aggression / Harassment	5	49	7	11	10	0	3	85
Burn / Scorch	0	3	1	0	0	0	0	4
Behaviour	0	0	0	0	0	0	1	1
Medication	1	744	101	149	7	6	5	1013
Unsafe situation	2	19	5	0	2	0	0	28
Prick incident	0	0	0	1	0	0	0	1
Bumps / Pinch / Clash	1	18	3	3	1	0	1	27
Fall	199	2575	855	333	202	41	87	4292
Missing resident	0	32	7	1	4	3	3	50
Other	9	104	30	20	7	6	8	184
Unknown	0	4	1	2	0	0	0	7
	<b>217</b>	<b>3548</b>	<b>1010</b>	<b>520</b>	<b>233</b>	<b>56</b>	<b>108</b>	<b>5692</b>

Mental damage after an incident is the highest after an fall incident, based on table 9. The red shaded cells are the incidents without observed mental damage. This means that there is no observed mental damage, the mental damage is not yet noticeable, or the mental damage is unknown. The latter means that there is no value registered for the mental damage after an incident, this option to register the mental damage was not implemented. For (5078/5692) 89,2% of the incidents there was no mental damage registered. Falling incidents causes the biggest number of mental damage, but for only (529/5692) 12,3% of all falling incidents there was mental damage reported. For only (614/5692) 10,8% of all incidents there was mental damage reported. Because the mental damage terms are very interpretable, it is still unclear what the impact of this result is. Anxiety for example could mean that the client does want to leave the bed anymore, but it could also mean that the client is somewhat afraid to walk alone.

TABLE 10: PHYSICAL DAMAGE PER INCIDENT TYPE

	Bruises	Fracture	Burn	No	Not yet noticeable	Unknown	Pain	Cut	Muscle complaint	Intoxication	Sprain	Other	Total
Aggression / Harassment	3	1	0	56	11	7	3	1	0	0		3	85
Burn / Scorch	0	0	2	1	1	0	0	0	0	0			4
Behaviour	0	0	0	1	0	0	0	0	0	0			1
Medication	0	0	0	784	132	79	3	0	0	0		15	1013
Unsafe situation	0	0	0	22	5	0	1	0	0	0			28
Prick incident	0	0	0	0	0	1	0	0	0	0			1
Bumps / Pinch / Clash	5	0	0	2	1	0	0	16	0	0		3	27
Fall	256	46	0	1791	838	28	559	562	5	2	13	192	4292
Missing resident	0	0	0	44	5	0	1	0	0	0			50
Other	6	1	1	91	45	4	9	11	0	0	1	15	184
Unknown	0	0	0	3	1	0	2	0	0	0	1		7
	<b>270</b>	<b>48</b>	<b>3</b>	<b>2795</b>	<b>1039</b>	<b>119</b>	<b>578</b>	<b>590</b>	<b>5</b>	<b>2</b>	<b>15</b>	<b>228</b>	<b>5692</b>

Also in table 10 the red shaded fields represent the incidents without any physical damage. The same rules apply as for the mental damage table. For both the mental as the physical damage applies that after a fall incident the biggest damage is done, based on table 9Table 10 and table 10. For (3953/5692) 69,4% of the incidents there was no physical damage registered. Most physical damage is done after a falling incident, by (1635/4292) 38,1% of all falling incidents there was physical damage reported. The large number of falling incidents has a great impact in the average percentage. For (1739/5692) 30,6% of all incidents there was some form of physical damage reported, which is mainly caused by falling incidents.

Because the dataset contains a lot of valuable information about incidents, the number of analysis is very broad. The causes of incidents is also registered, but the number of different causes prevents good analysis. For 5692 incidents the dataset contains 1242 different causes. In order to create proper analysis, the number of causes should be limited. Another interesting analysis could be which employees are more confronted with incidents. The current data is not suitable for these analysis, because the names of the involved employees are entered by the user. This means that every value could be entered, which resulted in 2179 different values.

## (2) Identify patterns in risk assessment

Risk assessment is a relatively new functionality in the software, which is introduced in 2011. Therefore, the first risk assessment for a client is done on August 18, 2011. For the analysis in the table below, the last risk assessment of the current clients is selected, in order to form a reliable picture of the risk assessment within the care institution. A query, that selects all the clients with a current subscription and no date of death, is used to select the 492 clients from the dataset. A subscription means that a client has a subscription for receiving some kind of product. Every risk assessment should be performed individually, for 121 clients no single risk assessment is performed. These 121 clients are included in the numbers in the table below.

TABLE 11: RISK ASSESSMENT

	No increased risk	Increased risk			No risk
		no adequate follow-up	adequate follow-up	total	
Falling	117 (23,78%)	12 (2,44%)	233 (47,35%)	245 (49,80%)	130 (26,42%)
Incontinence	96 (19,51%)	24 (4,88%)	241 (48,98%)	265 (53,86%)	131 (26,63%)
Depression	101 (14,92%)	184 (37,40%)	83 (16,87%)	267 (54,27%)	124 (37,23%)
Medication	283 (57,52%)	10 (2,30%)	68 (13,82%)	78 (15,85%)	131 (26,63%)
Problem behaviour	-	-	-	-	492 (100%)
Weight Extramural	-	-	-	-	492 (100%)
Weight Intramural	241 (48,98%)	14 (1,55%)	61 (12,40%)	75 (15,24%)	176 (62,89%)

Depression is the most performed risk assessment under the current clients. This is also the most interesting risk assessment. For (368/492) 62,77% of the clients, a depression risk assessment is done. More than (267/368) 70% of these 368 clients, are at an increased risk of depression. For (184/267) 68,9% of these increased risk clients, there was no adequate follow-up. Adequate follow-up means that the care institution includes for example treatment goals into the care plan of a client, in order to decrease the risk of incidents.

Another interesting discovery is the high percentage of clients that do not run an increased medication or weight risk. Medication risk includes the risk of taking the wrong medication, no medication, cognitive impairment or impaired hand function. This means that the client is assessed on various aspects in order to determine if there is an increased risk. Because a large number of medication incidents occurred, it is interesting to discover that only 78 clients run an



increased risk. Weight risk includes weight loss, loss of appetite and the need for assistance during eating. Only 75 clients run an increased risk for underweight. Data about the care related measures (for example the client's weight) could explain how many clients are actual high-risk clients.

In order to find valuable patterns or relationships in the risk assessment data, association rule mining is applied. For learning association rules, the Apriori algorithm is used (Agrawal, Srikant, & others, 1994). This Apriori algorithm is widely used and available in R Studio. For all the analysis, a minimal support level of 0,1 and a minimal confidence level of 0,8 is used. These parameters mean that at least 80% (confidence) of the clients that for example have an increased falling risk also have an increased medication risk, and at least 10% (support) of the clients have both.

The relationships between the different risks are investigated, based on the last risk assessment of the current clients. This means that the follow-up of an increased risk outcome is not taken into account for the first analysis and discovery of association rules.

TABLE 12: ASSOCIATION RULES INCREASED RISK

Rule	Support	Confidence	Lift
Medication → Incontinence	12,98%	82,05%	1,526
Weight intramural → Falling	12,37%	81,33%	1,637
Weight intramural → Incontinence	13,18%	86,67%	1,612
Falling → Incontinence	43,61%	87,76%	1,633
Incontinence → Falling	43,61%	81,13%	1,633
Falling, Medication → Incontinence	11,76%	95,08%	1,769
Incontinence, Medication → Falling	11,76%	90,63%	1,824
Falling, Weight intramural → Incontinence	11,56%	93,44%	1,738
Incontinence, Weight intramural → Falling	11,56%	87,69%	1,765
Depression, Falling → Incontinence	31,64%	88,64%	1,649
Depression, Incontinence → Falling	31,64%	82,11%	1,652

Table 12 provides insight in the risks a client run. In the last column the lift of association rule is elaborated. Lift is computed with the following equation:  $lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)supp(Y)}$ . A larger lift means a stronger association. Both support, confidence and lift are interesting measures to discover the most interesting association rules (Hahsler & Chelluboina, 2010).

The most interesting association rule based on the lift, states that 90,63% of the clients with increased incontinence and medication risks have also an increased falling risk, almost 12% of the clients run an increased risk on all three. This is the association with the highest lift, and is thus the strongest association.

Increased falling or incontinence risks are part of every association rule. A large number of clients run an increased risk on these aspects. Also a lot of clients run an increased risk at depression, but depression is only part of 2 association rules. That means that depression is not very related to the other risks.

### (3) Identify relationships between risk assessment and incidents

Previous analysis turned out that falling incidents are the most common incidents with the most damage (both mental and physical). For this reason, only the falling incidents will be analysed for identifying the relationships between incidents and risk assessment. Clients with an increased fall risk, are likely to experience more fall incidents. The following analysis will tell whether this assumption is true.

Risk assessment is, as already said before, a relatively new functionality in the software, that is taken into operation on August 18, 2011. In order to find a relationship between risk assessment and the incidents, data is selected from September 2011 onward. The gap of 13 days is taken into account, because the risk assessment cannot be done for all the clients in one day. A created query selects all the fall incidents, since September 2011, and the outcome of the last risk assessment done before the incident. This selection results in 937 falling incidents and 361 medication incidents with the outcome of the last risk assessment.

TABLE 13: NUMBER OF INCIDENTS WITH RESPECT TO RISK ASSESSMENT OUTCOME

	Fall incidents		Medication incidents	
No increased risk	71 (7,58%)		124 (34,35%)	
Increased risk – no adequate follow-up	62 (6,62%)	555 (59,23%)	12 (3,32%)	66 (18,28%)
Increased risk – adequate follow-up	493 (52,61%)		54 (14,96%)	
No risk assessment	311 (33,19%)		171 (47,37%)	
<b>Total</b>	<b>937 (100%)</b>		<b>361 (100%)</b>	

Table 13 shows that the risk assessment does not always create a good insight in the risk group. For fall incidents the risk group is quite good mapped, but that is not the case for medication incidents. This is most likely due to the fact that most medication incidents happen because care personnel forgets to provide the medication to the client. The other types of incidents and risks could not be analysed, because there is no further overlap between the two. Problem behaviour risk assessment could be used to gain insight in the target group for aggression incidents, but the risk assessment is not performed. The usefulness of risk assessment can, therefore, not be determined.

#### (4) Identify patterns in the number of clients

Figure 8 represents the actual number of clients. The y-axis of figure 8 is different than that of subsequent figures, in order to discover the differences in the number of clients. Since February 2008, the care institution is working with ResidentWeb as the leading ECR software. Incorrect data from the past has been corrected before the introduction. This means that the peak in the total number of clients in 2006 contains incorrect data. Clients who were out of care already, were still registered as ‘in care’ clients. The peak at the end of 2008 could not be substantiated and will be taken into account for that reason. The dip at the end of 2011 is caused by a deliberate reduction of clients due to construction plans.

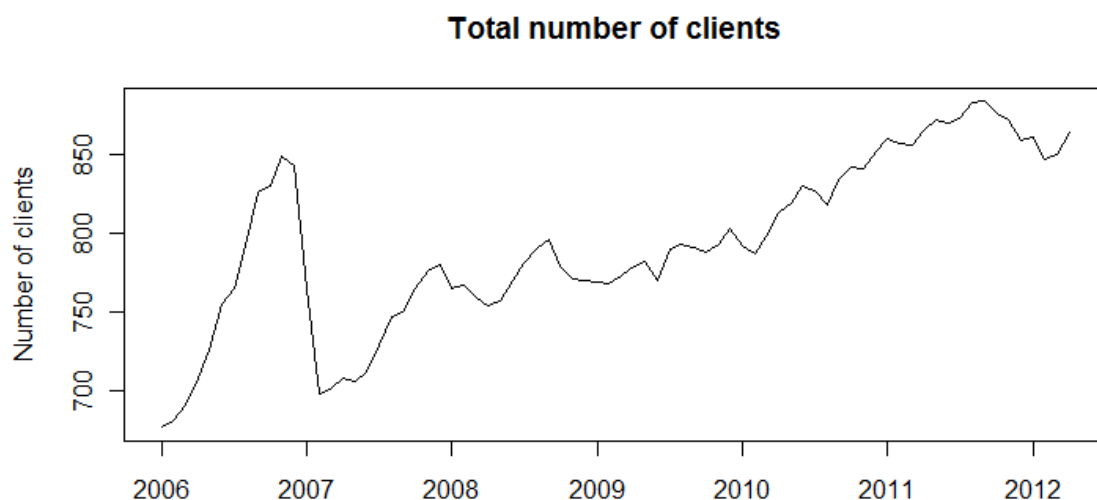


FIGURE 8: NUMBER OF CLIENTS IN CARE

The patterns and predictions for the number of clients, will be based on the data from February 2008 onwards. A query is constructed which counts the number of clients that receive care on the first day of every month, since 2008 and per location. This query shows the total number of clients that receive care per location.

Multiple plots have been created, based on data from the database, to gain insight in the patterns in the number of clients. Figure 9 shows the 5 different locations of the care institution. The colours represent the locations and the lines represent the number of total clients at the 1<sup>st</sup> day of every month. It is very clear to see the split-up of the red location, due to construction plans. This location has been split into two locations (red and green) in the first quarter of 2011. Also the black line represents a strange continues increasing line. This black line represent a new location, that has been taken over in January 2011. For the red location, the number of clients is deliberately reduced because of the construction plans. Only the blue location represents a decline in the number of clients, the other locations show a relatively stable line.

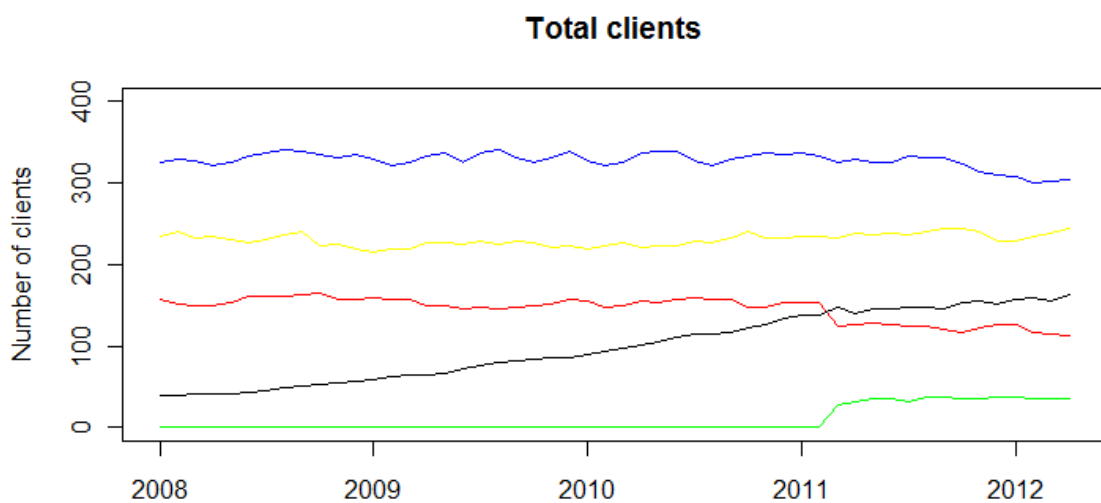


FIGURE 9: TOTAL NUMBER OF CLIENTS PER LOCATION

The query is extended with a filter that also looks if the client has resided in a room, to make a distinction between intramural and extramural. Because the number of intramural clients will show a stable line (no continues increase in the number of beds per location). In figure 10 the number of intramural clients per location can be found. Also in this figure it is visible that the red location has been split into 2 locations (red and green). All the lines represent a quite stable picture, which was already predicted. This means that the number of extramural clients should show a decline for the blue location.

### Intramural clients

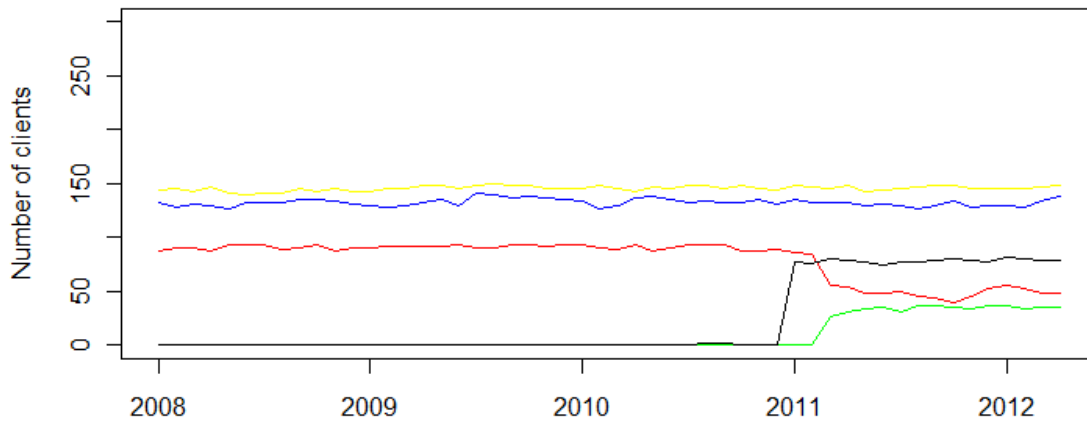


FIGURE 10: NUMBER OF INTRAMURAL CLIENTS PER LOCATION

The number of extramural clients is not limited to the number of beds, which could lead to a more clear increase or decrease in the number of clients. When the number of extramural clients increases, this will have consequences for the number of employees that is needed.

### Extramural clients

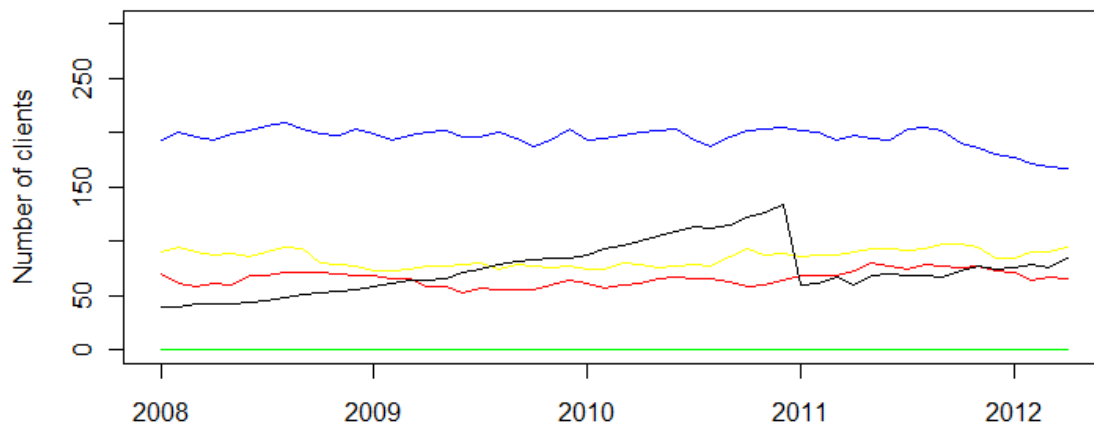


FIGURE 11: NUMBER OF EXTRAMURAL CLIENTS PER LOCATION

Also the number of extramural clients show a quite stable picture for each location. The green location does not deliver extramural care and is therefore the most stable line. For the black location applies that the clients were registered as extramural before the takeover. This means that care was already delivered to the client on an extramural basis. After the takeover, the clients that resides in the care institution were registered at the room in which they reside. This explains the decrease of over 50 clients for that location. The number of extramural clients is decreasing for the blue location.

#### (5) Predict the number of clients

Based on the information used for the previous data mining goal, a prediction will be made on the number of extramural clients per location. The number of intramural clients will remain almost the same of time, because of the number of available beds. Therefore it is more interesting to create a prediction for the future number of extramural clients. The

dataset is contains for each month the number of clients. This information is loaded into R Studio as a time series. R Studio contains a broad range of packages which enables one to create forecasts based on time series. For this prediction the automatic forecast() algorithm is chosen. This algorithm applies multiple different forecast algorithms to the data and selects the one with the lowest Akeike’s Information Criteria (AIC) score. This score calculates the in-sample fit, low AIC scores mean a better predicted model (de Gooijer & Hyndman, 2006).

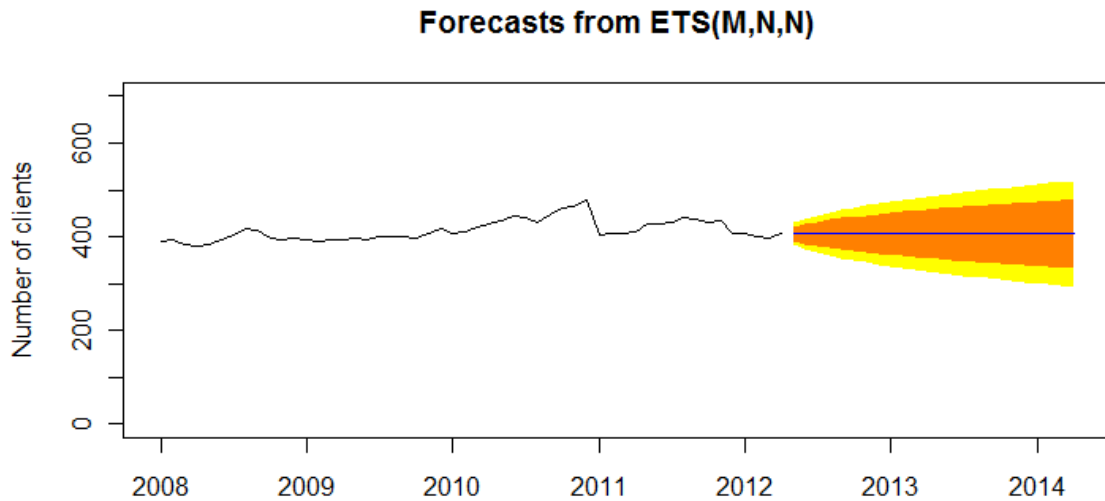


FIGURE 12: FORECAST FOR THE NUMBER OF EXTRAMURAL CLIENTS (TOTAL)

In figure 12 a forecast is visualized for the number of total extramural clients. The forecast() method picked the ETS method with multiplicative errors and no trend or seasonal component. This means that the algorithm does not discover a trend nor a seasonal component in the time series data. Because the number of extramural clients show a very stable line, the predicted number of extramural clients also show a stable line. The orange and yellow areas show the confidence intervals. For the orange area the confidence is 80% and for the yellow area this is 95%. This model predicts that the number of extramural clients in April 2014 will be between 292 and 518 clients at a certainty of 95%. With a certainty of 80% the model predicts that the number of clients in April 2014 will be between 332 and 479 clients. The point forecast for the whole forecasted period is 405 extramural clients.

As already visualized in figure 11, the number of extramural clients could also be split to locations. For each location a prediction can be made for the number of both intramural and extramural clients. Almost all locations show a stable line with regard to the number of extramural clients (Figure 11). Only the blue and black locations show respectively a decrease and increase in the number of extramural clients.

### Forecasts from ETS(M,N,N)

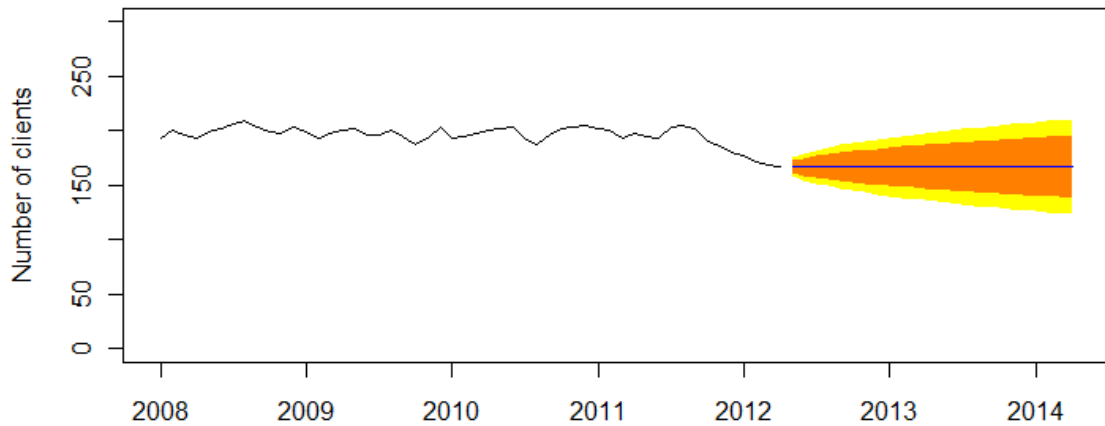


FIGURE 13: FORECAST FOR THE NUMBER OF EXTRAMURAL CLIENTS (BLUE LOCATION)

For this blue location the model predicts a point forecast of 167 extramural clients over the whole forecasted period. With a certainty of 80% the model predicts that the number of extramural clients will be between 138 and 195 in April 2014. For the 95% certainty the model predicts that the number of extramural clients will be between 123 and 210 clients.

### Forecasts from ETS(M,N,N)

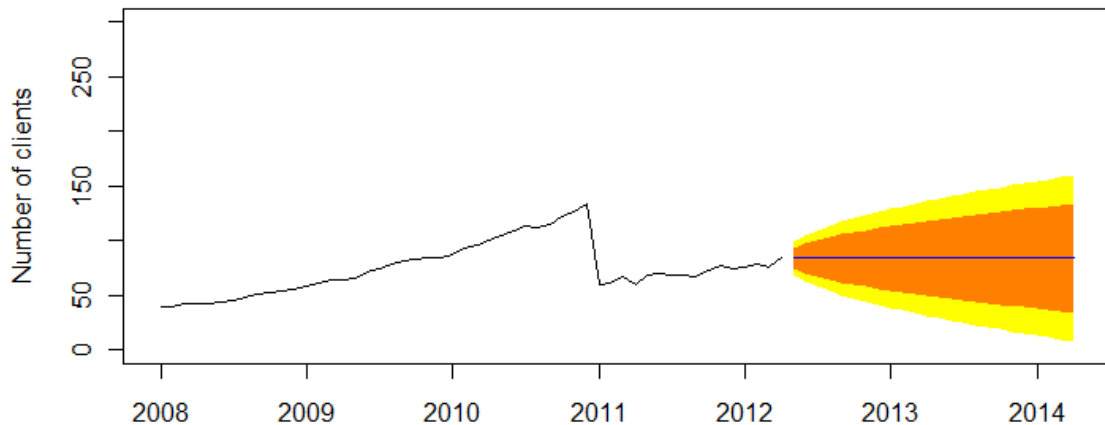


FIGURE 14: FORECAST FOR THE NUMBER OF EXTRAMURAL CLIENTS (BLACK LOCATION)

For this black location the model predicts a point forecast of 84 extramural clients over the whole forecasted period. With a certainty of 80% the model predicts that the number of extramural clients will be between 34 and 133 in April 2014. For the 95% certainty the model predicts that the number of extramural clients will be between 8 and 160.

For both the blue as the black location means that the prediction shows a stable line. The fitted model is the same as that for the total number of extramural clients. No seasonality and no trend could be discovered in the time series of the number of extramural clients for both locations. Also for these models means that the orange area represents the 80% confidence, whereas the yellow area represents the 95% confidence. The difference between the two models is the window of the predictions. For the black location the window is almost twice the window of the blue location, at the 95% confidence level.

#### **(6) Identify the patterns in care-related measurements**

Care-related measures are being registered from the start, which resulted in a very large table with 27174 rows of care-related measures. The care-related measures that could be registered are: blood sucker, blood pressure, heart rate, temperature, length and weight. However, the fields, in which the measures could be filled, are free text fields. This means that the software does not check whether the value is correct or incorrect. This results in lengths of 45 meter. Also many different employees are working with the software, which results in many different ways to type the same information (heterogeneity). The registered care-related measures are not suitable for proper analysis.

#### **(7) Identify the relationship between care-related measures and incidents**

Because the information is captured at a given moment in time, it is possible to create complex analysis by selecting the latest measure before a particular incident could gain insight in the relationship between the measures and the incidents. However, also for this data mining goal means that the registered care-related measures are not suitable for proper analysis. The quality and the homogeneity of the data is too low to create valuable analysis.

#### **(8) Identify the relationship between pharmaceutical use and incidents**

The dataset does not contain information about pharmaceutical use of clients. This information is mostly managed by the specific practitioner of the client (mostly the general practitioner). If care institutions register the pharmaceutical use information by themselves and use this information, they are responsible for having the correct information. That is the reason for care institutions to use the pharmaceutical use information from the practitioners. In order to create valuable analysis, a connection should be established between the two (or more) databases. This connection could not be established for this case study.

#### **(9) Identify the patterns in client related and non-client related time**

For this data mining goal time registration data should be available, in order to identify the patterns in client related and non-client related time. The dataset does not contain complex time registration information. All care plans are registered in which the operations are registered, including the default time to complete the task. This could be used to identify the time that is needed to perform all client-related activities. Every task is planned with the client, this means that one task could be performed daily whereas another task should be performed once in a month. The dataset contains, per task, the total time per week that is needed to perform the task. Data from the start of 2008 is available until April 2012, this is also the selected period of time. A query is constructed that sums all the time per week for the tasks that should be performed that week. This query returns the number of hours per week needed to perform all the tasks, rounded up (Figure 15). Because of the new location that has been introduced in 2011, the number of FTE needed to perform all the tasks has increased. The number of FTE needed has increased from 233, in the first week of 2008, to 349 in week 16 of 2012.

Another important consideration is the fact that ZZP indications include a time indication, with a margin. For example, ZZP 1 VV has a time indication between 3 and 5 hours care. That is an average time indication. Care institutions could plan care for all clients at the top margin (in this case 5 hours). The planned activities also contain an average time indication. The actual time needed to complete the activity is not gathered. Because the correctness and integrity of this data is very speculative, further conclusions with respect to care related time cannot be made based on this analysis.

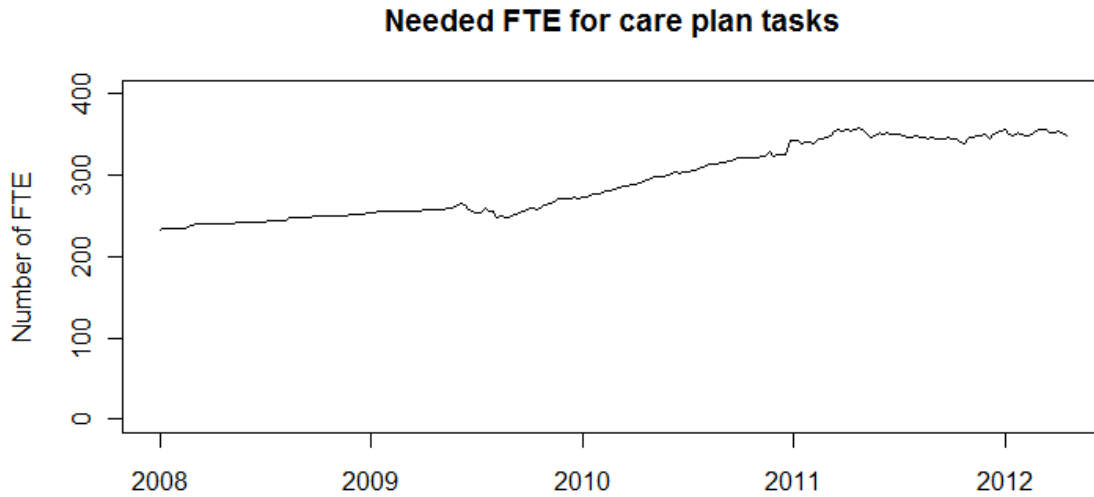


FIGURE 15: NEEDED FTE FOR CARE PLAN TASKS

**(10) Identify the patterns in obtained and not-obtained treatment goals**

The dataset contains information about 20725 treatment goals, that have been set up together with the client. As already stated in the data understanding sector, this care institution works with the 4 domain method in which the treatment goals are related to one of the 4 domains. Each goal is also linked to the client, and for every goal a description, activities, a date of evaluation and a finished date is registered. Both the description and activities of a treatment goal is registered in one field in the dataset. If, for one goal, multiple activities should be performed, these multiple activities are registered in one field. That means that the description and the activities are gathered by free input text fields, that resulted in 7229 different descriptions and 10740 different activities. The table contains also 36 entries without a description, and 1750 entries without any activities. That means that, on average, only 2,86 treatment goals have the same description. For the activities, only 1,77 treatment goals have the same activities. Lack of standardization makes it impossible to identify the feasibility of the treatment goals. Patterns in the obtained and not-obtained treatment goals could therefore, unfortunately, not be identified.

**(11) Identify the patterns in the average duration of stay**

A current trend in long-term care is that the duration of stay is decreasing over the years. Before an analysis is created based on the dataset, it is important to realize that the duration of stay is measured at a given moment in time. For this analysis the duration of stay is measured every first day of the month, from January 2008 onwards. That means that the duration of stay is measured from the start of the stay until the first of each month. In figure 16 the average duration of stay from January 2008 is elaborated.



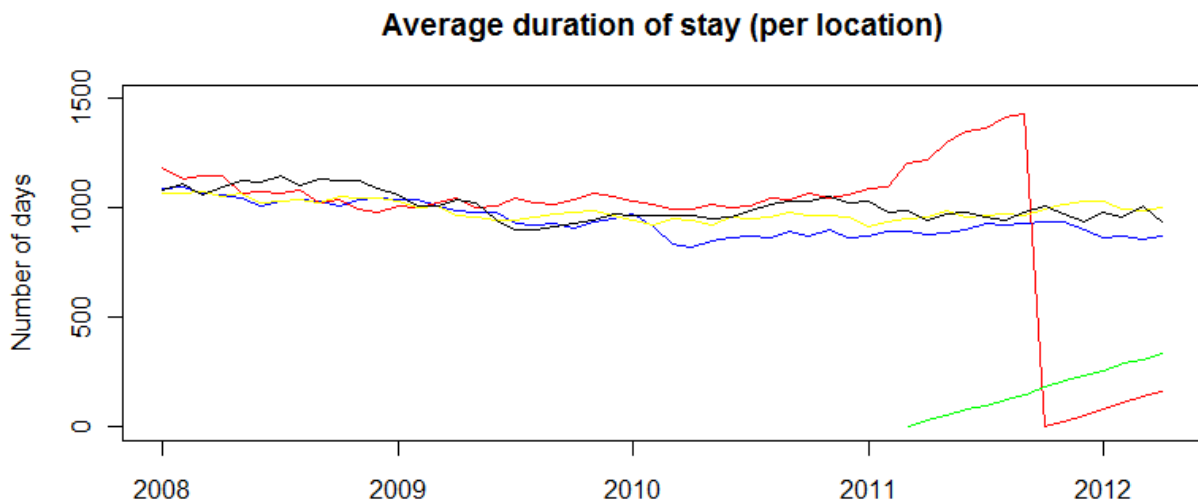


FIGURE 16: AVERAGE DURATION OF STAY (PER LOCATION)

In 2011, the red location has introduced a new room layout. This new room layout influence the average duration of stay, as one can see in figure 16. One remark could be that a move of a client from one room to another is no reason to reset the duration of stay. For this analysis, it is a reason to reset the duration of stay. Because of that, the dip is visible for the red location. The green location is a new location, which causes the constant increase in the average duration of stay. Another interesting insight is the slightly decreasing lines for the black, yellow and blue locations. Because this is the average duration of stay, it means that intramural clients stay shorter in a care institution.

It is impossible to subtract the average duration of extramural clients, but it is possible to identify the average duration for all clients(that means for intramural and extramural clients). As figure 17 shows, the average duration for all clients is quite stable. Only the green location shows a strong increase in the average duration of stay, which is due to the fact that that is a new location. Because this data is based on another table, and thus data, the red location does not show a dip in the average duration.

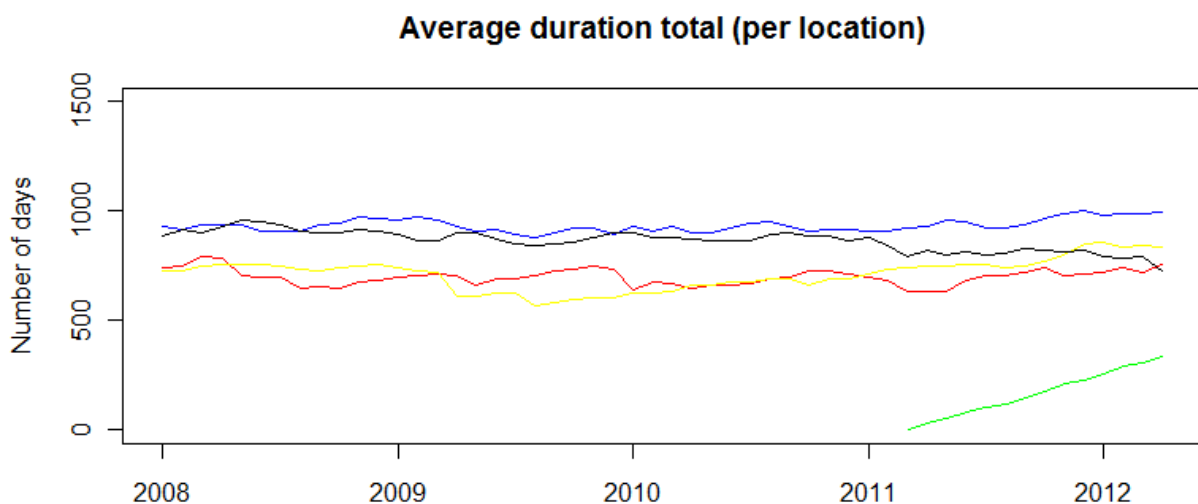


FIGURE 17: AVERAGE DURATION A CLIENT IS CUSTOMER

Because the average duration of stay is at a higher level than the overall average duration, it is clear that the intramural clients are longer customer of the care institution than the extramural clients. The extramural clients pull the average duration down.

### (12) Predict the average duration of stay

With the analysis created for the previous data mining goal, it is possible to create a forecast based on that historical data. Also for this prediction the automatic forecast() algorithm is chosen. This algorithm applies multiple different forecast algorithms to the data and selects the one with the lowest Akeike's Information Criteria (AIC) score. This score calculates the in-sample fit, low AIC scores mean a better predicted model (de Gooijer & Hyndman, 2006).

Because all the locations show a very stable picture with regard to the average duration, one forecast has been produced for both the average duration of stay of all locations and for the average duration total of all locations. In figure 18 the forecast for the average duration of stay is elaborated, whereas in figure 19 the average duration total is elaborated.

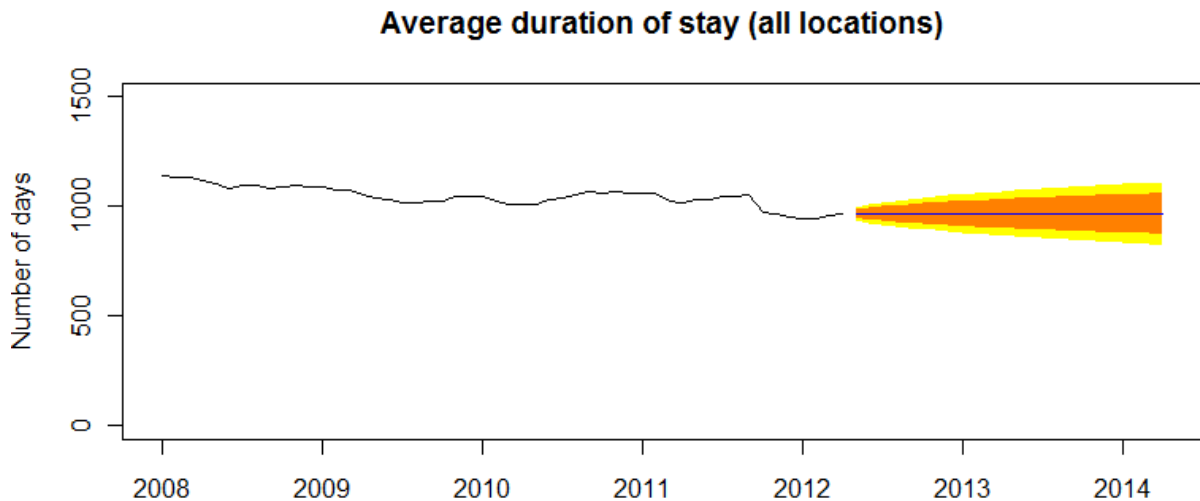


FIGURE 18: AVERAGE DURATION OF STAY FORECAST

For this first forecast, the point forecast lies at 968 days over the whole forecasted period. In April 2014 the algorithm forecasts with a confidence level of 95% that the average duration of stay will lie between 825 days and 1111 days. On the 80% confidence level the algorithm predicts that the average duration of stay will lie between 875 days and 1061 days.

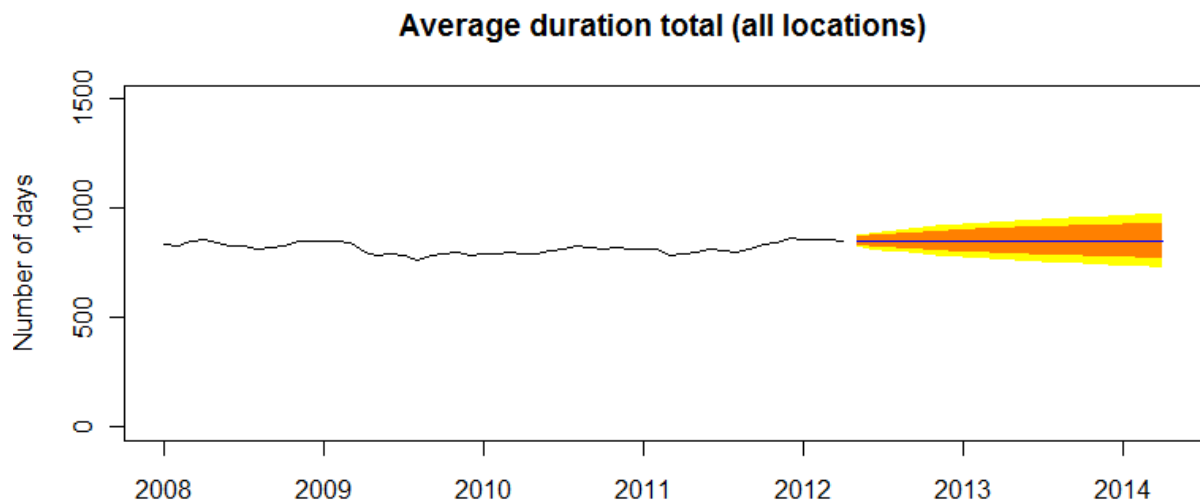


FIGURE 19: AVERAGE DURATION TOTAL FORECAST

Because the average duration for all clients is more stable, the forecast has a smaller error margin. The point forecast for the whole forecasted period is 853 days. With a confidence level of 95% the algorithms forecasts that the average

duration of all clients will lie between 729 days and 977 days in April 2014. On the 80% confidence level the algorithms predicts that the average duration for all clients will lie between 772 days and 934 days.

**(13) Identify the patterns in coercive measures / restraints**

As already elaborated in the data understanding sector, the dataset contains 25 coercive measures / restraints. Because the reason for applying the coercive measure / restraint it is hard to create valuable analysis. All the data will be used for further analysis, this means a time scope of June 2010 until April 2012. For only one measure / restraint there has been set up a goal. In table 14 the measures / restraints are elaborated per type, more complex analysis is impossible due to the lack of standardization and the few number of measures and restraints.

TABLE 14: COERCIVE MEASURES / RESTRAINTS

Type of measure / restraint	Amount
Separation	2
Fixation	15
Serve medication	8
Serve fluid and nutritionals	0

**(14) Identify the age distribution of clients**

In the figure below the age distribution of the current clients, with status “in care”, “new client”, “temporarily out of care” and “waiting list” and without a date of death, in the dataset is visualized. The data is not selected based on the location of stay or the type of care requested (intramural or extramural).

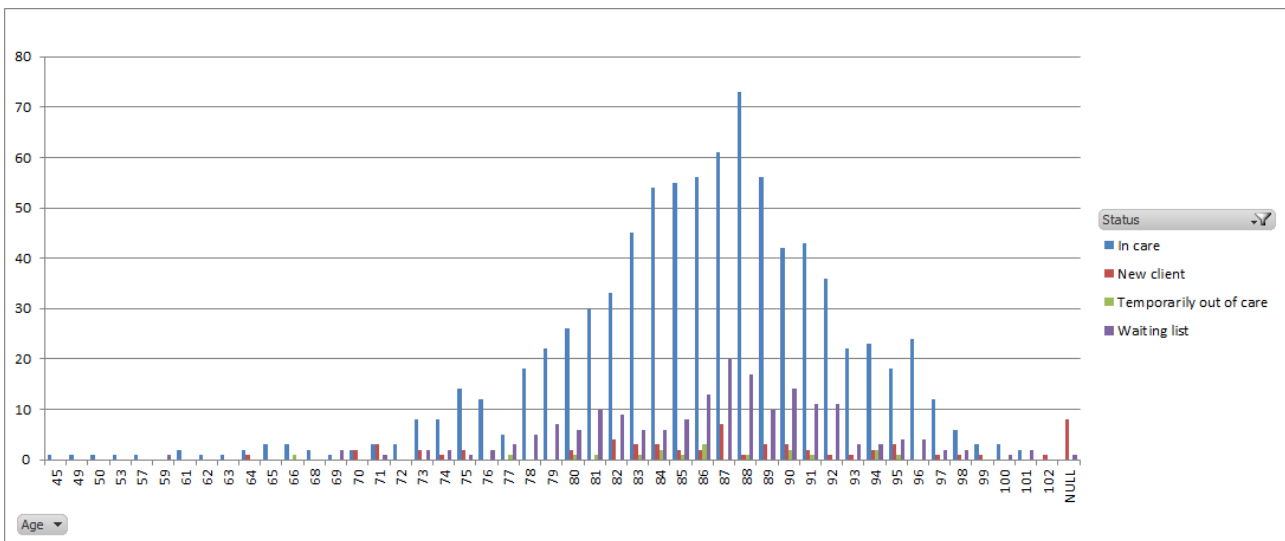


FIGURE 20: AGE DISTRIBUTION PER STATUS

More elaborate information about the information, visualized in the figure above, can be found in table below. This information consist of the minimal, mean and maximal values, but also the 1<sup>st</sup>, 2<sup>nd</sup> (median) and 3<sup>rd</sup> quartile. Also the number of not available values (NA's) could be found in the table below.

TABLE 15: AGE DISTRIBUTION PER STATUS

Type	Min	1 <sup>st</sup> quartile	Median	Mean	3 <sup>rd</sup> quartile	Max	NA's
New client	64,00	82,00	86,50	84,96	90,00	102,00	8,00
Waiting list	59,00	82,00	87,00	86,21	90,00	101,00	1,00

In care	45,00	83,00	87,00	85,84	90,00	101,00	-
Temporarily out of care	66,00	83,25	86,00	85,56	90,00	95,00	-
	<b>45,00</b>	<b>82,00</b>	<b>87,00</b>	<b>85,85</b>	<b>90,00</b>	<b>102,00</b>	<b>9,00</b>

From the table above it is clear that 50% of the clients are aged between 82 and 90. This is higher than the average age at death (for females: 79,6 years, for males 74 year) as stated by the CBS (Central Bureau of Statistics, 2012). Another interesting insight is that the age distribution is almost the same for clients with the status “in care” and clients with the status “waiting list”. This means that the age distribution will not change that much over time, and certainly will not change towards lower ages.

### (15) Identify the patterns in ADL information

One table in the dataset contains information about the Activities of Daily Living. For each client a broad range of aspects are registered, including an explanation, approach, nutrition, diet, allergies, syndrome, medication use, medication management, support tools, mobility, toileting, incontinence material, day, night, security, activities, care agreements and washing information. All this information describes what the client is able to do, and what the client is not able to do anymore. The table counts a total of 1504 rows with data about 1441 clients. That means that only for a small number of clients the ADL information is filled multiple times. All the columns are text fields that contain heterogeneous data. In the following table the number of different values per column are presented, together with the number of NULL values. In the last column the percentage of different values is presented, where a small percentage is better. Only the medication management has a percentage under 10%, which is the least valuable value in the table.

TABLE 16: NUMBER OF DIFFERENT VALUES

	Number of different values	Number of NULL values	Percentage
Explanation	1480	15	99,40%
Approach	1150	224	89,84%
Nutrition	898	175	67,57%
Diet	234	367	20,58%
Allergies	148	467	14,27%
Syndrome	1266	190	96,35%
Medication use	655	276	53,34%
Medication management	5	169	0,37%
Support tools	564	528	57,79%
Mobility	738	222	57,57%
Toileting	502	266	40,55%
Incontinence material	163	441	15,33%
Day	233	771	31,79%
Night	254	782	35,18%
Security	345	671	41,42%
Activities	489	852	75,00%
Care agreements	605	671	72,63%
Washing information	73	1419	85,88%

With the data that is available in the current dataset it is impossible to create analysis that could create a good insight in the patterns, the data is too fragmented due to the lack of standardization. Only for a small number of clients the ADL information is filled multiple time, so that it is impossible to gain insight in the decline or progress of clients. It would be advisable to create for all these (see table 16) aspects a point scale. Based on that point scale it is possible to gain insight in the decline or progress of all clients.

### (16) Identify & predict the ZZZ-mix

Figure 21 shows the current ZZZ-mix per location. In order to create this current ZZZ-mix, only the current delivered ZZZ's are selected. Every ZZZ has a start date and an end date, which makes it possible to select only the current ZZZ for each client. The mix is therefore based on the ZZZ's that are delivered by the care institution, and not the indicated ZZZ. It is possible that a client receives an ZZZ5 indication from the care office and receives ZZZ4 at the care institution for different number of reasons. It is therefore important to know that this ZZZ-mix is based on the delivered ZZZ's.

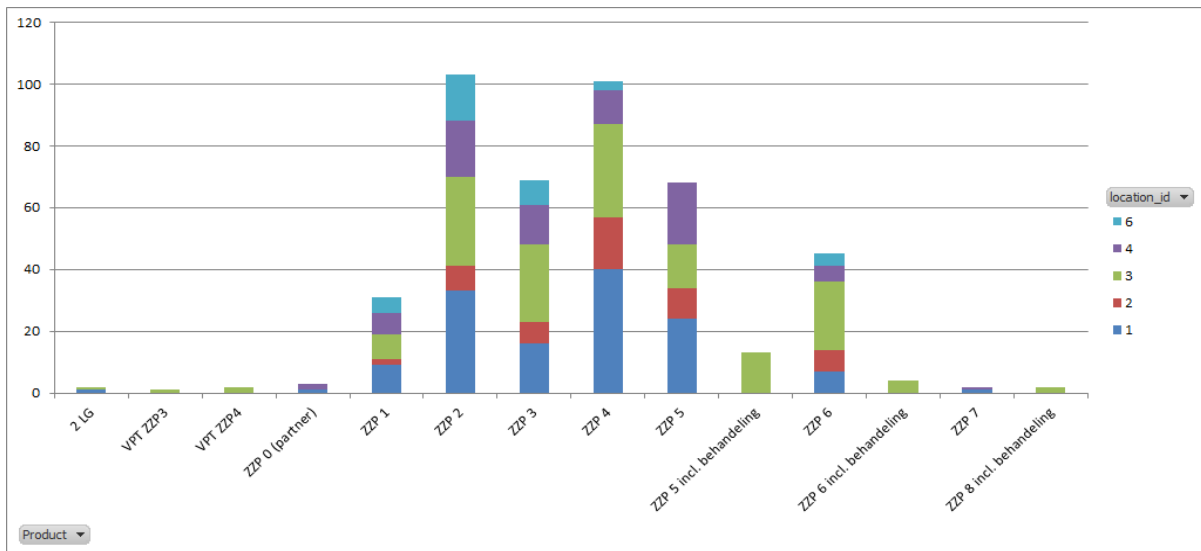


FIGURE 21: CURRENT ZZZ-MIX (DELIVERED)

In order to predict the future ZZZ-mix it is necessary to have historical data which represent the ZZZ-mix over a longer period of time. The dataset contains information about the delivered ZZZ's from January 2009. Before January 2009, care institutions did not work with ZZZ's. Changes in laws and regulations make it therefore hard to create a solid basis for data mining algorithms to create predictions. Nevertheless, the data that is present in the dataset will be used to identify the historical ZZZ-mix, and based on that information a prediction will be made about the future ZZZ-mix.

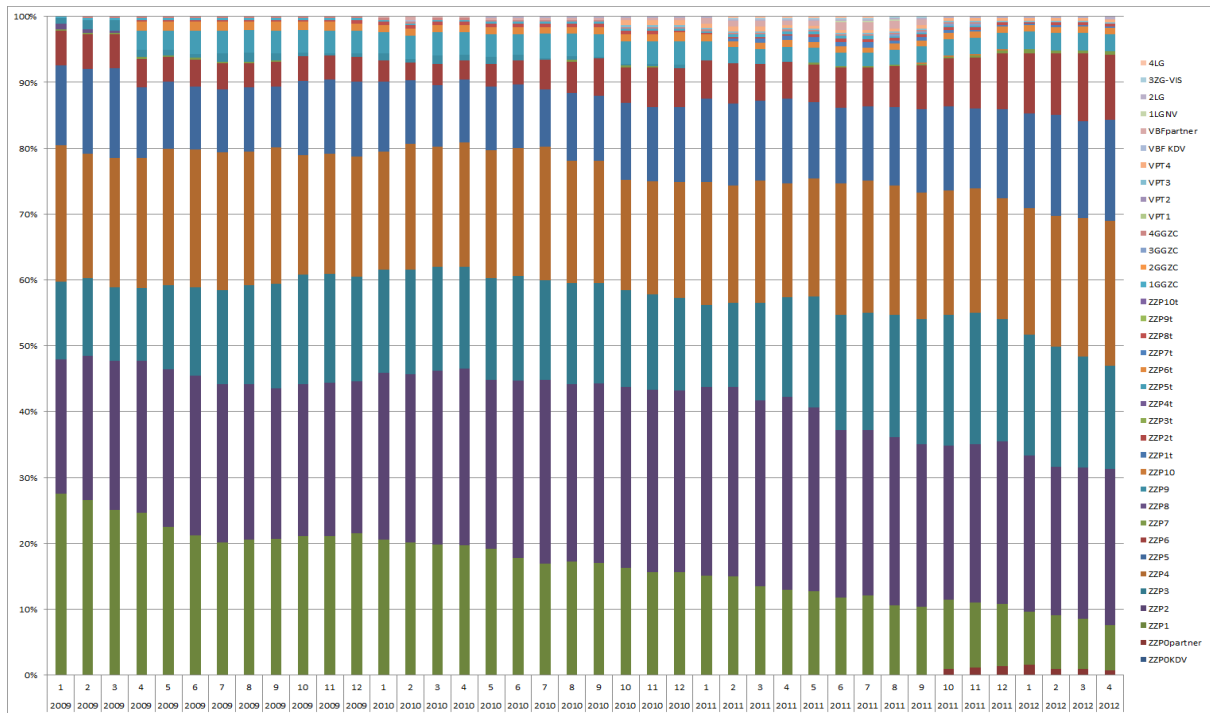


FIGURE 22: ZZZ-MIX OVER TIME (DELIVERED)

The dataset contains information about 36 different delivered ZZZ-types. A query is used to select the right data from the dataset. For every first day of every month, from January 2009 until April 2012, the number of distinct clients is counted that receive a particular ZZZ-type. In total 26 different ZZZ-type have been delivered to clients. Based on this information it is possible to create insight in the course over the last few years.

Figure 22 makes clear that the number of ZZZ 1, 2 and 3 clients declined from almost 60% to almost 45%. This is a decline of almost 15%, which means that the care institution needs to focus more on the higher ZZZ-type clients. The outcome is in line with the current trend that elderly stay longer at their own house and receive extramural care. This means that when the elderly turn to a care institution, they will need more care and will receive a higher ZZZ-type. Over 90% of the clients possess ZZZ1, ZZZ2, ZZZ3, ZZZ4, ZZZ5 or ZZZ6. Current plans of the Dutch government include, inter alia, to minimize the number of intramural clients with ZZZ1, ZZZ2 or ZZZ3. These plans could have a major impact for this care institution, because almost 50% of the clients possess such ZZZ-type.

With the current plans of the Dutch government in mind, it is important to have insight in the forecast for ZZZ1, ZZZ2 and ZZZ3 types. In order to create the forecasts, the standard forecast() method from the R Studio program is used. This method applies different algorithms to the times series and selects the best fitted model.

### Forecasts from ETS(A,A,N)

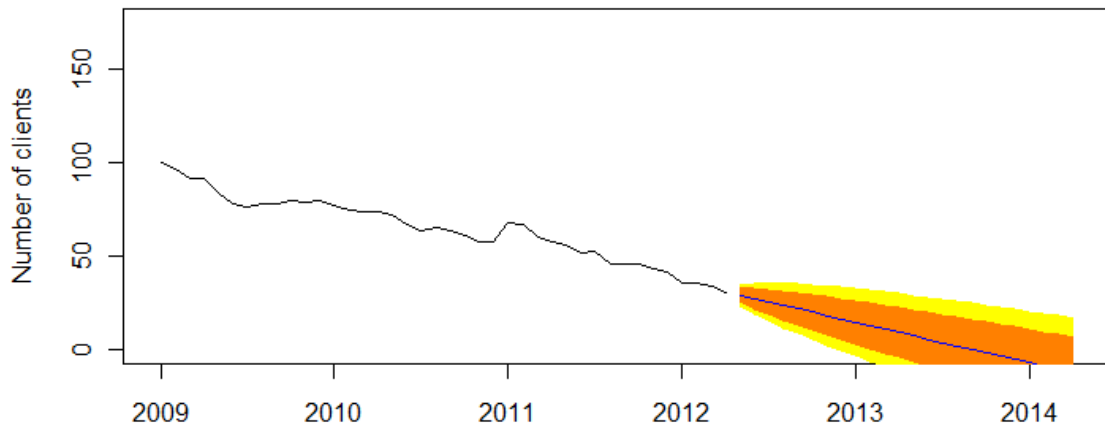


FIGURE 23: NUMBER OF ZZP1 CLIENTS (DELIVERED)

The forecast method selected a model with an additive trend component and no seasonality. This model is also named as the Holt’s linear method. The point forecast of this model predicts that the number of ZZP1 clients will be zero in September 2013, based on the historical trend. In December 2012 the number of ZZP1 clients could be zero based on the lower 95% confidence level. For the lower 80% confidence level the model predict zero ZZP1 clients at first in February 2013.

### Forecasts from ETS(M,N,N)

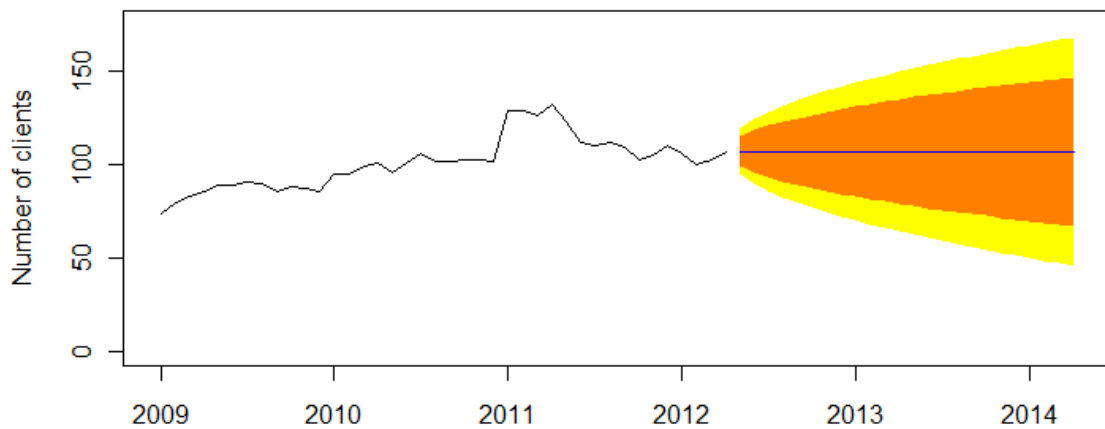


FIGURE 24: NUMBER OF ZZP2 CLIENTS (DELIVERED)

For this time series the forecast() method selected a model without trends or seasonality, which means that the prediction is stable. The point forecast of this model is stable at 107 clients, whereas the window at the 95% confidence level is between 46 and 168 clients.

### Forecasts from ETS(A,Ad,N)

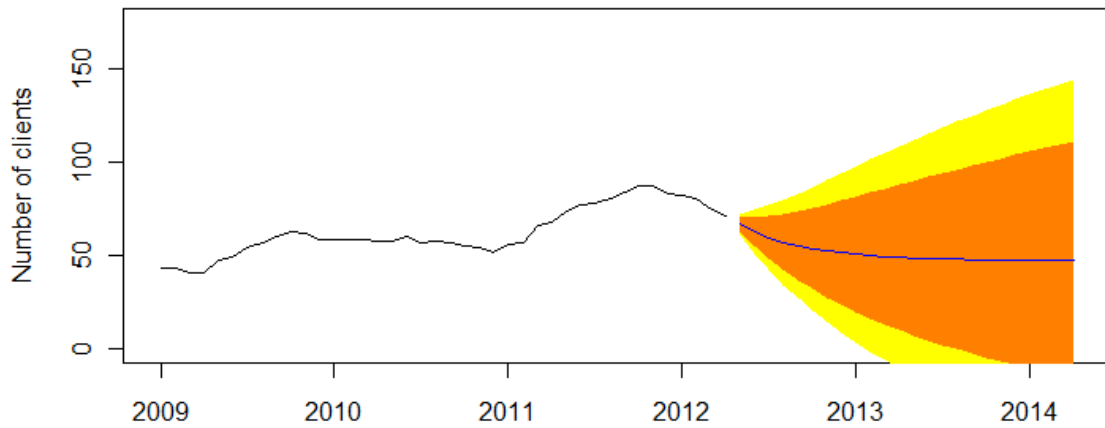


FIGURE 25: NUMBER OF ZZZ3 CLIENTS (DELIVERED)

For ZZZ3 clients the prediction is very uncertain. The prediction levels are very wide, which means that the window is very broad. In April 2014 the number of ZZZ3 clients could be between 0 and 145 clients, at a confidence level of 95%. This means that the number of ZZZ3 clients could be tripled or disappear. The forecast method selected a model with an additive damped trend component and no seasonality.

Another interesting forecast is that of ZZZ6 clients. From figure 22 it is clear that the number of clients with a ZZZ6 indication increases. The prediction for this ZZZ-type can be found in the following figure.

### Forecasts from ETS(A,A,N)

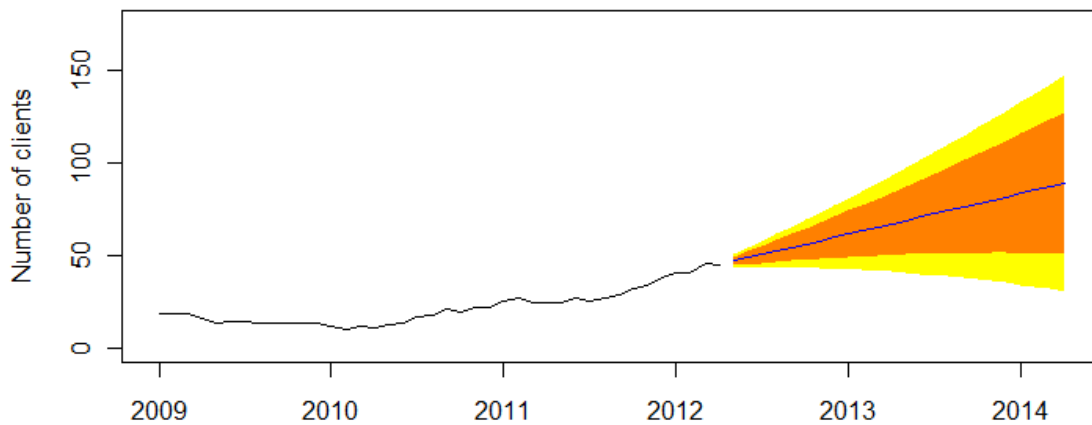


FIGURE 26: NUMBER OF ZZZ6 CLIENTS (DELIVERED)

For this time series the forecast method selected a model with a trend component and no seasonality. The number of ZZZ6 will increase, based on the point forecast. The 95% confidence window for April 2014 is between 31 and 147 clients. The 80% confidence window for April 2014 is between 51 and 127 clients. At this moment the number of ZZZ6 clients lies at 45. This means that one could state that it is for 80% sure that the number of ZZZ6 clients will increase.

#### (17) Identify the relationships between operations and ZZZ

All the operations performed for the clients are registered in the care plan of the clients. A total number of 23100 rows contain information about the operations in the care plan of the clients. 215 different tasks have been identified and used



for 2110 clients. Of these 2110 clients, only 450 clients possess a ZZP. As a result, the data is too fragmented for proper analysis, this time not primarily due to the lack of standardization. 215 different tasks is not that much, taken into account the numerous actions care personnel could perform for the clients. The operations are not directly linked to the ZZP of the client. A ZZP is just a product that is linked to the client, in this dataset. This means that the clients could possess multiple products, what makes it impossible to link the tasks from the care plan to the ZZP's of the client. Not only the lack of standardization is here a problem, also the inability to link the tasks / operations to the ZZP of the client.

#### **(18) Identify the patterns in production**

Production information is very complex information with a lot of dependencies. In order to create useful analysis it is important to gain understanding of all those dependencies. The dataset contains, as already stated before, more than 1.8 million records with information about the delivered production. Production information is registered by an automatic process, products delivered to the client contain product patterns that prescribed how often and when the product should be delivered. The automatic process registers the production based on those product patterns. But, this is raw data and could not directly be used for analysis. The software adds all kind of rules to create insight in the data, for example checking whether the client has the right indication, whether the production is within the indication, whether the production could be delivered at the conscious location, etc.

Due to the automatic process of registering production (standardization), the homogeneity of the data is good. But, because of all the implications and dependencies (laws and regulations) it is, for this study, too complex to create proper analysis.

#### **(19) Identify the patterns in financed and non-financed care**

For the registered production, no distinction has been made between financed and non-financed care. All care institutions declare their production at the care offices every month, for the previous month. The care offices validate this declared production and grant a certain amount of money to the care institution. By sending a return message to the care institution, care offices express the amount of money that is granted. This information is not registered in the dataset, which makes it impossible to gain insight in the patterns between financed and non-financed care.

#### **(20) Identify the patterns in unplanned care & (21) Predict the unplanned care**

The distinction between unplanned care and planned care is not registered in the dataset. That means that the distinction should be made afterwards. For every production entry in the dataset the validity should be defined. Planned production is production for which a client has a product pattern. Every production entry should therefore be matched with the product patterns of the client. Because clients could possess multiple product patterns at the same time and over time, it is too complex to validate every production entry. This data mining goal could therefore not be accomplished.

#### **(22) Predict the occupancy rate**

Data mining goal 4 created insight in the number of clients over time. This information could also be used to predict the future occupancy rate. The occupancy rate is the number of clients divided by the number of beds available. Because only intramural clients reside in the care institution, only the intramural clients are selected. The data that is used for Figure 10 is also used for this data mining goal. Each location had a different number of beds, this is already defined in the Data understanding chapter.

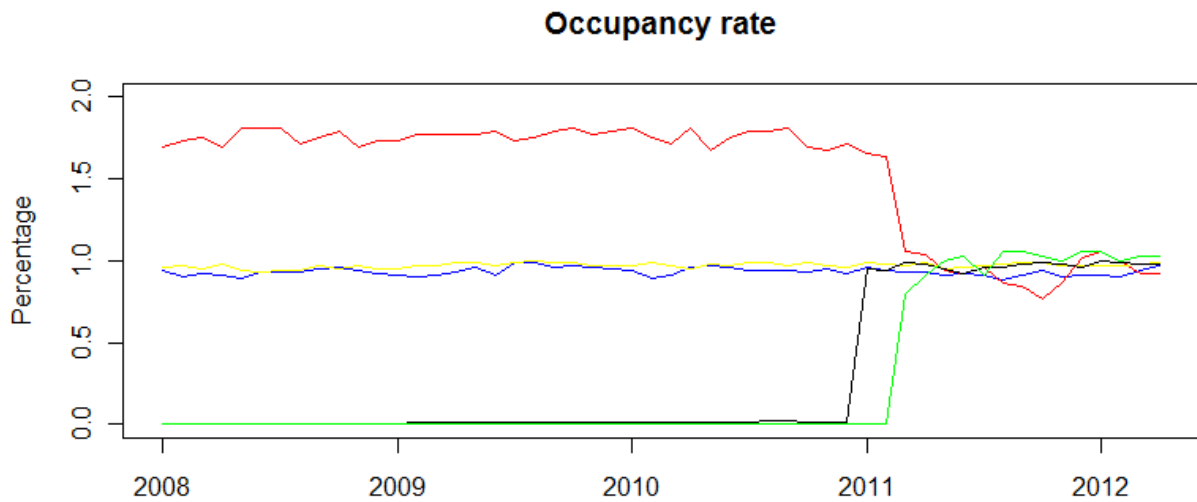


FIGURE 27: OCCUPANCY RATE PER LOCATION

The red location shows an occupancy rate of more than 100%, which is caused by the split in 2011. Before 2011, the red location contained more beds. All locations show a high occupancy rate which is higher than 90% in every month. For example, the blue location had an occupancy rate of 0,9718 in April 2012. Because the blue location contains 142 beds it means that 138 beds were occupied, and only 4 beds were available. Minimizing the number of available beds is a goal in order to increase the profit.

Only for one location a prediction will be presented, because all the occupancy rates show a very stable line. The yellow location shows the most stable line, which results in a very narrow forecast window. In figure 28 the forecast of this location is presented, based on the forecast() algorithm as it is used for previous data mining goals. This forecast predicts a point forecast of 0.97926 for the who forecasted period, without seasonality and trends.

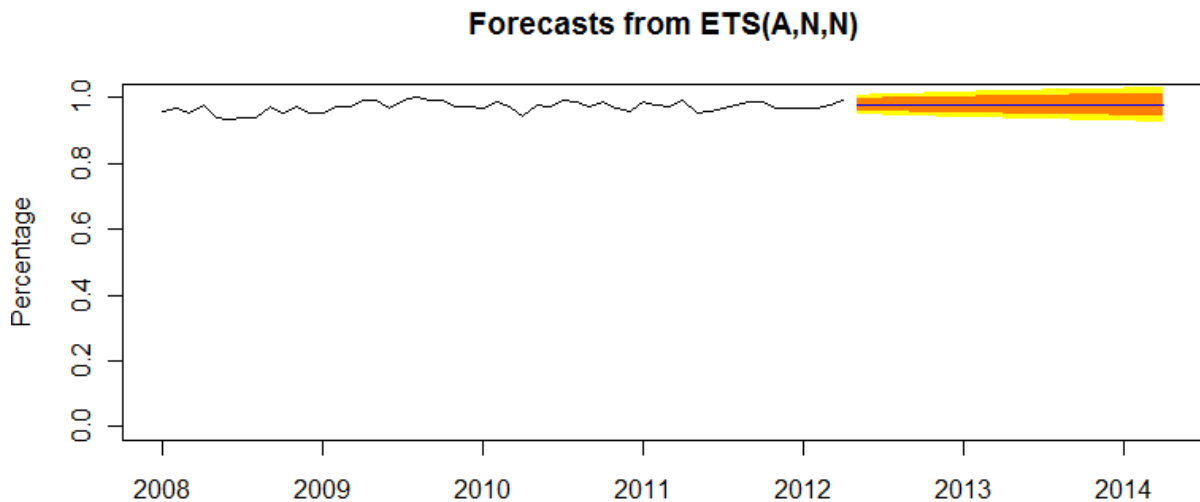


FIGURE 28: OCCUPANCY RATE FORECAST (LOCATION 3)

**(23) Identify the patterns in the waiting list information**

One table in the dataset contains information about the waiting list. A client can be on the waiting list for a location, department and for a product, where all combinations are possible too. Clients that are more dependent on care could receive a higher priority than others. Also the start and end date of the waiting list entry are registered, together with the end reason. The table contains 1315 rows with waiting list information, where every row is one entry. Because the care

institution uses the ECR software since 2008, the data from that date onwards is selected, resulting in 1303 rows. One client could be on the waiting list with multiple entries. The 1303 entries resemble 530 clients, 667 entries have been closed, whereas 636 entries are still open. These 636 open entries resemble 294 clients. That means that at this moment, read April 17, 2012, 294 clients are waiting for admission.

The closed 667 entries resemble 280 different clients. The average waiting duration for these closed entries was 230 days. That does not mean that on average a client is allowed after 230 days, an entry could also be closed because the client went to another care institution, the client deceased, etc. For every waiting list entry that is closed, an end reason could be assigned. Because it is not mandatory to register the end reason, analysis based on that information could create a bias because for 117 entries there was no end reason registered. That means that for 17,5% of the closed entries there was no end reason registered. With the data from the dataset it is therefore not possible to gain insight in the reason for ending the waiting list entries.

In order to gain insight in the progress of number of clients on the waiting list, the number of distinct clients on the first day of every month (since January 1, 2008) is selected per location. This means that the number of clients is counted at the 1<sup>st</sup> day of every month with a valid waiting list entry. This information resembles a time series, which is elaborated in Figure 29.

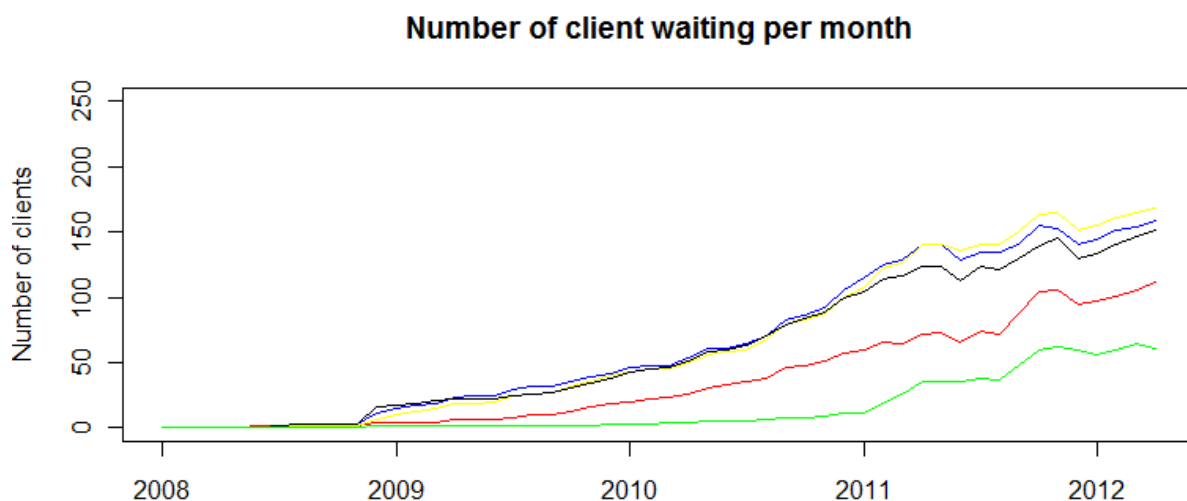


FIGURE 29: WAITINGLIST PER MONTH PER LOCATION

With the information that is resembled in figure 29 it is clear that the number of clients that reside on the waiting list is constantly increasing for each location. Waiting list information for the green and black locations (new locations) has been migrated from the previous software used by those locations. The cause of this increase cannot directly be related to the increasing number of clients that want to reside in this care institution. The increase could also be caused by the implementation of the ECR software / waiting list functionality.

Another interesting analysis could be based on the products for which the clients reside on the waiting list. All the waiting list entries are waiting for only 9 different products. 52 entries contain no product. For the next analysis the same information is selected as the previous analysis, but the number of clients with a valid registration are counted per product. Figure 30 creates insight in the fact that ZZP2 and ZZP4 resemble the largest increase over time.

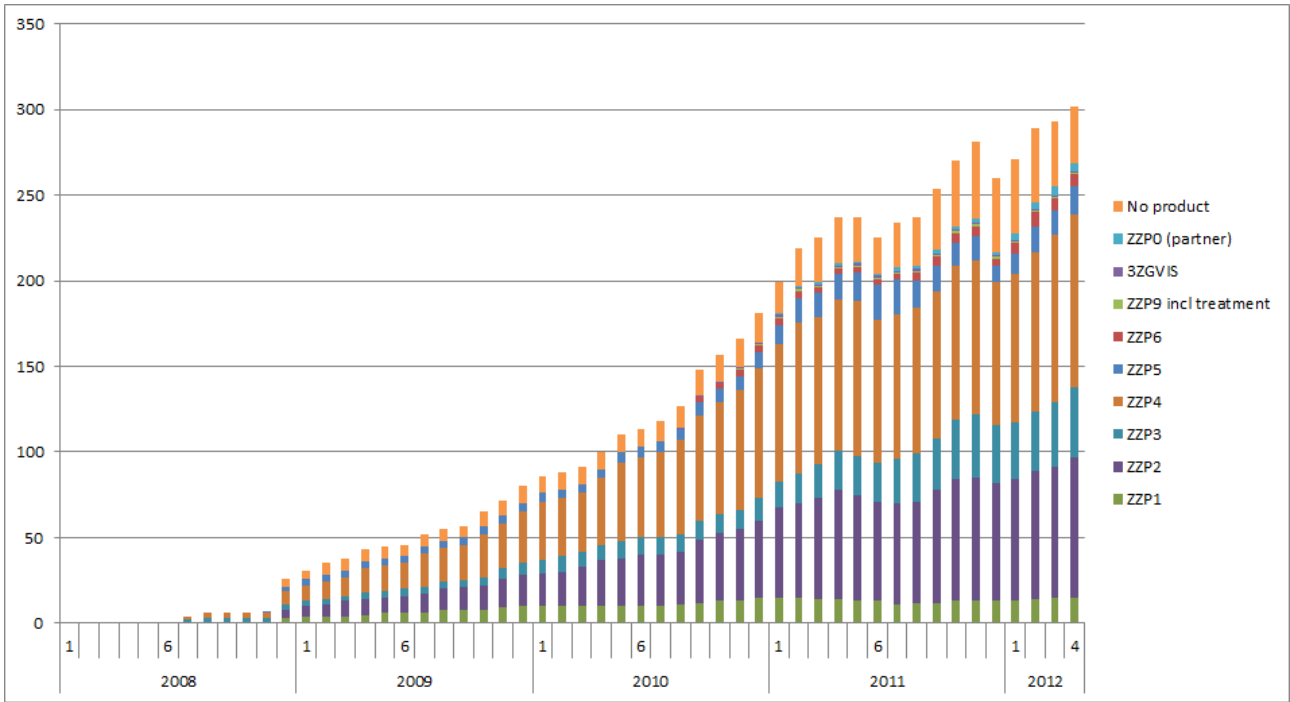


FIGURE 30: WAITING LIST PER PRODUCT

**(24) Identify the average duration of a ZZP**

Based on the historical data it is possible to identify the average duration of a ZZP. In the past 1384 ZZP's were registered and closed. The current ZZP's are not taken into account for this analysis, because the duration can only be determined when the ZZP is closed. A ZZP closes when a client request another ZZP, moves to another institution or dies. This last reason for ending a ZZP is also interesting for further analysis.

In order to identify the average duration of the ZZP's, all the registered ZZP's are selected. With this information the average duration of the closed ZZP's is selected from the dataset together with the number of clients on which the average is based. Because an average needs more than one value, all the average durations based on less than 2 clients are removed. The result of this selection is elaborated in the following figure.

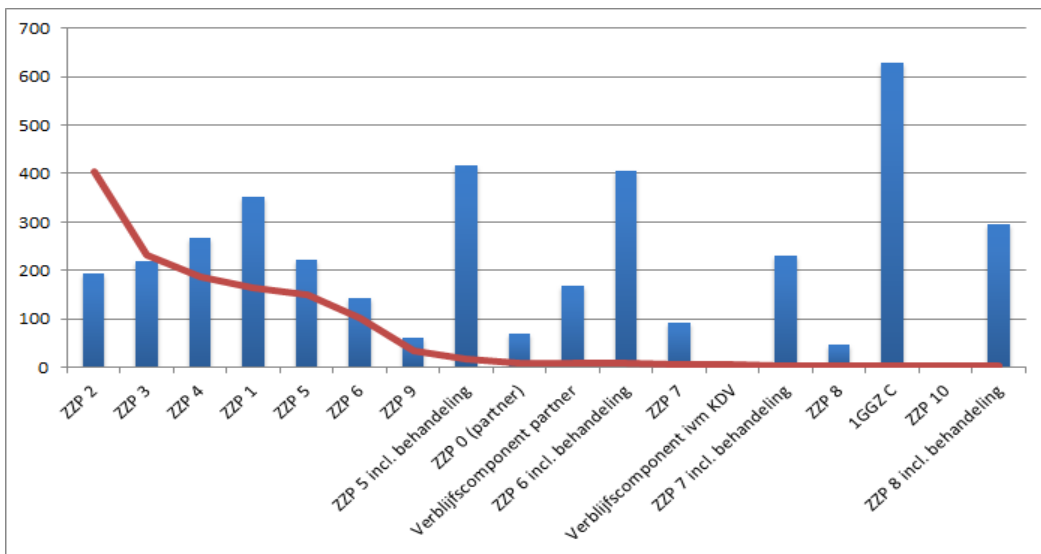


FIGURE 31: AVERAGE DURATION OF ZZPS

In figure 31 the blue bars represent the average duration of the ZZPs in number of days, whereas the red line represents the number of clients on which the average is based. The higher the number of clients on which the average is based, the more reliable is the average duration. The average duration of ZZP 0 (partner) is based on 9 clients, which means that all the ZZPs on the right of that ZZP are based on less than 10 clients.

**(25) Identify the patterns in mutations**

All mutations should be registered in order to preserve the quality of the data in ECR software. AZR messages are used to communicate between care offices and care institutions. Care offices should also possess up-to-date information about a care institution in order to validate declarations. Since August 2011, all AZR messages are saved in the dataset. But this information could not immediately be used for analysis. It is too complex (for this study), to prepare this data for data analysis. Also the usefulness of this data for proper data analysis could not be identified, due to the complexity of this data analysis with regard to the way the data is registered.

## 4.4. EVALUATION & VALIDATION

### 4.4.1. EVALUATE RESULTS

Together with the care institution that provided the data for this research, the results are evaluated. Because of the explorative character of this study, the results are superficial. In order to gain a deep understanding of the internal business, this project should be customized for a single care institution. That enables to create subjective data mining goals for that one care institution.

First of all, the care institution was pleased with the results of this study because it gave a very interesting view on the internal business. Patterns in the number of incidents gave the care institution insight in the progress of incidents over the week and during the day. The incident models gave also insight in the type of incidents per locations, which could be the reason for the care institution to implement changes. For example, the bathroom sizes are different over the locations, which could be the main reason for the larger percentage of incidents that occur in the bathroom in some locations. Insight in the progress of incidents during the day gave insight in the peaks between 8 and 9 am and between 5 and 6 pm. That information is of added value for the care institution in order to increase the quality of care, because the care institution could respond to these facts. Also the ZZP-mix information created a good insight in the progress of the mix over time. This information helps the care institution to create strategic plans and decisions.

Other performed analyses are less interesting. This is especially the case for the forecasts done in the modelling phase of this study. ZZP-mix, number of clients, occupancy rate could not be forecasted based on the historical data. It is too dependent on other factors that it is not funded enough to be useful. Also the analysis done for the care related time is not useful, because it is not based on the actual time.

The reason why the analysis that could not be performed are confirmed by the care institution. Lack of standardization and data are the two main reasons for not being able to create the proper analysis. In order to be able to create those analysis in the future, ECR software vendors should focus more on the easiness to register this data. Coercive measures and restraints are in this case an exception. Complexity is the reason that prevented from analysing the production information. Business rules, laws and regulations are programmed in the ECR software. The data does not contain that logic, which makes it difficult to analyse the data.

For all data mining goals applies that more elaborate analysis is possible when data is gathered more quantitative. At this moment most data is gathered qualitative, due to the personal process of delivering care. By adding a quantitative option, it is possible to gain more insight in the data. For example, implementing a point scale for ADL creates more insight in the decline or progress of a client.

### 4.4.2. REVIEW PROCESS

Some of the data mining goals could not be achieved due to the complexity of the data, lack of data and lack of standardization. Production information that is gathered by the care institution was too complex for immediate use of data mining algorithms. Business rules, laws and regulations should first be applied to the data, before proper analysis is possible. In order to create analysis based on the production information, the data preparation phase of the used CRISP-DM method is very important. Business rules, laws and regulations should be known before creating the analysis.

For data mining goals that could not be achieved due to the lack of standardization, it is important to create standards. A standard set of care goals could for example create more insight in the feasibility of the goals. Also standards for ADL

registration could create more insight in the decline or progress of a client. Incident information is to a certain level standardized, but further implementation of standards could gain even more insight in the patterns that exist. ECR software vendors should give care institutions the possibility to collect data on a high level standardization, without requiring the use of this high level standardization. Because not all care institutions are in favour of the high level standardization application.

Lack of data is another important reason for not being able to create analysis. Information about customer experience and complaints are not gathered by the care institution, used for the data analysis in this study, in ECR software. This makes it impossible to create analysis. ECR software vendors should research the need of care institutions to create such functionality in the software. Customer experience is now measured by external companies, with standardized questionnaires and answers. This results in quantitative data that could be used for data analysis.

Much information that is gathered by care institution is gathered in a qualitative way. Free text input is used to facilitate this. The main reason for this is the core process of care institutions, delivering care. This is a personal process with qualitative data as a result. No two clients are the same, which is also the case for the data that should be gathered for the clients. Coercive measures and restraints are rarely performed, so there is little data available. This is also the case for complaints, only 1 or 2 complaints per year could not be a basis for data analysis.

#### 4.4.3. VALIDATION

This research is validated using 4 different validation aspects, which are: construct validity, internal validity, external validity and reliability.

##### **Construct validity**

The research conducted is explorative of intent, because of the limited available research in this sector. In order to create an understanding of the information needs within the Dutch long-term care sector, multiple interviews have been performed with multiple stakeholders. This resulted in a list of information needs and a valuation score. The valuation score determines the importance of the information needs.

For the data analysis part of this research, operational data of one long-term care institution is used. The data from the dataset is collected from a single source, which is the ECR software. All the data is operational data and used to manage and finance the core processes of the care institution. Data analysis is based on the data mining goals that have been determined based on the information needs collected during the interviews.

Construct validity is met by using multiple interviews with multiple stakeholders. And the data used for data analysis is operational data used for managing and financing the core process of the care institution.

##### **Internal validity**

This research does not directly study causal relationships. Assessing the internal validity should be done when causal relationships have been studied (Yin, 2009). In the data analysis part of this study some relations have been researched. All the results of the data analysis part have been verified with experts, from the care institution that has made the data available. The purpose of the data analysis was not to discover and establish causal relations, the purpose was to gain insight in the possibilities of the gathered data.

**External validity**

This research has a clear focus domain, which is the Dutch long-term care sector. In order to create a clear understanding of this sector multiple interviews have been performed. Not only care institutions have been interviewed, also, for example, the Ministry of Health, Welfare and Sport is interviewed for this research. This means that the results of the interviews performed in this study are generalizable to the whole Dutch long-term care sector.

The data analysis part of this study is less generalizable. For data analysis, data from one care institution is used. Different care institutions use different software to gather data. This means that the results of the data analysis part of this study are generalizable to all the care institutions using the same software.

**Reliability**

All the steps taken in this research are described in detail in the Research approach section. Also the interviewees are elaborated and described in this research, together with a global description of the dataset and the way the dataset was obtained.

Repetition of this research will not automatically result in the same results due to time constraints. Current information needs, as addressed in this research, could differ from future information needs, due to the always changing environment. Care institutions have to justify itself to the government and their stakeholders. If this environment changes, the information needs of the care institutions will most probably also change.



## 5. CONCLUSION

All the sub research questions that have been described in the Introduction, are answered by performing interviews and data analysis. Based on the answers of the sub research questions, it is possible to answer the main research question of this study:

*“ How can knowledge discovery support Dutch long-term care institutions to manage the internal organization? ”*

As already described, knowledge discovery could support Dutch long-term care institutions by gaining insight in their internal business. This study focussed on the global information needs of care institutions. These information needs were then translated into data mining goals. All the data mining goals are therefore based on the overall information needs of care institutions, not only in the VVT subsector of long-term care but also in the GGZ and GZ subsectors of long-term care. Due to the wide scope that is used for this research, the information needs are high level. This wide scope is chosen, because knowledge discovery is a new field of study for the Dutch long-term care. There is little research on applying knowledge discovery processes and techniques in the Dutch long-term care. With this explorative research a first step has been made to increase the efficiency of the gathered data of care institutions.

All the data mining goals are placed in a matrix (Figure 32). The size of the circle represents the importance of the data mining goal, based on the times mentioned in the interviews with experts. In the top row of the matrix all the goals that are related to the quality of care goal are elaborated, whereas in the bottom row the goals that are related to the financial state goal of care institutions are elaborated. Based on the data analysis the data mining goals are presented on a time scale. The data mining goals on the left side could already be satisfied, goals on the right side should be taken into account for future developments.

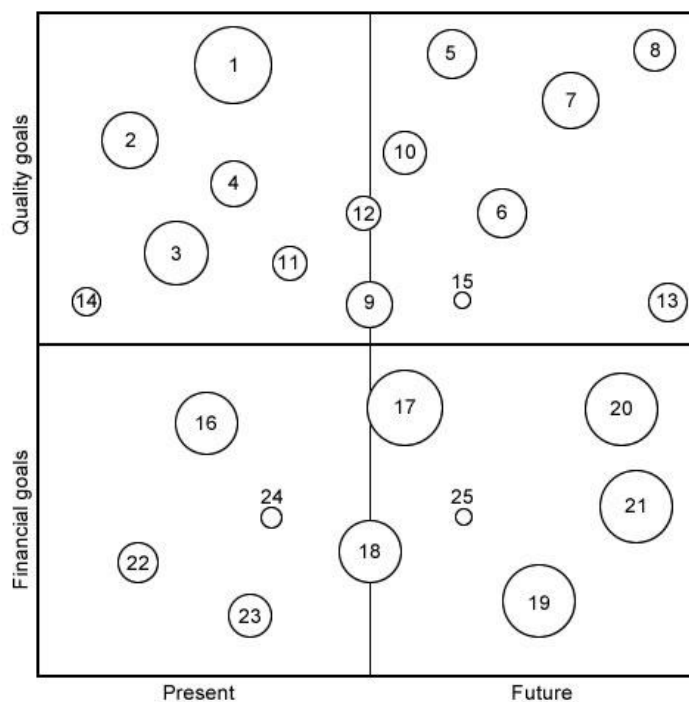


FIGURE 32: DATA MINING GOALS MATRIX

A large amount of data that is gathered by care institution is qualitative of nature. This is due to the fact that delivering care is a personal process. Standardization could cause a less personal approach of care personnel, which should be prevented. Besides the possibility to gain data in a qualitative way, it should be possible to collect data in a more quantitative way. All kinds of data collection techniques exist and could be used to increase the usability of the gathered data. Point scales could be added for collecting ADL information, where one point means that a client is very dependent on the help of care personnel and five points means that the client is independent. Multiple ADL measurements per year could create insight in the decline or progress of a client. Another example are the measurements that are gathered. Standardization of the data collection could increase the effectiveness of the data for data analysis.

Data gathered by care institution is very dependent on the laws and regulations. Long-term care is a public facility in The Netherlands, and every citizen in The Netherlands can invoke it. This makes the Dutch government responsible for the long-term care. Changes in, and new, laws and regulations has consequences for the information needs of care institutions and for the data that is gathered.

Based on the evaluation of the data analysis results, it is clear that predictive models are not that valuable for care institutions. Predictions are too dependent on environment variables. The number of clients is for example dependent on acquisition plans, laws and regulations, population, etc. It is impossible to predict the number of clients in the future based on historical data, due to these dependencies. Descriptive modelling on the other hand is very interesting for care institutions, because of the large quantity of valuable information. Staffing could be based on the activities of the clients care plan, but also the number of incidents that occur during the day could be used to plan personnel. In the future, predictive modelling could be of increased value when data is gathered in a more quantitative way. This enables care institutions to gain insight in the progress or decline of clients, feasibility of care goals, etc.

Not only care institutions, but also the demanding stakeholders should have a stronger focus on their information needs. Long-term care, especially VVT, should focus on increasing the quality of care. Quality of care could not be measured by registering all incidents and compare all institutions with each other. One could also argue whether an incident is a fall incident or not, for example when a client slips out of a wheelchair. Registration of these events results in large quantities of data of which the value is questionable. However, registration takes time which in result costs money. Care personnel, on the other hand, are required to register all this data without a clear reason.

Overall it is clear that the large quantity of valuable information that is gathered by care institutions could be used more efficient by being more aware of the information needs. Current data is not gathered to gain insight in the internal business, current data is collected to support the core process of the care institutions. Care institutions should pay more attention to their information needs and the purpose to gather data.

## 6. LIMITATIONS

This study contains a couple of limitations that will be described in this sector. These limitations are the result of staying within the scope of this research. The goal of this research was to gain insight in the business and data of care institutions. Care institutions collect a lot of (valuable) data, which is not used as such. The overview of the long-term care sector contains the most important stakeholders for long-term care institutions. A complete overview of all the stakeholders in the long-term care sector was not the goal of this research. Stakeholders of individual care institutions are therefore not researched.

For the data analysis part of this study is chosen for a single case study. This decision could be seen as a limitation, because the collected data is one hand dependent on the use of ECR software. Not every care institution uses the same ECR software. But, this explorative study is aimed at the possibility to use knowledge discovery to gain insight in the internal business of a care institution. Another argument is that a lot of data is important for care institutions, in order to get financed, which makes the outcome of the performed analysis generalizable for other care institutions that use ECR software.

Future research should focus more on the subsectors in the long-term care sector. For this study, only one GGZ and one GZ institution are interviewed to gain insight in their information needs. This results in a strong focus on the VVT subsector. Not only are the target groups completely different, also the laws and regulations differ within each subsector. Also the data analysis is based on the data gathered by a VVT care institution.

Due to the fact that long-term care is financed by public money, laws and regulations play an important role in the information needs of care institutions. Changes in these laws and regulations have therefore consequences for these information needs, and thus data mining goals. The interviews in this research are performed in the first 2 months of 2012, all the changes in and new laws and regulations since then are not taken into account. Increasing expenditures in healthcare and the current economic situation cause more changes in laws and regulations for the complete healthcare sector. ECR software vendors should execute on-going research of those changes and translate it to the specific information needs and data mining goals.

## 7. BIBLIOGRAPHY

- Agrawal, R., Srikant, R., & others. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Databases vol.1215*, (pp. 487-499). Santiago de Chile.
- Azevedo, A., & Santos, M. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *Proceedings of the IADIS European conference data mining*, (pp. 182-185). Amsterdam.
- Berendt, B. (2012). More than modelling and hiding: towards a comprehensive view of Web mining and privacy. *Data Mining and Knowledge Discovery*, 1-41.
- Cadsand, R. v., Cazemier, G., Koster, J., Roose, J., & Vos, F. (2011). *De toekomst van de langdurige zorg*. Amersfoort: BMC.
- Central Bureau of Statistics. (2012, 4 19). *Sterfte; kerncijfers naar diverse kenmerken*. Retrieved from Statline CBS: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37979ned&D1=0-1,6-8,11,14,17,20-21,23-25,28-29&D2=0,5,10,15,20,25,30,35,40,45,50,57-1&HD=100614-1440&HDR=G1&STB=T>
- Centrum Indicatiestelling Zorg. (2012, 4 4). *Over het CIZ / CIZ*. Retrieved from CIZ: <http://www.ciz.nl/over-het-ciz>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*. SPSS.
- Chen, M., Han, J., & Yu, P. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data engineering vol.8(6)*, 866-883.
- CIBG. (2012, 4 4). *CIBG in het kort*. Retrieved from CIBG: <http://www.cibg.nl/overhetcibg/cibginhetkort/>
- Cios, K., & Moore, G. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine 26*, 1-24.
- Cios, K., Pedrycz, W., & Swiniarski, R. (2007). *Data mining: a knowledge discovery approach*. Springer Verlag.
- de Gooijer, J., & Hyndman, R. (2006). 25 years of time series forecasting. *International journal of forecasting vol.22(3)*, 443-473.
- Duin, C. v., & Garssen, J. (2010). *Bevolkingsprognose 2010-2060: sterkere vergrijzing, langere levensduur*. Den Haag: Centraal Bureau voor de Statistiek.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine vol. 17 (3)*, 37-54.
- Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management 37*, 271-281.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine Vol.13 (3)*, 57-70.
- Freitas, A. (2003). A survey of evolutionary algorithms for data mining and knowledge discovery. *Advances in evolutionary computing*, 819-845.
- Giraud-Carrier, C., & Povel, O. (2001). Characterising Data Mining software. *Intelligent Data Analysis 5*, 1-12.

- Goodwin, L., VanDyne, M., Lin, S., & Talbert, S. (2003). Data mining issues and opportunities for building nursing knowledge. *Journal of Biomedical Informatics* 36, 379-388.
- Hahsler, M., & Chelluboina, S. (2010). Visualizing Association Rules: Introduction to the R-extension package arulesViz.
- Harding, J., Shahbaz, M., Kusiak, A., & Srinivas, S. (2006). Data mining in manufacturing: a review. *Journal of Manufacturing Science and Engineering* Vol.128, 969-976.
- Het CAK. (2012, 4 4). *Het CAK - Organisatie*. Retrieved from Het CAK: <http://www.hetcak.nl/smartsite.dws?id=58621>
- Inspectie voor de Gezondheidszorg. (2011). *Organisatie*. Retrieved 12 15, 2011, from Inspectie voor de Gezondheidszorg: <http://www.igz.nl/organisatie/>
- Inspectie voor de Gezondheidszorg. (2012, 5 2). *Vragen over klachten*. Retrieved from Inspectie voor de Gezondheidszorg: [http://www.igz.nl/actueel/veelgestelde-vragen/vragen\\_over\\_klachten/](http://www.igz.nl/actueel/veelgestelde-vragen/vragen_over_klachten/)
- Koh, H., & Tan, G. (2011). Data Mining Applications in Healthcare. *Journal of Healthcare Information Management* vol. 19(2), 64-72.
- Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review* Vol. 21(1), 1-24.
- Lindell, Y., & Pinkas, B. (2002). Privacy Preserving Data Mining. *Journal of cryptology* vol. 15(3), 177-206.
- Lucas, P. (2004). Bayesian Analysis, Pattern Analysis and Data Mining in Health Care. *Current Opinion in Critical Care* vol. 10(5), 399-408.
- Lutz, W., Sanderson, W., & Scherbov, S. (2001). The end of world population growth. *Nature* vol. 412, 543-545.
- Marban, O., Mariscal, G., & Segovia, J. (2009). A data mining and knowledge discovery process model. *Data Mining and Knowledge Discovery in Real Life Applications*, 1-16.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2012, 01 09). *Regeling zorgaanspraken AWBZ*. Retrieved from Overheid.nl: [http://wetten.overheid.nl/BWBR0014855/geldigheidsdatum\\_09-01-2012#Bijlage2](http://wetten.overheid.nl/BWBR0014855/geldigheidsdatum_09-01-2012#Bijlage2)
- Ministerie van Sociale Zaken en Werkgelegenheid. (2010). *A short survey of Social Security in the Netherlands*. Den Haag.
- Ministerie van Volksgezondheid, Welzijn en Sport. (2011). *Programmabrief langdurige zorg*. Den Haag.
- Ministerie van Volksgezondheid, Welzijn en Sport. (2011). *Vaststelling van de begrotingsstaten van het Ministerie van Volksgezondheid, Welzijn en Sport (XVI) voor het jaar 2012*. 's-Gravenhage: Ministerie van Volksgezondheid, Welzijn en Sport.
- Ministerie van Volksgezondheid, Welzijn en Sport. (2011). *Voortgangsrapportage Hervorming Langdurige Zorg*. Den Haag.
- Morée, M. (2006). Combining work and long-term care in the Netherlands. *International Federation on Ageing 8th Global Conference*. Copenhagen: The Netherlands Institute of Care and Welfare.
- Mot, E., Aouragh, A., Groot, M. d., & Mannaerts, H. (2010). *The Dutch system of long-term care*. Den Haag: CPB Netherlands Bureau for Economic Policy Analysis.

- Nederlandse Zorgautoriteit. (2011). *The Dutch Healthcare Authority*. Retrieved 12 15, 2011, from Nederlandse Zorgautoriteit: <http://www.nza.nl/organisatie/sitewide/english/>
- Nederlandse Zorgautoriteit. (2012). *Beleidsregel CA-300-510: Prestatiebeschrijvingen en tarieven zorgzwaartepakketten*. NZa.
- Norèn, G., Hopstadius, J., Bate, A., Star, K., & Edwards, I. (2010). Temporal pattern discovery in longitudinal electronic patient records. *Data mining and knowledge discovery vol.20(3)*, 361-387.
- Onwubolu, G. (2009). An Inductive Data Mining System Framework. *IWIM*, 108-113.
- Prather, J., Lobach, D., Goodwin, L., Hales, J., Hage, M., & Hammond, W. (1997). Medical data mining: knowledge discovery in a clinical data warehouse. *Proceedings of the AMIA Annual Fall Symposium* (pp. 101-105). Nashville: American Medical Informatics Association.
- R Project. (2012, 05 25). *The R Project for Statistical Computing*. Retrieved from R-Project: <http://www.r-project.org/>
- R Studio. (2012, 05 25). *RStudio*. Retrieved from RStudio: <http://rstudio.org/>
- Rijksoverheid. (2011). *Welke zorg valt onder de Algemene Wet Bijzondere Ziektekosten (AWBZ)?* Retrieved 2 17, 2011, from Rijksoverheid.nl: <http://www.rijksoverheid.nl/onderwerpen/algemene-wet-bijzondere-ziektekosten-awbz/vraag-en-antwoord/welke-zorg-valt-onder-de-algemene-wet-bijzondere-ziektekosten-awbz.html>
- Rijksoverheid. (2012, 5 9). *Hulp en ondersteuning | WMO | Rijksoverheid*. Retrieved from Rijksoverheid: <http://www.rijksoverheid.nl/onderwerpen/wet-maatschappelijke-ondersteuning-wmo/hulp-en-ondersteuning>
- Rijksoverheid. (2012, 5 9). *Wat is een persoonsgebonden budget (pgb) en waar kan ik het voor krijgen?* Retrieved from Rijksoverheid: <http://www.rijksoverheid.nl/onderwerpen/persoonsgebonden-budget-pgb/vraag-en-antwoord/wat-is-een-persoonsgebonden-budget-pgb-en-waar-kan-ik-het-voor-krijgen.html>
- Rijksoverheid. (2012, 04 15). *Zorgzwaartepakket: beschrijving van de zorg*. Retrieved from Rijksoverheid: <http://www.rijksoverheid.nl/onderwerpen/zorgzwaartepakketten/zorgzwaartepakket-beschrijving-van-de-zorg>
- Schäfer, W., Kroneman, M., Boerma, W., van den Berg, M., Westert, G., Devillé, W., & van Ginneken, E. (2010). The Netherlands: Health system review. *Health Systems in Transition 12(1)*, 1-229.
- Treacy, M., & Wiersema, F. (1992). Customer Intimicy and Other Value Disciplines. *Harvard Business Review*, 83-94.
- Tweede Kamer der Staten-Generaal. (2012). *Vaststelling van de begrotingsstaten van het Ministerie van Volksgezondheid, Welzijn en Sport (XVI) voor het jaar 2012*. 's Gravenhage: Tweede Kamer der Staten-Generaal.
- van Rijn, A. (2011). *Het Nederlandse zorgstelsel in een notendop*. Deventer: Kluwer.
- Yin, R. (2009). *Case study research: Design and methods*. Sage publications, INC.
- Zichtbare Zorg. (2011). *Programma Zichtbare Zorg*. Retrieved 12 20, 2011, from Zichtbare Zorg: <http://www.zichtbarezorg.nl/page/Programma-Zichtbare-Zorg>

# APPENDIX I – INTERVIEW I

## CHARACTERISTICS

Function:	Board of directors member
Type of institution:	VVT
Number of locations:	6
Number of clients:	300 - 500 intramural / 100 - 300 extramural
Number of FTE:	400-500
Duration:	45 min

## SUMMARY

Due to some uncertainties about the appointment this interview was short, compared with the other interviews. Nevertheless, the interviewee kept it short in terms of information needs. The interviewee was interested in some basic information about the care-related information. The list of care-related information needs for this expert is short and additionally the expert was not interested in a daily update. Responsibility for care-related cases is located at the managers level. Therefore the information needs of this interviewee are:

- Production information
- Incident information
- Waiting list information
- Occupancy rate of clients
- ZZP-mix with a forecast of the future ZZP-mix
- Complaint information
- What kind of care does the client demand?

This information is high level. If irregularities are observed, the interviewee is interested in a deeper analyses of the above information. Incident and complaints information is used as a basic indicator about the quality of the institution. Information about the waiting list is important in order to create a stock to insure the care institution. Personnel is hired to support a full occupation. When full occupation cannot be achieved this has to have consequences for the personnel. This information needs is care-related but is used for financial safety of the institution. Occupancy rate is also used for the managing the business.

ZZP-mix information consist of information about the amount of care a client needs and the budget a care institution receives. It is important for management to know what the ZZP-mix is within their institution. High ZZP clients requires a different institution and personnel as low ZZP clients. Also for future investments it is good to know what kind of clients are supported. This has for example consequences for the construction of a new location.

Finally, the interviewee is interested in the care that a client demands. Personnel that is hired must be able to give the care that a client demands. From another perspective this has also to do with the finances. Clients that demand more care than their indication prescribes should consider to request a new indication.

## APPENDIX II – INTERVIEW II

### CHARACTERISTICS

Function:	Location manager
Type of institution:	VVT
Number of locations:	2
Number of clients:	100 - 300 intramural / < 100 extramural
Number of FTE:	
Duration:	2 hours

### SUMMARY

The interviewee is location manager of one of the locations of this care institution. The interviewee underlined my statement that care institutions do not know what kind of information is needed. Information is gathered and presented for stakeholders and the institution of the interviewee do exactly what comes to them. Another remark of the interviewee is that the size of the institution will affect the need for information.

Information about personnel with respect to the ZZP-mix is used at this moment. Because of this information it is clear that staffing exceeds the available time within an ZZP. But, because of the efficient structure of the institution it is affordable to deploy more personnel. This information is not used to steer the institution, as long as it is affordable. The information about staffing, absenteeism, production, ZZP-mix and occupancy is important for the institution to get insight in their internal business. Also the experience of client is very important to manage the institution.

The interviewee finds it very useful to use the data in order to subtract the possible cause of an incident. For example a client falls, but what is the cause of this incident? Is it because of age, high blood pressure, medication, or something else? Prediction of incidents would also add value, because it could prevent incidents. For example, client X has medication Y which could lead to fall incidents. With the prediction information, care personnel could take precautions. This creates an opportunity to use the data to exclude risks or at least minimize the risk for the clients.

Actions taken after risk identification are registered. This information can be used to create an understanding of the effectiveness of certain actions. This could increase the effectiveness of the actions taken after risk identification and incidents. Some actions will not decrease the amount of incidents where some actions will significantly decrease the amount of incidents. These last actions are more effective and could increase the quality of care. For example in case of fall incidents, after such an incident care personnel could choose to place a crash pad by the client. Such a crash pad does not decrease the number of incidents, it only reduce the impact. Care personnel could also choose to raise the bed rails of a client, with approval of the client. This action will significantly reduce the number of incidents, which makes it a more effective action. Managers of care institutions could therefore use the analysis to decide which actions should be performed after an incident occur or after the risk identification.

For every client treatment goals are created in dialogue with the client. Achieved treatment goals are important for the client and also for management because it improves the quality of care. The amount and the types of achieved and not achieved treatment goals is interesting and important to increase the quality of care.



## APPENDIX III – INTERVIEW III

### CHARACTERISTICS

Function:	Projectleader ECD & Care administrator
Type of institution:	VVT
Number of locations:	18
Number of clients:	700 – 900 intramural / 100 – 300 extramural
Number of FTE:	700 – 900 FTE
Duration:	1,5 hours

### SUMMARY

The care administrator is mainly interested in production summaries. Four types of production is important to know, which are: planned production, realized production, declared production and non-financed production. This information should be available per business unit. Other interesting numbers are the bed capacity and the number of care operations. Also the number of and the type of incidents is important to know. Not only because it is mandatory to justify itself towards the IGZ.

One of the interviewees stated that the average age of the clients is interesting information. At this moment it is impossible with their ECD software to subtract such information. The ECD projectleader does not see the added value of such information, because it is not used to improve the quality of care.

Connections with other software packages is seen as important for the future in order to create an even larger knowledge base about the clients. One can think of systems of hospitals, general practitioners, pharmacists, domotics and other systems and devices. For example a physiotherapist that create a training program for a rehabilitation client. The client has to train every day for one hour on the treadmill. The treadmill registers all the information and this information could be extracted from the device in order to control and steer the rehabilitation.

The interviewees are very interested in the direct and indirect client related time. Information about the time related to the client (direct or indirect) is important to increase the effectiveness and efficiency of the core process. Without the information about client related time it is very complicated, if not impossible, to improve the efficiency and effectiveness.

Another important topic that has been discussed is the ability to analyse the data by themselves. Standard reports are interesting, but the information needs are too complex to build all the reports to satisfy all those needs. Some basic information needs include: progress of clients, average length of stay, reason for admission and number of client per demand for care.

From next year, long-term care institutions will also work with the Diagnosis Treatment Combinations (Diagnose Behandel Combinatie; DBC), as it is used in the Dutch hospitals. These DBC's will be used for, for example rehabilitation. Rehabilitation treatments follow a certain treatment pattern. For each DBC it is interesting how this treatment pattern proceeds. This information need is used as an equivalent for obtained treatment goals, as it is used for normal AWBZ care.

## APPENDIX IV – INTERVIEW IV

### CHARACTERISTICS

Function:	Quality manager
Type of institution:	VVT
Number of locations:	4
Number of clients:	300 - 500 intramural / 300 - 500 extramural
Number of FTE:	400-500
Duration:	1 hour

### SUMMARY

The interviewee is quality manager and is interested in all information concerning the quality of care. Some additional information needs, that support mainly the financial state of the care institution, are interested for quality reasons.

- Care-related measures & incidents with respect to risk assessment
- Risk assessment interrelationships
- Age distribution of clients
- Average duration of stay & ZZP indication
- Development of the ZZP-mix (+ future predictions)
- Occupancy
- Average duration on waiting list
- Effectiveness of the actions taken
- Insight into the incidents & complaints
- Care within & outside ZZP indication
- Staffing with respect to ZZP & operations
- Progress of care-related measures
- Number of obtained treatment goals
- Average client-related time
- Unplanned care prognosis
- CQ-measure
- Incident prevention actions along with their effectiveness

Other data sources could be linked to get more understanding of the internal organization. One can think of the data of delusion detection systems and restaurant systems. If a large part of clients is ill, this could be of wrong food. With the data of the restaurant it is possible to understand this pattern.

Another thing that is discussed during this interview was the ability to intervene during the registration of information. When care personnel for example registers a new incident, the system could intervene and tell the care giver that this is the fourth incident in one week and that the client need more attention. Or that the care-related measure is out of sync. The care giver could than immediately take action.

## APPENDIX V – INTERVIEW V

### CHARACTERISTICS

Function:	Head finances
Type of institution:	VVT
Number of locations:	8
Number of clients:	500 - 700 intramural / 500 - 700 extramural
Number of FTE:	500 - 700
Duration:	1,5 hour

### SUMMARY

The interviewee expects that care institutions will disappear in the future. On one side care institutions will be absorbed in home care and on the other side in nursing homes. This means that the care institution of the interviewee focuses on the nursing homes and home care.

Every care institution must create an annual report and a social report every year. The interviewee stated that almost all the information that they include in the reports, is included because it is mandatory. This means that they would like to include other information, information that is really relevant for their stakeholders. At this moment the annual and social reports could be created on the basis of a directive. This directive is created by the Ministry of Health, Welfare & Sport, and consists of more than 300 pages. The social report especially, is aimed at the stakeholders or the (new) clients, which means that the content should be focused. Reports should link to the processes within the healthcare. At this moment this link is not in place.

The interviewee is interested in information about the incidents in order to minimize the number of incident and prevent incidents. Number of incidents is seen as a parameter for quality of care. Minimizing the number of incidents will therefore lead to higher quality care. In order to minimize the number of incidents it is important to analyse the effectiveness of follow-up actions and the cause of incidents. Risk assessment figures are interesting for analysing the impact of measures to prevent incidents. Some measures will be more effective than others. At this moment this knowledge is implicitly known by the care givers. By extracting the information from the collected data, it could be possible to make that knowledge explicit. CQ-index information is used to assess the quality of care and the experience of the clients.

Another important task for care institutions is to minimize the number of complaints of clients. The interviewee finds it very useful to use the collected data to prevent complaints, or at least note them previously. When the complaints could be noticed on forehand, the care institute could prevent that the client files a complaint. Other important figures are: production information, occupancy, waiting list information, age distribution over a period of time, client milestones, ZZP mix information and mutations (amount, type and causes). The mutations lead to fluctuations, which affects the processes of the care institution. Client milestones are milestones with respect to the age or marriage of the client. It is a milestone for clients that turn for example 100 years or are married 60 years. Also staffing with respect to ZZP mix and operations are important numbers in order to control the deployment of minimal qualified employees. Employee incidents are incidents that must be registered by the IGZ.

## APPENDIX VI – INTERVIEW VI

### CHARACTERISTICS

Function:	Board of directors
Type of institution:	VVT
Number of locations:	2
Number of clients:	100 - 300 intramural / < 100 extramural
Number of FTE:	300-400
Duration:	1,5 hour

### SUMMARY

The interview started with a discussion about the annual report. Every year the care institution creates an annual report, in which the institution is described. The content is largely the same every year. A new annual report is created by modifying the report of last year. The information that is described in the report is not interested for the institution itself, because they consult that information on a regular basis. This includes information like the number of clients, production information and absenteeism of personnel.

Complaints that will result in official complaints at the complaint commission are the complaints that are interested for the interviewee. Other complaints, for example that the food was not good enough, are handled by the location manager. Information about the incidents is of interest of the interviewee. One remark is that some incidents happen more frequently because clients do not want other measures in order to prevent incidents. These incidents are registered at this moment, but will cloud the understanding. Care institutions are required to publish the number of incidents and are judged on the number of incidents. This results in the fact that some care institutions do not register some incidents, in order to create a better picture of the institution. Minimizing the number of incident reports should not be the goal, the goal should be to improve the quality of care. In order to improve the quality of care it could be valuable to register all the incidents to get a good understanding of the real problem.

The interviewee is interested in the link between the staffing and the ZZP-mix, and indicated that this information is also one of the most important figures. In order to support the higher ZZP clients it is important to have a certain level of personnel. If the ZZP-mix changes, it will affect the staffing. The interviewee indicated also that it is not primarily used to create a more efficient staffing. Staffing should be within the budget, but it is not required to generate profit by reducing the manpower. Also the customer experience is an important information need of the interviewee. For insight in the customer experience, the institution of the interviewee uses the CQ index.

Because of the complexity of the care, the interviewee indicated that there should be a possibility to do their own analysis. When, for example, the number of medication incidents increases, the institution wants to have more insight in the problem. These analysis are situational and could therefore not be described on beforehand. The interviewee underpinned the fact that the large amount of data is not fully used at this moment.

The interviewee indicated that care personnel should not be alienated from the process. This means that the software should not intervene too much in order to prevent that the employees will act too lazy. When the employees analyse the data by themselves this result in more involvement with the client.

## APPENDIX VII – INTERVIEW VII

### CHARACTERISTICS

Function:	Assistant director
Type of institution:	VVT (Geriatric Rehabilitation Care)
Number of locations:	1
Number of clients:	40 rooms (>500 clients)
Number of FTE:	-
Duration:	1,5 hour

### SUMMARY

This institution is a rehabilitation institution, which is not a usual care institution. Clients stay for a short period of time in this institution, which is different than the long stay in a normal care institution. Within this rehabilitation institution the care is aimed at the rehabilitation of the client. Before the client enters the institution a clear and measureable goal is created which has to be achieved. In order to increase the quality of care, the institution should have insight in the number of obtained treatment goals. The treatment goal is obtained when the treatment was a success and the client is pleased. Information about treatment goals is important for the interviewee. Not only information about the obtained treatment goals, but also information about the not-obtained treatment goals. Information about the reason behind the not-obtained treatment goals is information that could create added value in order to increase the quality of care.

Besides the information about the treatment goals, the interviewee is interested in the customer experience. This information is gathered by surveys. This survey is derived from the CQ index, that means that parts of the CQ index are used for their own internal tool to measure the customer experience.

The interviewee indicated that these are the most important information needs to manage the quality of care. Information about the production, waiting list and staffing is information that is handled by the controller. Because this institution is differently financed than general care institutions, the interviewed institution can focus on the quality of care. Care provided by this institution is partly private financed and partly financed by the health insurance (*Zorgverzekeringswet; ZVW*). Besides the institution there is a friends foundation. 90% of the clients treated by the institution are member of this foundation. This makes it possible to manage primary on the quality aspects within the financial possibilities. Another difference between this institution and a 'normal' care institution is the composition of management. Only the board of directors function, the institution consists of a one headed board of directors, is a full-time employee. The interviewee and the financial controller are both part-time employees.

Other information that is of interest of the institution for managing the institution is the client distribution within the institution. When a large proportion of clients is female the institution should arrange other facilities than when a large proportion of clients is male. This makes it an interesting, not very important, information need that is also in place to increase the quality of care.

## APPENDIX VIII – INTERVIEW VIII

### CHARACTERISTICS

Function:	Board of directors
Type of institution:	VVT
Number of locations:	4
Number of clients:	300 - 500 intramural / 300 - 500 extramural
Number of FTE:	400-500
Duration:	1 hour

### SUMMARY

The interviewee indicated that the registration of information is important if the information is used to improve the process. If the information is not used, it is useless to register that information. The information used by the interviewee to manage the internal institution and processes is:

- Production information
- Financing information
- Personnel & volunteer formation
- Waiting list information & information about the flow of clients
- Absenteeism information
- Complaints information
- Incident information
- Coercive measures information

At this moment the institution has the goal to reduce the number of incidents and coercive measures with 15%. That means that information about the incidents and coercive measures is important. Other information that could help management to manage the internal business is information about the occupancy, production and staffing. For staffing one can think of the absenteeism and the formation of qualified personnel. Daily information is not interesting for the interviewee, these information is used by care givers.

The interviewee finds it very useful to have the possibility to forecast some scenarios. For example, changes in laws & regulations could affect the care institutions. At this moment residence is financed in lower ZZP's. When residence is not financed anymore in lower ZZP's this has impact for the care institutions. The interviewee would like to use the gathered data to forecast different scenarios to create an understanding of the impact of this change.

Ambition of the interviewed care institution is to deliver the best care in 2014. In order to achieve this goal the customer experience is measured, with 2 questions, every half year. The goal is to get in 2014 more than 80% 'yes' answers on the first question. The questions are: Would you recommend our care institution to your friends and family? & What score do give our care institution?

The interviewee indicated also that only the necessary data should be collected, which is hard to capture. If the basic set of information represent some deviations, it is necessary to zoom in. This zooming creates a more complex information need.

## APPENDIX IX – INTERVIEW IX

### CHARACTERISTICS

Function:	Quality manager
Type of institution:	VVT
Number of locations:	2
Number of clients:	< 100 intramural / 100 - 200 extramural
Number of FTE:	< 100
Duration:	45 min

### SUMMARY

This interview did not run smoothly what affects the outcome of it. The interviewee works for a small care institution with less than 100 intramural clients. This makes computerization less important, since almost all clients are known. The interviewee works as quality manager at the institution and is responsible for increasing the quality of care. Improvements come from all sides and are registered and maintained. Systems and processes are in place to get grip on the quality and to improve it.

- External audit
- Internal audit
- Quality team
- Risk assessment institution-wide
- Client council
- Complaints committee
- Other researches
  - CQ
  - Supplier assessment
  - Care-related measurements

Improvements come from these systems and processes. Also other errors within the institution are recorded as improvements. The quality manager is responsible for good monitoring of these improvements. If improvements are not properly picked up, then the quality manager intervenes. When there is a need to organize a research for a certain problem within the institution, this will be undertaken by the quality manager. The data that is being gathered during that research will be analysed afterwards to come up with improvements.

Specific information needs regarding the data stored in the ECD have not been discussed. The interviewee is interested in improving the quality of care. The information needed to complete this task comes from the systems and processes which are in place and mentioned in the enumeration above. Risk assessment, complaints, CQ information and care-related measurements can still be read from the enumeration.

## APPENDIX X – INTERVIEW X

### CHARACTERISTICS

Function:	board of directors member
Type of institution:	VVT
Number of locations:	16
Number of clients:	> 2000 (intramural & extramural)
Number of FTE:	> 1000
Duration:	1,5 hour

### SUMMARY

The interview started with a discussion on the current project of the Ministry of Health, Welfare and Sport to create competition in the care sector in order to make care affordable. However, the interviewee says that the growing demand for care and the stable supply of care make it impossible to create such a market, at this moment. All care institutions have waiting lists, which are kind of the insurance of care institutions. Care personnel is deployed on the basis of full occupancy. When full occupancy cannot be achieved it has consequences for care personnel. These dynamics is difficult for care institutions because they have contracts with their personnel. So, the waiting list is very important for a care institution.

One of the questions the interviewee had in order to raise the quality of the delivered care was: “How do you keep the care personnel as much as possible with the client?”. Care personnel should be walking outside with the client, instead of registering things about the care and client. The quality of care is very hard to measure. Currently, the CQ measure is used to get insight in the user experience. This is a questionnaire for clients, or responsible person, with questions about their experience in the care institution. Some care offices use this measure to judge care institutions and use it during the budget discussion. Clients can for this reason be instructed to fill in the CQ questionnaire in a more positive way, in order to avoid the possibility that the care institution will receive less budget. An argument towards the clients could be that a bad score results in less budget and thus in less personnel. This means that the clients, in a way, determine the quality of care. Using the CQ measure to judge the quality of a care institution is not desirable. Institutions that score worse on this CQ measure receive less budget than good scoring institutions. Improving the quality with less budget is a hard task. Another aspect is that the CQ is measured two yearly.

Some of the information needs the interviewee had are: production figures on locations or other organizational unit, amounts of products, client-related time / non client-related time, travel time (direct time), schooling (indirect time), personnel costs. These information needs are all aspects that support the financial goal of a care institution. Other information needs were about the risks, incidents and complaints of the clients. At this moment it is not known what the most common causes of incidents are and what the most common actions after an incident are. However, for every incident the caregiver must register in plain text what the cause is and what action he/she has taken. In order to create some reports it would add value to create a standard list of causes and actions that can be selected.

The interviewee has made a map for the care personnel to create a care plan for each client. Care plans are created on an individual client basis. This means that the care plan is not digital. Reason for this is the clarity of such a written care plan for the care personnel.



## APPENDIX XI – INTERVIEW XI

### CHARACTERISTICS

Function:	Chairman of board of directors
Type of institution:	VVT
Number of locations:	6
Number of clients:	> 1000 intramural / > 2000 extramural
Number of FTE:	> 1000
Duration:	1,5 hour

### SUMMARY

The interviewee is chairman of the board of directors of a wide care institution. More than 3000 clients in total and also a wide variety of care that is offered. Besides elderly care the institution is for example also provides maternity and child health care. Managing the institutions property is being done by their own housing association.

Target group analyses are not required at all times. When the institution is planning a new location or other major investments are explored, target group analyses could be very interesting. This is also the case for demographics of a certain location. Also the National Institute for Health and Environment (Rijksinstituut voor Volksgezondheid en Milieu; RIVM) is doing research on the target groups in the Netherlands. These analyses could be used by the institution when they explore new activities. Feasibility could then be determined using secondary information together with their own (primary) information.

The integration of ECD software is estimated by the interviewee at 25%. This means that still 75% of the care institutions in the Netherlands have not been provided with an ECD. The market of ECD software suppliers is therefor far from saturated. Which also is a sign that ECD software is still not mature enough. A large percentage of the care institutions do not see the added value of an ECD, do not have the money for an ECD or are too small for such a tool.

Also this institution is almost done with creating a Business Intelligence tool to get insight in their own organization. Currently, the institution do not use a full ECD with a care dossier and treatment dossier. The data warehouse consist therefor of data from the HRM and finance software. Combining this information results in a dashboard with up-to-date information for management.

In order to get insight in the core process of the institution there are some processes in place. One of them is the care plan consultation with the client twice a year. During this consultation the client may indicate the care he or she wishes the coming half year. This is also the moment to get feedback from the client about their experience in the institution. Information from these consultations is qualitative and is hard to make quantitative and do analyses on the data. The care plans of the clients nevertheless, are also important for internal steering. These plans are normally registered in an ECD (which is not the case for this institution). The interviewee thinks it is a good idea to measure the quality of care during the care plan consultations.

Another important process for managing the internal business is to spend a day with the care personnel. This gives the best insight into what care personnel must do and how the clients experience this. Quality of care is very subjective and hard to get grip on. Spending a day with care personnel provides valuable insight and knowledge.

The information needs of the interviewee are also based on the financial state of the institution but also on care related topics. The interviewee came up with the following information needs:

- Care plan information
- Incident reports (Incidents, MIC, MIM)
- Complaints
- Reports of the confidential counsellor
- Amount of coercive measures
- CQ Index information
- Non-attendance information in relation with other information
- Per location the mix of personnel and ZZP

Above information needs are the basic information needs of the interviewee and is not the complete in-depth information needs. These basic information needs are important for getting an overview of the situation. If further analysis is needed, the information needs are more extensive. The interviewee indicate that in their opinion the responsibility of quality should be on the lowest level as possible. That enables the institution to immediately seize and resolve errors made by care personnel.

## APPENDIX XII – INTERVIEW XII

### CHARACTERISTICS

Function:	director
Type of institution:	GGZ
Number of locations:	7
Number of clients:	100-300 intramural / unknown extramural
Number of FTE:	100-200
Duration:	2 hours

### SUMMARY

This institution is a private company which is licensed to give AWBZ care. The institution gives supportive care to mentally handicapped. They have a tight defined target group on which they focus. Now and then it is important to check how the current client group relates to the target group. Client information and prognoses of the CBS are important for these checks.

At this moment the institution is almost done with the development of a Business Intelligence solution. This has been developed in cooperation with an external organization. For this BI solution they have identified all the information needs of the organization. These information needs are nothing more than visualising data in a structured and understandable way. This means that the raw data has to be displayed in an user-friendly way. The information needs that have been defined are:

- Actual FTE per location and total FTE
- Total gross hours per location
- Total net hours per location
- Jobs per location / expert levels
- Average level of salary per expertise
- Payroll overhead (general salary expenses)
- Payroll care (care-related salary expenses)
- Number of intramural clients (per location/overall/zzp)
- Number of extramural clients (per location/overall/products)
- Number of care days (per location/overall/zzp)
- Number of care hours per extramural client (location/overall)
- ZZP-budget per location and overall
- Extramural production per location and overall
- Number of new clients
- Number of leaving clients
- Number of days between application and first intake (registration delay) compared to the standard
- Number of days between first intake and placement (placement delay) compared to the standard
- Number of hours client-related time (allocated to expertise levels)
- Number of incidents

- Causes of incidents
- Number of hours care provided by subcontractors
- Turnover (ZZP-budget, extramural budget, capital cost allowance, reallocation)
- Average ZZP per location and total
- Average ZZP rate per location and total
- ZZP distribution per location and total
- Costs (personnel costs, expenses, client-related costs, capital costs, depreciation, overhead)
- Total number of care FTE per location
- Total number of required care FTE per location
- Comparison between actual number of care FTE per location and required care FTE per location
- Occupancy per location and total
- Percentage overhead
- Total client-related time per location and total on the basis of care arrangements
- Total net working hours per location and total
- Percentage client-related time

This is a long list of the information needs that are identified in order to get insight in the organization. The data can be extracted from different applications, for example: Electronic Client Record software, Financial software, Planning software and Human Resource Management software. The information will be visualised using a Business Intelligence dashboard.

All the information needs are related to the financial goal of the care institution. This does not mean that the institution is not aware of the care quality. The responsibility lies on another level and if some major problems occur the director(s) will be informed. The interviewee finds the way in which the quality information of a care institution is being used too negative. If a care institution scores low (or lower than other institutions) this will be taken into consideration during the budget negotiations. This results in an apparent reality wherein institutions will inform the clients that they need to assess the quality of care good. That means that the care institution scores better on quality and it will receive more (or at least not less) budget.

The quality of their institution can also be assessed using the achieved treatment goals. Personnel will establish treatment goals in consultation with the clients. If such a treatment goal is achieved this means that the client can be satisfied. The next step is to establish a new treatment goal.

## APPENDIX XIII – INTERVIEW XIII

### CHARACTERISTICS

Function:	Operations division manager & Project manager
Type of institution:	GZ
Number of locations:	60
Number of clients:	700 – 900 intramural / 900 - 1000 extramural
Number of FTE:	500 – 700
Duration:	1,5 hour

### SUMMARY

The care institution has a clear target group defined. This target group are limited auditory and non-congenital brain injury clients. The interviewee indicated that this is a very narrow target group, especially because this market is relatively stable. For VVT institutions the market is growing when the number of elderly grows, this is not directly the case for the target group of this institution. The target group of this institution does not focus on the age of clients, that means that clients at every age could be supported by this institution. Both interviewees indicated that care institutions will go up in nursing institutions and home care. The operations division manager has an IT background and indicated that the use of data mining is not (fully) applied to long-term care at this moment.

First of all, the care institution is not managed based on daily or weekly information. Information on a monthly basis is good enough to manage the internal organization. Information that is most important are the production figures to determine the turnover and expenditures. Expenditures consist for 80% staffing costs. Other important and interesting numbers are the ZZP-mix information related to staffing in order to manage the deployment of personnel. Personnel of the correct discipline at an adequate qualification should be deployed. Not only to control the expenditures, but also in order to control the quality of care. These are the information needs that are of interest of the operations division manager on a monthly basis.

Besides the information that is needed on a monthly basis indicated by the operations division manager, the absenteeism of personnel is important to know. Not only the percentage of absenteeism, but also the frequency and insight in short- and long-term absenteeism. Other information about personnel, such as contract information and deployment of personnel. The operations division manager stated that all the expenditures of a location should be linked to the location, in order to calculate the efficiency of a location. Also cost prices of all the different products delivered by the institution are of interest of the operations division manager.

The project manager indicated that the treatment goals of clients are very important to manage. On one side it is important to track the progress of treatment goals. On the other side it is important to track the downfall of clients. Every half year the client and the care givers negotiate about the desired care (care plan) and the treatment goals. Care personnel report frequently about the current situation of the client including the situation with respect to the treatment goals. With this information it should be possible to manage the progress of the treatment goals. Also the information about the abilities and the disabilities of the client are reported. This information could be used to adjust the care at forehand for the clients. Care related measures of the client could support the decisions about the needed care.

Other information, related to the risks and incidents is indicated as important by the interviewees. The link between risk assessment and incidents, and more interesting the actions taken after an incident, is very important to improve the quality of care. At this moment there are no standard actions defined. This means that heterogeneity of the data prevents to do proper analysis. In order to enable proper analysis, standard actions should be defined.

An interesting statement during the interview was about the risk assessment. In care institution with a large proportion of dementia clients, it is obvious that a large proportion of clients have a higher risk at dementia. The interviewee indicated that a care institution should look at a more elaborate risk assessment in order to enable more complex analysis. New clients could be diagnosed at forehand based on historical information about the risk assessment.

Information about how the client is financed, is important in the near future. In the near future the laws & regulations will change which means that some care will shift from, for example, AWBZ to the ZVW. The interviewees would like to have an overview of the clients and the way in which the care is financed.

## APPENDIX XIV – INTERVIEW XIV

### CHARACTERISTICS

Function:	specialized inspector
Type of company:	Healthcare Inspectorate (Inspectie voor de Gezondheidszorg; IGZ)
Duration:	1 hour

### SUMMARY

The interviewee works for 4 years at the IGZ, with the focus areas on care institutions, nursing institutions and home care. In the past the interviewee has worked for more than 10 years as director of a care institution. The IGZ maintains around 25 laws and their work is based on laws and regulations and field standards. The field standards are standards and guidelines are the standards that have been put-up for professions and the IGZ reviews those field standards.

Information that could be requested by the IGZ is inexhaustible. IGZ could request information for every standard, guideline, law or regulation. For example, if clients receive medication that is not prescribed (particularly risky and restricted actions) there must exist an execution request of the general practitioner. The IGZ would like to receive an overview of the execution requests and an overview of the restricted actions. This is one of the many examples that are possible, in order to verify that care institutions comply with the field standards and laws and regulations. Also the risk assessment information, incident information, restraints, means and measures, ZZP information with respect to staffing, and other information about the client (for example dental care arrangements and decubitus information). Dental care is at this moment a hot item for IGZ. The number of dental visits is then a basic figure that could help the IGZ to inspect the care institutions. ZZP information with respect to staffing is for IGZ seen from the quality perspective. Clients with a higher ZZP need more extensive care which should be performed by trained personnel.

If care personnel identifies an issue, then care personnel should report on that issue. IGZ is interested in this information. Another interesting figure is that of blood pressure in combination with the medication. These reports, as stated before, are inexhaustible. One could also think of the risk of malnutrition, the weight measure is registered in an ECD. If these values decrease every time, the system could inform care personnel that they should keep track of the client.

The interviewee stated that less than 50% of the care institutions have access to ECD software. Some of the employees of the IGZ even indicated that only 10% of the care institutions use ECD software. Lack of uniformity in ECD software is a problem for the IGZ, and most likely also for other stakeholders that require information. When IGZ requests information from care institutions, they are dependent on the way ECD's register data.

For care institutions it is important to collect the information from the hospitals, if a new client transfers from the hospital to the care institution. When the client enters the care institution, the care institution is for 100% responsible for the client. This means that the care institution should have all the required information of the client available. The information history of a client could also help the care giver in the delivery of high quality care.

## APPENDIX XV – INTERVIEW XV

### CHARACTERISTICS

Function:	senior advisor
Type of company:	knowledge centre within the long-term care sector
Duration:	45 min

### SUMMARY

This interview can be left outside of the information needs study. This organization does not request information directly from the care institutions. It is a knowledge centre that could support care institutions to improve the quality of care.

The interviewee started by telling something about the Zorgmonitor of Unit4's Cura ECD software. This is a Business Intelligence solution on top of ECD software, which enables care institutions to get a visualized view of their institution. The Zorgmonitor is a standardized tool, but also personalized reports are possible. A lot of care institutions use this software according to the interviewee.

During the interview we discussed the possibility of care institutions to measure their quality. The interviewee came up with a couple of important measures that could indicate the quality of care. These measures are: treatment of the client, medical quality by the amount of incidents, hospitality, has the client the feeling that someone is listening to him or her, is the life story of the client known and other aspects of perception. All these points are part of the CQ Index measure which should be conducted every 2 year.



## APPENDIX XVI – INTERVIEW XVI

### CHARACTERISTICS

Function:	quality policy advisor & innovation manager
Type of company:	Care insurer
Duration:	1,5 hour

### SUMMARY

From 2013, the care insurers are responsible for implementing the AWBZ care. Care insurers will then purchase care for their clients at the care institutions. In order to fulfil this task it is important to have some information about the care institutions. Because care insurers will be responsible for purchasing care for their clients, it is important to have purchase information. This information consists largely of quality information. Another part is the financial part of the process. The care insurer should pay the care institution and therefore it is important to know the production information of the care institutions.

The interviewees state that the first year (2013) the current process will be changed slightly. This means that the information that should be delivered to the care offices at this moment, is almost the same information that the care institutions should deliver to the care insurers next year. Care insurers will use the production information of the care institutions to finance the care institutions. Quality information is available in the form of the CQ index.

In the future the care insurers will go to a process in which the care institutions need to determine their own quality of care. If the care institution can indicate that their quality is better than other care institutions, they could negotiate about the fees. Different measurement tools (such as RAI, PROM and CQ) are in place to determine the quality of a care institution. At this moment the CQ index is used to determine the quality. The CQ index is also used by care offices in the negotiations about the fees. The objectivity of the CQ index is often doubted. The interviewed care insurer wants to go to a model in which care institutions get a standard fee, that could be increased.

The production information used by the care insurers to pay the care institutions is handled by AZR messages. These AZR messages are used to implement mutations and to transfer the production information. For each mutation, the care institution should send an AZR message to the care office. A reply message will return to the care institution to verify the receipt. If the client receives a (new) indication, the care office will send an AW message to the care institution, that in return needs to send an AW message to verify the receipt. Other messaging standards are in place to inform the care insurers about the waiting lists of care institutions.

Another interesting point that has been discussed during the interview is that the care insurer wants to decrease the administrative burden of care givers. This could result in more time for the client and cost savings.

## APPENDIX XVII – INTERVIEW XVII

### CHARACTERISTICS

Function:	program manager
Type of company:	Dutch Health Inspectorate (Inspectie voor de Gezondheidszorg; IGZ)
Duration:	1 hour

### SUMMARY

The Dutch Health Inspectorate is part of the Ministry of Health, Welfare and Sport. This inspectorate promotes public health by enforcing the quality of care, prevention and medical products. All care in the Netherlands should be checked with approximately 250 inspectors. This means that it is impossible to inspect every care institution on yearly or even a half-yearly basis. The IGZ focusses therefore on risk assessment. Care institutions with a higher risk, based on some variables, are more likely to be inspected than institutions that are stable. Based on this knowledge, IGZ needs certain information about the risk that a care institution runs.

In 2007, the Ministry of Health, Welfare and Sport established the Program Visible Care (Programma Zichtbare Zorg). Two years later, in 2009, the IGZ and the Ministry concluded that this program could not meet the expectations. They created a new framework in order to assess the risk of care institutions, that is based on 4 pillars: Care organization continuity information, System information, Care-related information and CQ Index information. Information needed to assess the position on those 4 pillars is described below. This is also the information that will IGZ demand from the care institutions (in the near future).

#### **Care organization continuity**

- Solvability / Liquidity
- Turbulence within the Board of Directors
- Vacancies
- Absenteeism
- New housing / mergers
- Large variations in the middle management

#### **System**

- Complaints
- Incident reports (clients / personnel) (MIC / MIM)
- Risk assessment
- Client council
- BHV
- Care reports

### **Care-related**

- General Daily Necessities (ADL) follow-up system
- Fall incidents
- Use of psycho-pharmaceutics
- Freedom restrictive measures
- Medication errors
- Participation in activities
- Depression

### **Consumer Quality Index**

- Participation
- Information
- Treatment

Based on all the above information the IGZ will categorize care institutions on a risk scale. Care institutions that run a high risk will need to prepare for a visit by the IGZ. Care institutions running a medium risk and a low risk are not necessarily free from a visit. IGZ will also continue the random control of care institutions.

Information demands should, from the perspective of IGZ, be made based on a “need to know” basis. At this moment the stakeholders in the field of care institutions demand information on the basis of “nice to know”. Result is that care institutions register all kind of information that is only registered because of those kinds of demands.

At this moment the data that is handed in by the IGZ is edited data. This means that the collected data from for example the CQ Index measure is transformed into a nice overview of the outcome. The mean of response per question is being translated into a 5-point scoring scale. IGZ is not content with this current method and wants raw data instead so they can do the statistics themselves.

Another important change within the long-term care is the role that care insurers get next year (2013). At this moment all the care insurers out-sourced their task to purchase care. From January 2013 this task will be carried out by care insurers, which give them more power in the sector. The care insurers are responsible for purchasing care for her clients. Care is purchased by the care insurers (at this moment care offices) because individual people are too vulnerable, especially elderly people. The care is being purchased collectively by the care insurers for her clients.

## APPENDIX XVIII – INTERVIEW XVIII

### CHARACTERISTICS

Function:	two senior policy makers
Type of company:	Ministry of Health, Welfare and Sport
Duration:	2 hours

### SUMMARY

First of all, the interviewees indicated that the Ministry of Health, Welfare and Sport does not demand (extra) information from care institutions. Information, that is used by the ministry, is nothing more than the information available from the IGZ or other authorities. The interviewees endorsed that the ministry is in place to create policies and not to demand all kinds of information from the care institutions.

Information, nevertheless, is important for good selection information, improve information, supervisory information and purchasing information. These four types of information use must be fully implemented and utilized.

- **Selection information**

At this moment the selection information is being used by kiesBeter.nl. That is a public care portal where inhabitants of the Netherlands can find answers in the field of health care. Another interesting functionality is to compare different care institutions with each other. kiesBeter.nl is made by the National Institute for Health and Environment (Rijksinstituut voor Volksgezondheid en Milieu; RIVM). Ministry of Health, Welfare and Sport is the initiator of this portal and the only financier of kiesBeter.nl.

- **Improve information**

Information to improve the care is handled by the program Zorg voor Beter. This program is also an initiative of the Ministry of Health, Welfare and Sport in consultation with branch organizations, professional associations and client organizations. Zorg voor Beter is being directed by ZonMw. ZonMw is an organization that funds health care research and stimulates the use of it. Main contractors of ZonMw are the Ministry of Health, Welfare and Sport and The Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; NWO).

- **Purchasing information**

Care offices use the information, collected by the care institutions, at this moment to purchase the care. As already said before, the care insurers will take over this roll and will then use the information. Care Insurers Netherlands (Zorgverzekeraars Nederland; ZN) is the umbrella party for both care offices and insurers. The information that these stakeholders demand from care institutions is not completely generic. At this moment, AZR messages hold the information that care offices need. Some care offices use a slightly modified AZR standard, which makes it more complex for care institutions and software suppliers. AZR is a standard which is managed by the Ministry of Health, Welfare and Sport.

- **Supervisory information**

Supervisory information is collected and used by the IGZ. The information that the IGZ demands is information in order to supervise all the care institutions. The interviews with the IGZ are more elaborate on this topic.

Care institutions have different systems and processes in place to fulfil the current demand for information. AZR messages, CQ Index information, annual report and reports with regard to Zichtbare Zorg are all in place to meet the information demand from stakeholders. The AZR method is still being developed into a more strategic system. At this moment the method is too complex and old fashion. The whole flow of messages is predefined and every variety in that standard flow will cause a problem. In order to work like a care chain it is important to make it a bit more flexible.

The Ministry of Health, Welfare and Sport is policy maker in the field of health care, but rules are necessary to provide the sector some guidelines. The care institutions are insecure about what they have to do when the rules leave too much room. This is at least the opinion of the interviewees. That is the reason for the ministry to make more detailed rules and regulations. The ministry is also the founder of competition in the health care sector. Result of this competition is the sector as it is now. Care institutions are not communicating with each other because they are competitors now. Communication between care institution could give the institutions new insights in problem solutions and other ideas. This is very important for the care sector. The ministry therefore tries to bring the care institutions together to let them communicate.

With the new care client right law (Wet Clientrechten Zorg; WCZ), the ministry wants to lift the quality of care to a higher level. One of the important changes to make this possible is by making a member of the board of directors responsible for quality. At this moment quality is not the biggest issue within a care institution. The responsibility for quality lies by the care personnel and the board of directors trust them. When some major (or in some cases less major) problems occur the board of directors intervenes. With the new law, a member of the board of directors is responsible for quality, the ministry hopes that this responsibility improves the quality of care.

# APPENDIX XIX – INFORMATION NEEDS

Quality of care	(Board of) directors							Management					Stakeholders					
	Interview I	Interview V	Interview VI	Interview VII	Interview VIII	Interview X	Interview XI	Interview XII	Interview II	Interview III	Interview IV	Interview IX	Interview XIII	Interview XVIII	Interview XIV	Interview XVII	Interview XVI	Interview XV
Number of clients		✓	✓	✓				✓	✓	✓	✓		✓					
Age distribution of clients		✓		✓					✓	✓	✓							
Percentage male/female				✓						✓	✓							
Client milestones (ageing & marriage milestones)		✓		✓						✓								
Progress of clients		✓		✓	✓			✓		✓	✓							
Average duration of stay			✓		✓				✓	✓	✓							
Reason for admission										✓								
Registration delay (# days between application and first intake)																		✓
Placement delay (# days between first intake and placement)																		✓
Number of clients per demand for care	✓								✓	✓	✓							
Number of clients at an increased risk		✓					✓		✓	✓	✓	✓	✓		✓	✓		
Types of risk the clients run		✓					✓		✓	✓	✓	✓	✓		✓	✓		
Progress of care-related measures		✓							✓	✓	✓	✓	✓		✓	✓		
Number of incidents occurred	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
Types of incidents occurred	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
Causes of the occurred incidents	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
Incidents with respect to risk assessment		✓				✓	✓	✓	✓	✓	✓		✓		✓	✓		
Incidents with respect to care-related measures		✓						✓	✓	✓			✓		✓	✓		
Incidents with respect to pharmaceutical use		✓						✓	✓	✓			✓		✓	✓		
Patterns in occurred incidents	✓	✓				✓	✓	✓	✓	✓	✓		✓		✓	✓		
Number of complaints from the clients	✓	✓	✓		✓	✓	✓	✓			✓	✓			✓	✓		
Types of complaints from the clients	✓	✓	✓		✓	✓	✓	✓			✓	✓			✓	✓		
Causes of complaints from the clients	✓	✓	✓		✓	✓	✓	✓			✓	✓			✓	✓		
Pharmaceutical use by the clients									✓	✓	✓				✓	✓		
Treatment goals (obtained & not-obtained)				✓					✓	✓	✓		✓		✓	✓		
Care plan information								✓		✓	✓		✓		✓	✓		
Customer experience	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	
Reports of confidential counsellor													✓					
Client-related time & non client-related time		✓			✓	✓		✓				✓						
Activities of Daily Living tracking													✓					
Number of coercive measures					✓		✓		✓	✓	✓				✓	✓		
Types of coercive measures					✓		✓		✓	✓	✓				✓	✓		
Reason for coercive measures					✓		✓		✓	✓	✓				✓	✓		

	<b>(Board of) directors</b>								<b>Management</b>					<b>Stakeholders</b>					
	Interview I	Interview V	Interview VI	Interview VII	Interview VIII	Interview X	Interview XI	Interview XII	Interview II	Interview III	Interview IV	Interview IX	Interview XIII	Interview XVIII	Interview XIV	Interview XVII	Interview XVI	Interview XV	
<b>Financial state</b>	Production information (planned, realized, declared)	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓				✓	
	Care within & outside ZZP indication (planned, realized)	✓	✓	✓		✓	✓	✓	✓		✓	✓		✓					
	Unplanned care	✓	✓	✓		✓	✓	✓	✓		✓	✓		✓					
	Number of mutations		✓			✓												✓	
	Types of mutations		✓			✓												✓	
	Causes of mutations		✓			✓													
	ZZP-mix (per business unit)	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓					
	ZZP-mix prognosis	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓					
	Average duration of ZZP			✓		✓						✓							
	Number of clients on waiting list	✓	✓	✓		✓					✓	✓						✓	
	Average duration on waiting list	✓	✓	✓		✓					✓	✓						✓	
	ZZP-mix on waiting list	✓	✓	✓		✓					✓								
	Current occupancy	✓	✓	✓		✓			✓	✓	✓	✓							
	Occupancy prognosis	✓	✓	✓		✓			✓	✓	✓	✓							
	Short-term absenteeism personnel			✓		✓		✓		✓				✓			✓		
	Long-term absenteeism personnel			✓		✓		✓		✓				✓			✓		
	Frequency absenteeism personnel			✓		✓		✓		✓				✓			✓		
	Staffing (per business unit)		✓	✓		✓	✓	✓	✓	✓	✓			✓					
	Staffing with respect to ZZP-mix	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
	Staffing with respect to operations		✓	✓		✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
Bed capacity										✓									
Operations per function		✓	✓		✓	✓	✓	✓	✓	✓			✓		✓				
Operations per ZZP	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓		✓				

## APPENDIX XX – INFORMATION NEEDS BY SIZE

		Small (< 500 clients)	Medium (500 - 1000 clients)	Large (> 1000 clients)
<b>Quality of care</b>	Customer experience	2	5	4
	Number of incidents occurred	2	4	4
	Types of incidents occurred	2	4	4
	Causes of the occurred incidents	2	4	4
	Number of complaints from the clients	3	2	3
	Types of complaints from the clients	3	3	3
	Causes of complaints from the clients	3	3	3
	Patterns in occurred incidents	2	3	4
	Incidents with respect to risk assessment	2	2	4
	Number of clients	3	3	2
	Incidents with respect to care-related measures	2	2	3
	Incidents with respect to pharmaceutical use	2	2	3
	Number of clients at an increased risk	2	2	3
	Types of risk the clients run	2	2	3
	Client-related time & non client-related time	1	3	3
	Progress of care-related measures	2	2	2
	Treatment goals (obtained & not-obtained)	2	3	1
	Progress of clients	1	4	1
	Care plan information	1	2	2
	Number of coercive measures	1	3	1
	Types of coercive measures	1	3	1
	Reason for coercive measures	1	3	1
	Age distribution of clients	1	3	1
	Average duration of stay	1	3	0
	Number of clients per demand for care	1	3	0
	Pharmaceutical use by the clients	1	2	0
	Client milestones (ageing & marriage milestones)	0	2	1
	Percentage male/female	0	3	0
	Activities of Daily Living tracking	0	1	1
	Reports of confidential counsellor	0	0	2
Registration delay (# days between application and first intake)	1	0	0	
Placement delay (# days between first intake and placement)	1	0	0	
Reason for admission	0	1	0	



		ZZP-mix		
		Small (< 500 clients)	Medium (500 - 1000 clients)	Large (> 1000 clients)
<b>Financial state</b>	Staffing with respect to ZZP-mix	3	4	4
	ZZP-mix (per business unit)	3	4	4
	ZZP-mix prognosis	3	4	4
	Staffing with respect to operations	3	3	4
	Operations per ZZP	3	3	4
	Production information (planned, realized, declared)	3	3	2
	Care within & outside ZZP indication (planned, realized)	2	4	4
	Unplanned care	2	4	4
	Operations per function	3	2	4
	Staffing (per business unit)	3	2	4
	Current occupancy	3	4	1
	Occupancy prognosis	3	4	1
	Number of clients on waiting list	1	4	1
	Average duration on waiting list	1	4	1
	Short-term absenteeism personnel	2	1	2
	Long-term absenteeism personnel	2	1	2
	Frequency absenteeism personnel	2	1	2
	ZZP-mix on waiting list	1	3	1
	Average duration of ZZP	1	2	0
	Number of mutations	0	0	1
	Types of mutations	0	0	1
	Causes of mutations	0	0	1
	Bed capacity	0	1	0

## APPENDIX XXI – DATA MINING GOALS VS. INFORMATION NEEDS

1 Identify the patterns in incidents	Number of incidents occurred
	Types of incidents occurred
	Causes of the occurred incidents
	Patterns in occurred incidents
2 Identify the patterns in risk assessment	Number of clients at an increased risk
	Types of risk the clients run
3 Identify relationships between risk assessment and incidents	Incidents with respect to risk assessment
4 Identify the patterns in the number of clients	Number of clients
	Age distribution of clients
5 Predict the number of clients	Progress of clients
	Number of clients
6 Identify the patterns in care-related measurements	Progress of care-related measurements
7 Identify the relationship between care-related measures and incidents	Incidents with respect to care-related measures
8 Identify the relationship between pharmaceutical use and incidents	Pharmaceutical use by the clients
	Incidents with respect to pharmaceutical use
9 Identify the patterns in client related and non-client related time	Client-related time & non client-related time
	Care plan information
10 Identify the patterns in obtained and not-obtained treatment goals	Treatment goals (obtained & not-obtained)
	Care plan information
11 Identify the patterns in the average duration of stay	Average duration of stay
12 Predict the average duration of stay	Average duration of stay
13 Identify the patterns in coercive measures / restraints	Number of coercive measures / restraints
	Types of coercive measures / restraints
	Reason for coercive measures / restraints
14 Identify the age distribution of clients	Age distribution of clients
	Client milestones (ageing & marriage milestones)
15 Identify the patterns in ADL information	Activities of Daily Living tracking
16 Identify & predict the ZZP-mix	Number of clients per demand for care
	ZZP-mix (per business unit)

		ZZP-mix prognosis
17	Identify the relationships between operations and ZZP	Operations per ZZP
18	Identify the patterns in production	Production information (planned, realized, declared)
		Care plan information
		Operations per function
19	Identify the patterns in financed and non-financed care	Care within & outside ZZP indication (planned, realized)
20	Identify the patterns in unplanned care	Unplanned care
21	Predict the unplanned care	Unplanned care
22	Predict the occupancy rate	Current occupancy
		Occupancy prognosis
		Bed capacity
23	Identify the patterns in the waiting list information	ZZP-mix on waiting list
		Number of clients on waiting list
		Average duration on waiting list
24	Identify the average duration of a ZZP	Average duration of ZZP
25	Identify the patterns in mutations	Number of mutations
		Types of mutations
		Causes of mutations

## APPENDIX XXII – ZZP PRODUCTS

<b>ZZP</b>	<b>Sector</b>	<b>Description</b>
ZZP0	VV	Partner residence
ZZP1	VV	Sheltered residence with some support
ZZP2	VV	Sheltered residence with support and care
ZZP3	VV	Sheltered residence with support and intensive care
ZZP4	VV	Sheltered residence with intensive support and extended care
ZZP5	VV	Sheltered residence with intensive dementia care
ZZP6	VV	Sheltered residence with intensive care and nursing
ZZP7	VV	Sheltered residence with intensive care, because of specific disorders, with emphasis on support
ZZP8	VV	Sheltered residence with intensive care, due to specific disorders, with emphasis on care / nursing
ZZP9a	VV	Geriatric rehabilitation care
ZZP9b	VV	Restorative treatment and nursing care
ZZP10	VV	Protected residence with intensive palliative-terminal care

ZZP1 B	GGZ	Continued residence with support
ZZP2 B	GGZ	Continued residence with extensive support and structure
ZZP3 B	GGZ	Continued residence with intensive support
ZZP4 B	GGZ	Continued residence with intensive support and care
ZZP5 B	GGZ	Continued residence with intensive support and regulation of behaviour
ZZP6 B	GGZ	Continued residence with intensive support and intensive nursing and care
ZZP7 B	GGZ	Secure continued residence due to extreme behavioural problems with intensive support

ZZP1 C	GGZ	Sheltered residence with support
ZZP2 C	GGZ	Structured sheltered residence with extensive support
ZZP3 C	GGZ	Sheltered residence with intensive support
ZZP4 C	GGZ	Structured sheltered residence with intensive support and care
ZZP5 C	GGZ	Sheltered residence with intensive support and regulation of behaviour
ZZP6 C	GGZ	Sheltered residence with intensive support and intensive nursing and care

ZZP1 VG	GZ	Residence with some support
ZZP2 VG	GZ	Residence with support
ZZP3 VG	GZ	Residence with support and care
ZZP4 VG	GZ	Residence with support and intensive care
ZZP5 VG	GZ	Residence with intensive support and intensive care
ZZP6 VG	GZ	Residence with intensive support, care and regulation of behaviour

ZZP7 VG	GZ	(Private) residence with intensive support, care and regulation of behaviour
ZZP8 VG	GZ	Residence with support and complete care and nursing
ZZP1 LVG	GZ	Residence with some treatment and support
ZZP2 LVG	GZ	Residence with treatment and support
ZZP3 LVG	GZ	Residence with intensive treatment and support
ZZP4 LVG	GZ	Residence with very intensive treatment and support
ZZP5 LVG	GZ	Private residence with very intensive treatment and support
ZZP1 SGLVG	GZ	Treatment in an SGLVG treatment centre
ZZP1 LG	GZ	Residence with some support and some care
ZZP2 LG	GZ	Residence with support and some care
ZZP3 LG	GZ	Residence with some support and care
ZZP4 LG	GZ	Residence with support and care
ZZP5 LG	GZ	Residence with support and intensive care
ZZP6 LG	GZ	Residence with intensive support and intensive care
ZZP7 LG	GZ	Residence with intensive support and very intensive care
ZZP1 ZGA	GZ	Residence with support and some care
ZZP2 ZGA	GZ	Residence with intensive support and care
ZZP3 ZGA	GZ	Residence with intensive support and intensive care
ZZP4 ZGA	GZ	Residence with intensive support and some care
ZZP1 ZGV	GZ	Residence with some support and some care
ZZP2 ZGV	GZ	Residence with support and some care
ZZP3 ZGV	GZ	Residence with intensive support and care
ZZP4 ZGV	GZ	Residence with intensive support and intensive care
ZZP5 ZGV	GZ	Residence with very intensive support and very intensive care
ZZPX	-	Undetermined intensity of care package

Source: (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2012)