Aalto University
School of Business

High-Cost Users in Social and Health Care - Mental Health and Substance Abuse Customers

Case City of Tampere

MSc program in Information and Service Management

Master's thesis

Hanna Nyberg

2016

Department of Information and Service Economy
Aalto University
School of Business

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Nyberg Hanna	
Title of thesis HIGH-COST USERS IN SOCIAL AND HEALTH CARE – MENTAL HEALTH AND SUBSTANCE ABUSE CUSTOMERS	
Degree Master of Science in Economics and Business Administration	
Degree programme Information and Service Economy	
Thesis advisor(s) Merja Halme, Tomi Malmströn	
Year of approval 2016 **Number of pages** 88	**Language** English

Abstract

Social and health care costs accumulate to a small number of people, called high-cost users. High-cost use correlates with poor health status and complex service needs and thus causes significantly higher expenditures than the rest of the population. It is not uncommon that the costliest 10 % causes 80 % of annual social and health care expenditures.

This study analyses the high-cost users of the City of Tampere, and in more detail, mental health and substance abuse customers. The aim of this study is to gather understanding of these high-cost users, their demographics, service use and the persistence of the high-cost use.

The gathered understanding is then used to segment the population into distinct groups in order to create a framework of potential methods to manage these segments, and to achieve cost savings, better services, effectiveness, productivity, and better quality.

In addition to managing the current high-cost users of mental health and substance abuse services, this study tackles the topic of predicting future high-cost users by exploring the possibilities in predictive modeling in social and health care.

Keywords social care, health care, mental health, substance abuse, disease management, care management, cost analysis, high-cost users, service use, predictive modeling, logistic regression, cluster analysis

Table of Contents

List of Figures

List of Tables

1 Introduction

It is widely acknowledged that social and health care costs accumulate to a small number of people. The same trend is also clear in Finland, where 10 % of population account about 80 % of total social and health care expenditures (Leskelä, Komssi, Sandström, Pikkujämsä, Haverinen, Olli & Ylitalo-Katajisto, 2013). Studying social and health care expenditures is very important, especially in economic downturn as the resources are limited and every resource, for example labor, service units, money and time, should be used where it generates maximum overall welfare. In this paper, these people causing substantial share of annual expenditures are defined as high-cost users (HCUs), and this paper aims to better understand these people, what kind of services do they use, what diagnoses they have, and how persistent the HCU status is.

As the commonly agreed aim is to reduce the costs of social and health care, or at least slower the annual increase from rather stable 6 %, it is important to understand the effect of skewed distribution of the expenditures in order to better allocate resources and allocate budget in a more targeted manner.

As small population uses most of the resources and account for majority of costs, it is extremely important to understand how social and health care system should provide care and manage these HCUs. This research tackles the challenge of HCU management by suggesting potential methods to manage HCUs.

In addition, it is important to be able to identify and predict people at risk of becoming HCUs in the future, in order to offer preventive interventions and services. Accordingly, this research explores the possibilities of predictive modeling in social and health care. If prevention and light services, which are relatively cheap, are neglected the problems easily accumulate and people end up being HCUs, which is very expensive. Moreover, being able to predict future HCUs benefits especially people using mental health and substance abuse services (Leskelä et al., 2013).

Mental health (MH) and substance abuse (SA) problems have become more common and the costs from MH/SA services are increasing. These services make up a major share of expenditures especially among young adults and middle-aged people. The costs from MH/SA services are relatively highest among the age group between 25 and 59 years, where MH/SA problems also often accompany each other. Moreover, the co-occurrence of MH and SA problems with possibly other diseases creates challenges in coordinating and collaborating for institutions offering care.

Thus, this research focuses on high-cost users of mental health and substance abuse services. Even though high-cost users, or high utilizers, and mental health and substance abuse problems have received a lot of academic interest, MH/SA customers have not been studied from this point of view earlier.

1.1 Background of the Study

Social and health care systems in many developed countries are facing similar challenges including ageing population, increasing number of people living with long term conditions, rising rates of emergency hospital admissions and financial pressures (Lewis, Curry & Bardsley, 2011). In addition, as people live longer and have increasingly complex conditions, social and health care systems must adapt in a sustainable way (Panattoni, Vaithianathan, Ashton & Lewis, 2011). The skewed distribution of social and health care spending may indicate that overall social welfare is not optimized, and it could be enhanced through a reallocation of resources from HCUs to non-HCUs (Berk & Monheit, 2001). As mentioned, better preventive and light services can be good investments in the long run. This doesn't mean that HCUs care should be any worse, vise versa.

In Finland, mental health (MH) services are offered in many institutions, as these services can be offered under specialized health care, primary health care or social care. A municipality's responsibility is to offer MH services appropriate to be offered at health care center, i.e. primary health care. In 2013, there were 591 000 outpatient visits to health care centers' MH services by 81 000 different customers. Moreover, more difficult MH patients are treated at specialized psychiatric care, which is a specialty of specialized health care. After an episode at specialized psychiatric care, patients can be moved to specialized health care's outpatient care, for example psychiatric outpatient department, day center or rehabilitation home. Patients can also be moved back to health center's responsibility, to mental health clinic, to occupational health care or to private provider. (Kuntaliitto, 2015)

In Finland, substance abuse (SA) services are both social and health care system's responsibilities and the law states that a municipality has to offer basic SA services according to its SA patients' service needs.

SA services can be divided into three categories (Kuntaliitto, 2015)

1. Outpatient care (e.g. A-clinics, youth centers and day centers)
2. Intermediate services (e.g. first homes, nursing homes and supported housing)
3. Institutional care (e.g. detoxification)

MIELI - National Plan for Mental Health and Substance Abuse Work is a program by The National Institute for Health and Welfare (THL) and it defined the core principles and priorities for the future of mental health and substance abuse work until 2015, and outlined common national objectives, as these problems have a great significance for the public health. The plan emphasizes reinforcing client's status, promotion, prevention and focuses on basic and outpatient services. (The National Institute for Health and Welfare, 2015)

In the recent years, the trend has been to reduce heavy and expensive inpatient and institutional care and substitute it to outpatient services. The trend can be seen also in MH services, as in many countries, including Finland, policies have supported the development of community care and outpatient care to avoid costly hospitalization of people with severe MH problems (Calver, Brameld, Preen, Alexia, Boldy, & McCaul, 2006). Furthermore, community-focused, integrated care is seen as an important way to achieve the pressure to slow the level of public spending (Elissen, Struijs, Baanc & Ruwaardaa, 2015).

1.2 Research Objectives and Questions

There is no clear understanding of MH/SA HCUs' service use, demographics and indicators of becoming a HCU in the future. This research aims to tackle these problems to achieve better understanding, and thus management, of HCUs, that ultimately can lead to better health and cost savings.

First, the initial goal is to understand the distribution of MH/SA costs, i.e. how large share of costs the costliest 10 % of the population using MH/SA services causes.

Second, after understanding the basic distribution of MH/SA costs, this study aims to identify the characteristics of MH/SA HCUs in order to understand the factors that contribute to high costs. This includes the analysis of demographic variables, for example gender, age and diagnoses, together with the services used, including for example MH/SA services and other social and health care services.

Third, in addition to creating understanding of these costly customers, another aim is to form segments from the population, to enable better management. Moreover, this research aims to understand how MH/SA HCUs should be managed by segments, including better control, coordination and integration, by exploring social and health care operations management literature.

Finally, this study aims to find indicators for predicting future MH/SA HCUs among the patients by exploring predictive modeling literature. Due to the continuing growth of social and health care spending, the need to be able to predict future HCUs has been acknowledged in many studies (see e.g. Leskelä et al., 2013; Panattoni et al., 2011; Shenas, Raahemi, Tekieh & Kuziemsky, 2014).

Research questions are as follows

1. How large share of MH/SA costs the costliest 10 % of the population using MH/SA services causes?
2. How can the service use and costs of MH/SA HCUs be described by different demographic variables?
3. What kind of segments can be formed?
4. How can MH/SA HCUs be managed by segment?
5. What kind of indicators can be found for predicting future MH/SA HCUs?

1.3 Research Environment

This research is conducted for The Institute of Healthcare Engineering, Management and Architecture (the HEMA Institute) at the Department of Industrial Engineering and Management (DIEM), at Aalto University School of Science. HEMA is a research group that concentrates on the production of health services and their development.

In addition, this study is part of JYVÄ project (Public Private Co-Operation – Effective Models in Social- and Healthcare Service Value Networks), which aims to build understanding about service innovations, and how these innovations affect productivity, effectiveness and business models. JYVÄ project also aims to improve how people use social and health care services and enhance co-operation of public and private partnerships. JYVÄ project began in 2014 and is funded by Tekes, the Finnish Funding Agency for Innovation.

JYVÄ project aims to tackle the challenge of the lack of co-operation, especially between public and private actors, in social and health care services. Either actors' value networks don't coincide, or they overlap serving same customers. Business, production and financial models are needed to be promoted in order to increase the productivity, quality and effectiveness in these customer-oriented value networks.

JYVÄ project consortium includes three research centers

1. Aalto University's HEMA Institute
2. Oulu University School of Economics' micro-entrepreneurship research group
3. Oulu University of Applied Science's School of Health and Social Care

JYVÄ project has also service provider partners including

1. City of Tampere
2. City of Espoo
3. Joint municipalities of Kallio
4. Laastari Lähiklinikka
5. Megaklinikka
6. Doctagon
7. Omasairaala

City of Tampere (from now on as Tampere), which is the case provider for this study, has several objectives for the project

1. Identification and segmentation of HCUs
2. Cost analysis of HCUs
3. Development of new service models for HCUs' needs

1.4 Structure of the Thesis

The structure of this paper is following. In the second section, literature review presents current literature on the topics of HCUs in social and health care, predictive modeling and identifying future HCUs, and managerial models for managing various customer groups, with the focus being on MH/SA services and customers. In the third section, research methods, including two-step cluster analysis and logistic regression, are shortly discussed. Fourth section is reserved for the discussion about the case study and data. Fifth section is reserved for results and analysis. Sixth section discusses managerial models and suggestions to provide care for each segment. The last part is reserved for discussion and conclusions, including strengths and limitations, and suggestions for future research.

2 Literature Review

Literature review is divided into three parts. First, earlier research on HCUs in social and health care is presented, including also different definitions of high-cost use and characteristics of MH/SA problems. Second, identifying future HCUs is discussed with the focus being on predictive risk modeling. Risk segmentation, data requirements and two other identification methods, threshold modeling and clinical assessment, are also shortly discussed in that section. Third, selected management tools and frameworks to better organize the treatment and services of MH/SA HCUs are presented. The structure of the literature review is sequential and aims to achieve deeper understanding of the topic gradually, as demonstrated in Figure 1.

Figure 1. Structure of the literature review

2.1 High-Cost Users in Social and Health Care

HUCs in social health care have received a lot of academic attention for decades. In this section, different definitions of high-cost use are first discussed, followed by exploring earlier research on the topics of HCUs and characteristics of MH/SA problems.

2.1.1 Definition of High-Cost Users

HCUs (or high utilizers, HUs) can be defined in various ways based on the purpose of the study, and there is no commonly agreed way to define it. However, the most common way to define the intensive use of social and health care services seems to be the cost. Service use, or utilization, is another way to define the population of using majority of services, but it seems to be a less used method. However, there is a relationship between the cost and utilization, as high utilization correlates strongly with high expenditures, but not necessarily if the customer uses a lot of cheap services instead of costly services. Moreover, a third way to define the population is to use a combination of utilization and expenditures. For the simplicity and purpose of the empirical part of the study, the term HCU is used over HU. In this study the heavy use of social and health care services is defined based on costs, not based on number of visits.

Furthermore, there is no single criterion for defining the lower limit for high-cost use. The most common way seems to be to use a percentage value between 1 % and 15 %, which means that the customers are ordered based on their annual costs and then high-cost use is defined based on the costliest 1-15 % of the customers and their costs. Additionally, some studies have simply used absolute amount, for example costs over 75 000 € per year, to define the limit for high-cost use. In addition, some studies have used two criteria, for example costliest 5 % with at least one hospitalization per year. As it can be seen, it is common that the costliest 10 % have caused around 75-80 % of the total costs and the costliest 5 % about 40-60 %, depending on the services included. See Table 1 for examples of selected studies.

Table 1: Different definitions of HCUs with the percentage of costs caused

Location	Criterion	Cost / service use	% of Costs Caused	Services Included	Authors
Uusimaa, Finland	Top 15 %	Service use	70 %	Specialized health care	Leskelä et al. (2015)
Oulu, Finland	Top 10 %	Cost	81 %	Social and health care	Leskelä et al. (2013)
Ontario, Canada	Top 5 %	Cost	61 %	Hospital and home care	Rais et al. (2013)
The United States	Top 5 %	Cost	49 %	All health care	Ehrlich et al. (2010)
Capital Region, Finland	>75 000€ (0,1 %)	Cost	4 %	Health care and elderly care	Kapiainen et al. (2010)
Manitoba, Canada	Top 1 %	Cost	35 %	All hospital and physician care	Deber & Lam (2009)
Manitoba, Canada	Top 5 %	Cost	41 %	Prescription expenditures	Kozyrskyj et al. (2005)
Western Australia	Top 5 % (at least 1 hospitalization)	Cost	38 %	Inpatient care	Calver et al. (2006)
The United States	Top 10 %	Cost	75 %	All medical care	Garfinkel et al. (1988)

2.1.2 High-Cost Users in Earlier Research

In this section, the results and findings from selected literature is presented. There are many studies covering HCUs in health care, while social care seems to be a less studied field. Furthermore, as the focus of this study is on MH/SA customers, if MH/SA services or customers are included in the specific study, the focus is on the results and findings of them.

One of the most common finding concerning HCUs is that high-cost use strongly correlates with poor health status, and the spending distribution is remarkably stable over time (see e.g. Garfinkel, Riley & Iannacchione, 1988; Berk & Monheit, 2001). This means, that HCUs are often really sick and this status continues for a long time, i.e. several years.

Rais, Nazerian, Ardal, Chechulin, Bains & Malikov (2013) studied hospital and home care services in Ontario Canada and discovered that 5 % of the population caused 61 % of all costs, and their average cost per customer was 12 times compared to other customers. In their study, the elderly (over 65 years-old) accounted for 60 % of HCUs and 56 % of HCUs' costs, while the number of customers increased with increasing age, and the average cost per patient seemed to decrease with an increasing age. There were no significant differences among females and males, even though males had slightly higher average cost per patient and bigger share of HCUs' costs. Rais et al. (2013) included MH as a care type in their research and the results show that MH patients' average cost per customer was the highest among seven different care types, and 89 % of the total MH costs were caused by HCUs. Another research from Canada found similar results, as 1 % of the population accounted for 35 % of all hospital and physician costs (Deber & Lam, 2009).

A study from the United States found out that 5 % of the population accounted for 49 % of total health care spending, and MH problems accounted the largest share of chronic conditions' spending (Ehrlich, Kofke-Egger & Udow-Phillips, 2010). An earlier study from the United States (Garfinkel et al., 1988) found similar distribution of medical expenditures, as the costliest 10 % was responsible for 75 % of all incurred costs. The study also contributed interesting findings about the characteristics of HCUs. For example, the probability of being a HCU increased if the person was unemployed, employed only part time, married or lived in central city communities.

Calver et al. (2006) studied HCUs of inpatient care in Australia and discovered that HCUs accounted for 38 % of inpatient costs and care days. In addition, people with mental and behavioral disorders comprised 14 % of the HCUs.

Lately, HCUs have also received academic attention in Finland. Leskelä, Silander, Komssi, Koukkula, Soppela & Lehtonen (2015) studied patients using specialized care services in Uusimaa Hospital District (HUS) and discovered that 15 % of patients caused 70 % of the total costs. In addition, they discovered that 41 % of these costs were accounted by patients using more than one specialty, and when a patient was among the costliest 15 % for two years in a row, the share of these costs increased to 53 %. Specialized psychiatric care was included in the study, and they discovered that typically patients using many specialties had similar psychiatric diagnoses (e.g. bipolar disorder, depressive episode, schizophrenia, and psychosis), and these diagnoses were often together especially with heavy service use from surgery and internal medicine specialties. In addition, Leskelä et al. (2015) found out that specialized psychiatric care patients were 25 times more likely to be in the top 15 % for two years in a row, and psychiatric diagnoses clearly increased the risk of using more services.

Another study conducted in Finland by Leskelä et al. (2013) found out that the most expensive 10 % caused 81 % of social and health care costs in the city of Oulu. They discovered that these HCUs were the main users of specialized psychiatric care and also other social services. Leskelä et al. (2013) contributed more to understanding MH/SA patients, as these customers were included as two of the main segments. SA service (including housing) users' and specialized psychiatric care patients' average annual costs were 18 222 € and 13 600 € respectively. SA service users' annual costs were mainly from elderly care and social care housing (47 %), mental health and psychiatric care (30 %), and specialized somatic health care (14 %), while psychiatric care patients' annual costs came from mental health and psychiatric care (63 %) and specialized somatic health care (26 %).

Kapiainen, Seppälä, Häkkinen, Lauharanta, Roine & Korppi-Tommola (2010) studied HCUs in Finland's Capital region and used different kind of definition for HCUs – an absolute value. Extreme HCU's definition was over 75 000 € spending in a year and HCU's definition was over 50 000 € spending in a year. They discovered that 40 % of extremely HCUs' total costs were caused by psychiatric care ward, and 44 % of extremely HCUs had mental disease and/or depression. In addition, majority of these extremely HCUs with a mental disease were males (~55-65 %, depending on the city) and their prevalence was highest in the age group between 19 and 64 years old. The differences between Helsinki, Espoo and Vantaa, indicate that remarkable cost savings can be achieved by reducing mental patients' need for long term psychiatric ward early during the episode (Kapiainen et al., 2010).

The phenomenon of skewed cost distribution is also present in pharmaceuticals, as top 5 % contributed 41 % of total costs. (Kozyrskyj, Lix, Dahl & Soodeen, 2005). It is also noteworthy that HCUs were more likely to have mental health problems, as 25 % of the HCUs had depression and 9 % had a schizophrenia. Non-HCUs' figures were 13 % and 1,5 % respectively.

2.1.3 Characteristics of Mental Health and Substance Abuse Problems

"Behavioral health problems, such as depression, anxiety, alcohol and substance abuse, are among the most common and disabling health conditions worldwide" (Unützer, Harbin, Schoenbaum & Druss, 2013). As we know from the previous section, MH/SA customers form a major share of all HCUs and their costs seem to be rather significant. MH problems seem to be most common among 19-64 year-old males. In addition, MH/SA HCUs seem to use heavily also other services, such as surgery and internal medicine specialties. MH/SA HCUs are also likely to stay in the HCU group at least two years in a row. In this section the understanding of MH/SA customers and problems is deepened.

MH/SA problems and illnesses occur often together. It is estimated that 43 % of people aged 15-54 with a SA disorder also have a MH disorder, and 15 % of people with a MH disorder also have a SA disorder (Kessler, Nelson, McGonagle, Edlund, Frank & Leaf, 1996; Kessler, 2004).

MH/SA problems are also often accompanied with chronic general illnesses, such as diabetes, heart disease, neurologic illnesses, and cancer (Katon, 2003). MH problems can also cause somatic symptoms, for example headache, fatigue, dizziness, and pain, which in turn may lead to increased outpatient medical visits (Koenke, 2003). Moreover, chronic heavy alcohol use is associated with liver disease, immune system disorders, cardiovascular diseases, and diabetes, while substance abuse, especially injection drug use, is associated with hepatitis C, HIV, and hepatitis B (Institute of Medicine US, 2006).

Common MH disorders are also often associated with less privileged social position, for example low education, unemployment, low income or material circumstances, and low social status (Fryers, Jenkins & Melzer, 2004).

2.2 Identifying Future High-Cost Users

Previous section discussed characteristics of high-cost use and MH/SA problems. This section is reserved for the topic of identifying future HCUs. This topic is very important, because ideally, high risk patients should be identified before the occurrence of increased social and health care costs.

Identifying the current HCUs of a given year is the basis for understanding the situation, distribution of the costs, and characteristics of these customers. However, the problem often is in identifying people who will accrue majority of the costs in the future years (Cousins, Shickle & Bander, 2002). The problem arises due to the fact that a specific year’s HCUs might not be HCUs in the following year, because “individuals move in and out of the HCU group” (Ash, Zhao, Ellis & Kramer, 2001).

“Any effort to control costs must focus on those who are receiving large amounts of care” (Berk & Monheit, 2001). It is obvious, that it is more beneficial to go after low-hanging fruits by focusing on 80 % of the expenditures caused by 10 % of the population, than focusing on only 20 % of the expenditures caused by 80 % of the population. The success of any managerial model, especially for HCUs, lies in identifying the right people for these programs, in addition to being able to segment people into homogenous groups based on some important factors. These programs must target high-risk patient populations, with a specific condition, where interventions can theoretically lead to reduced resource utilization (Charlson, M., Charlson, R.E., Briggs, W. & Hollenberg, J., 2007). Thus, in the bottom line of being able to offer preventive interventions and services to the right people, we must be able to find the high-risk users accurately, who are in danger of becoming HCUs in the future. “Enrolling people who are not in reality very high risk in intensive interventions, such as community matrons or virtual wards, is not cost-effective” (Lewis et al., 2011).

The following sections present predictive risk modeling method in general, followed by risk segmentation and data requirements. In addition to predictive risk modeling, two other methods to predict future high-cost use, threshold modeling and clinical assessment, are also shortly presented.

2.2.1 Predictive Risk Modeling

Predictive risk modeling is a method used to identify potential future HCUs, who could be enrolled in management program or offered preventive interventive services. Predictive risk models are case-finding tools enabling the identification of patients at risk of becoming HCUs. Predictive risk models are already in use in many countries, for example PARR (Patients-at-Risk-of-Rehospitalisation) tool and Combined Predictive Model used by the National Health Service (NHS) in England. (Cousins et al., 2002; Panattoni, et al., 2011)

It has been demonstrated that predictive risk modeling using routine health and social care data can be used to predict individuals who will start using intensive social care in the coming 12 months. However, it is important to acknowledge that the fact that a specific individual becomes a HCU depends not only on their social and health care needs, but also on decisions made by the professionals, availability of services, and local, regional and national policy decisions. (Bardsley, Billings, Dixon, Georghiou, Lewis & Steventon, 2011).

Predictive risk models vary in four principal ways (Panattoni et al., 2011)

1. Predicted event
2. Set of predictor variables
3. Time period
4. Statistical technique

It is important to acknowledge that all of the four dimensions, event, variables, time period, and statistical technique, affect each other's fit and the prediction power of the model. (Panattoni et al., 2011).

Predicted Event

Selecting a target variable, i.e. predicted event, is the first step in the process of predictive modeling (Cousins et al., 2002). Predictive risk modeling can be used to predict any event (e.g. becoming a MH/SA HCU) if it meets the four criteria set by Lewis et al. (2011). First, the event is undesirable to the patient, and by offering preventive services the quality of life and health status can be improved. Second, the event is significant (i.e. costly) to the health or social service, because in order for the preventive services to break even it must generate future cost savings. Third, the event itself is preventable. Fourth, predictive risk model can be built on routine administrative data, by analyzing historic data for correlations between the outcome of interest (e.g. becoming a MH/SA HCU) and a range of potential explanatory variables, for example age, deprivation, patterns of health service use, and a range of different diagnoses. Cousins et al. (2002) give examples of predicted events, including adverse medical events or conditions, expensive or risky medical procedures, hospital readmissions, and high health care costs, this study focusing on the latter.

Time Period

The second step in building a predictive model is selecting "when" for the predicted event, meaning that deciding whether the goal is to predict the occurrence of the predicted event in six months, one year, or many years into the future (Cousins et al., 2002). Time horizon of a predictive risk model has a significant impact on the prediction accuracy and it is widely agreed upon that a time horizon of greater than one year decreases the prediction power significantly (Panattoni et al., 2011).

Predictor variables

The third step in building a predictive model is to decide on predictor variables, so called candidate drivers (Cousins et al., 2002). The most commonly used predictive variables are diagnoses and prior utilization combined with demographic data, and as discussed earlier, there is a wide range of possible variables available depending on the scope of the data, and the selection of the variables affect largely the accuracy of predictive risk model (Curry, Billings, Darin, Dixon, Williams & Wennberg, 2005).

Panattoni et al. (2011) divide possible predictor variables into six groups

1. Socio-demographic data
2. Diagnostic data
3. Prior utilization or costs data
4. Pharmacy data
5. Health status and functionality data
6. Clinical data

Most of the earlier research agrees that demographic variables alone do not yield a high prediction power, but when combined with diagnostic and prior utilization or costs the accuracy increases significantly (Panattoni et al., 2011). Ash et al. (2001) state that combined cost and diagnostic data are most powerful when used together. To conclude, socio-demographic data, diagnostic data and prior cost data seem to be sufficient combination of predictor variables when predicting future high-cost use.

Moreover, Shenas et al. (2014) suggest that identifying non-trivial and proactive factors enable better proactive identification of future HCUs, whereas trivial attributes identify costs after the fact. They identified five non-trivial predictor attributes to predict future HCUs

1. Individual's overall health perception
2. Age
3. History of blood cholesterol check
4. History of physical, sensory or mental limitations
5. History of colonic prevention measures

Moreover, Elissen et al. (2015) suggest additional determinants to be used at community level, including people’s lifestyle and social network, for example income level, income source, household income, geographic disparities in terms of population and environment, available care services, type of household, size of household, housing circumstances, and degree of loneliness.

Statistical Technique

There is a lot of literature on the topic of predictive risk modeling in social and health care, yet there is no single consensus on the best statistical technique. Regression techniques can be divided into linear and logistic techniques, while linear regression is more suitable when the outcome variable is continuous, whereas logistic regression is used when the outcome is binary. Time series models are useful in mapping changes over time, neural networks are useful when interactions between predictors are not explicitly identified, and classification trees are useful when data is incomplete and contains a lot of missing values (Cousins et al., 2002). However, regression models seem to be the most commonly preferred technique, while other methods, for example artificial intelligence and data mining techniques, for example neural networks, decision trees, and Bayesian nets, are also emerging and becoming more popular. (Curry et al., 2005; Panattoni et al., 2011; Shenas et al., 2014)

The final outcome of a predictive risk model is a ranking of customers according to their risk for future high-cost use. Then, the decision maker has to decide the cut point, which means how many of the riskiest customers are enrolled to preventive services. There is a lot of literature on the topic of deciding the ideal cut point, and these methods include for example an evidence-based “optimal” cut point, an “arbitrary” threshold, condition-specific cut point, and uniform screening method. Then, this rank-ordered list is used to decide on the level and type of intervention these persons should receive. At this point, it is important to emphasize that this risk-based ranking is not the same as the ranking based on some year’s costs and limit for being a HCU (e.g. the costliest 10 %), which were discussed earlier. These high-risk patients haven’t become HCUs yet, and the goal is to prevent it from happening. (Cousins et al., 2002; Murphy, Castro & Sylvia, 2011)

2.2.2 Risk Segmentation

As mentioned earlier, targeting the right people for preventive intervention services and other management programs is important, because these people with high predicted risk present the greatest opportunity for making cost savings. In addition, there is a relatively small amount of HCUs, thus the predicting and identification has to be done accurately. If interventive services are offered to a large population with lower risk, it is likely that some of the people are targeted correctly and intervened early, but the lower the risk threshold (i.e. the cut point), the cheaper or more effective the intervention has to be in order to achieve cost savings. In addition, some preventive interventions and services are typically very expensive and thus make accurate targeting even more important. (Bardsley et al., 2011; Lewis et al., 2011)

Kaiser Permanente Pyramid of Care (Figure 2) divides population into four groups according to their relative risk level. The pyramid shows that very high and high relative risk groups are only 5 % of any given population, but as we know from earlier, social and health care costs accumulate to a very small share of population. Thus, Kaiser Pyramid of Care seems to be a good tool to use as a basis of risk segmentation. (Goodwin, 2006; Lewis et al., 2011)

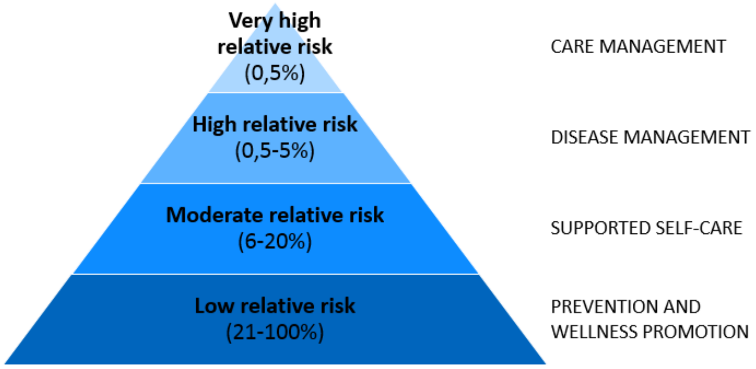


Figure 2. Kaiser Permanente Pyramid of Care

In addition to identifying high risk patients, predictive risk modeling can also guide in deciding on the appropriate level of care and ideal combination of resources (Cousins et al., 2002). Pyramid of Care suggests care management, disease management, supported self-care, and prevention and wellness promotion as different methods to manage people at different risk levels. Different ways to manage and organize care are discussed in more detail in section 2.3.

2.2.3 Data Requirements

Social and health care has been referred as "data rich" but "knowledge poor" (Abidi, 1999), where large volumes of data is constantly created but little is done to really take an advantage of it to create real knowledge. Predictive risk models are based on data, thus data quality is very important factor contributing to the prediction accuracy (Cousins et al., 2002; Curry et al., 2005). Additionally, data availability may also form limitations and barriers to designing predictive risk models (Elissen et al., 2015). In Finland, social and health care information systems don't currently support managing HCUs, including identification and prediction (Leskelä et al., 2013). The main obstacles include siloed data, which means that the right data exists but it is not available where needed, and simply non-existing data due to people's privacy or unwillingness to share information for example about marital status or annual income.

Prediction models typically use information from medical records or claims databases (Fleishman & Cohen, 2010). Kansagara, Englander, Salanitro, Kagen, Theobald, Freeman & Kripalani (2011) divide possible datasets into three groups

1. Retrospective administrative data
2. Real-time administrative data
3. Primary data collection

It is important that predictive risk models are built on routine data elements so that it doesn't increase burden on doctors and nurses (Shenas et al., 2014). Additionally, Panattoni et al. (2011) point out that implementing a predictive risk model may only require minor adjustments to the existing data infrastructure to possibly offer fast return on investment.

2.2.4 Alternative Methods: Threshold Modeling and Clinical Assessment

In addition to predictive risk modeling, there are alternative methods for identifying future HCUs and two of them are shortly discussed here – threshold modeling and clinical assessment. However, more accurate predictive risk modeling is preferable to clinical assessment and threshold modeling and thus it receives more attention in this study.

Clinical assessment simply means that a doctor or nurse makes a prediction based on her or his knowledge, training, and judgement, whether a certain patient is likely to become a HCU. There are three disadvantages making clinical assessment a worse option compared to predictive risk modeling. First, in contrast to predictive risk modeling, clinical assessment can't screen whole populations at the same time and repeatedly, and it is limited to those already in contact with the service. Second, it can't take contacts with every part of the social and health care system into account. Third, it is susceptible to different cognitive biases. Actually, clinicians' predictions are found to be statistically no different from chance. (Allaudeen, Schnipper, Orav, Wachter & Vidyarthi, 2011; Curry et al., 2005; Lewis et al., 2011; Panattoni et al., 2011)

Threshold modeling is a simple rule-based approach, meaning that if certain criteria is met, a person is identified to be in high risk. However, threshold modeling is susceptible to the effects of selection bias and regression to the mean. Selection bias means that people are selected because they are outliers and extreme, whereas regression to the mean means that if the criteria is met this year, then next year the criteria is not met simply due to random fluctuation, because after one extreme event, the next event is statistically likely to be less extreme. In addition, threshold modeling is not very accurate within general population. (Lewis et al., 2011; Panattoni et al., 2011)

2.3 Managerial Models for Mental Health and Substance Abuse Customers

Managerial models for MH/SA customers, or any other social and health care customers, aim towards improved care, care pathways, and outcomes by better control, coordination, and integration, thus by better management of processes. "Improving the quality of MH/SA care, and general health care, depends upon the effective collaboration of all MH, SA, general health care, and other human service providers in coordinating the care of their patients" (Institute of Medicine US, 2006). Lillrank (2012) defines integration as "combining several specialized and differentiated resources and contributions to create an output that is a system consisting of several parts". World Health Organization (2008) proposes six main definitions and usages for integrated health services

1. A package of preventive and curative health interventions for a particular population group
2. Multi-purpose service delivery points
3. Continuity of care over time
4. Vertical integration of different levels of service
5. Integrated policy-making and management
6. Working across sectors

Segmenting population based on the state of health and priorities have got a lot of attention in health care operations management literature. "Segmenting patient populations can lead to more creative and effective strategies for safe, efficient, effective, timely, patient centered, and equitable health care, and thus lead to a better understanding of how to achieve better health for both the individual and all people" (Lynn, Straube, Bell, Jencks & Kambic, 2007). As health and social care is too huge, complex and diversified to be treated as one industry, it is important to segment it into smaller parts which are homogeneous but large enough to be managed (Lillrank, Groop & Malmström, 2010). In social and health care, segmentation can be based on urgency, severity, demographics, clinical categories, cause types, and treatment type (Lillrank, 2012). Moreover, predictive risk modeling discussed in the previous section is a tool enabling efficient allocation of resources to make these managerial models more successful and cost-effective (Mukamel, Chou, Zimmer & Rothenberg, 1997).

MH/SA customers meet the criteria (Coons, 1996) of selecting diseases and conditions to be managed. First, high expenditures and preventability is associated with MH/SA problems, at least in the majority of the cases. Second, the outcomes of better management are measureable, for example by measuring the number of visits or hospitalizations. Third, rapid return on invest, as benefits can be captured within a relatively short time, for example in a year. Fourth, there is a large practice variation in the services for MH/SA customers, meaning that there is a wide range of approaches to treat the condition including a wide range of costs and outcomes.

In this section, a literature review is conducted on the topics of health care operation management. The aim is to understand the current relevant frameworks to be able to contribute on the better management of MH/SA customers based on the understanding achieved from the case study. The frameworks and managerial models are presented in an order based on their scope from high-level models, for example Bridges to Health model, towards more specific models, for example case management.

2.3.1 Bridges to Health Model

First framework discussed is the Bridges to Health model (Lynn et al., 2007), which divides the population into eight segments, with own definitions of optimal health and own priorities among services. The segmentation is done based on people’s needs, state of health and priorities. The framework helps in resource planning, care arrangements, and service delivery, aiming at meeting each person’s health needs effectively and efficiently. The eight groups are

1. People in good health
2. Maternal/infant situations
3. Acute illnesses
4. Stable chronic conditions
5. Serious but stable disability
6. Failing health near death
7. Advanced organ system failure
8. Long-term frailty

Lynn et al. (2007) proposes three considerations to take into account when using Bridges to Health model. First, in order for the social and health care system to be able to offer a sensible array of integrated services for each segment, available almost everywhere, the set of population segments must be limited. Second, every person must belong into one, and only one of the segments. Third, the people in each segment must have sufficiently similar service needs, rhythm of needs, and priorities, in order to make the segment useful for planning.

2.3.2 Demand and Supply-Based Operating Modes Model

Second framework discussed is the Demand and Supply-Based Operating Modes (DSO modes), which also divides population into seven homogeneous segments with different demand-supply combinations. The idea is to enable methods of mass production by segmenting individuals into large enough groups by some important aspect. It is important to notice, that the modes describe a set of integration, coordination, and control principles, and not an organization or a system. The modes and demands on integration are (Lillrank et al., 2010; Lillrank, 2012)

1. *Prevention*, aiming at preventing an event likely to happen (integrating current costs and future gains, as well as clinical and behavioral medicine)
2. *Emergency*, dealing with severe cases within an urgency situation (integrating rapidly at the triage)
3. *One visit*, dealing with non-urgent and non-severe cases (integration considering what can be done during one visit to complete the intervention)
4. *Project*, dealing with a poorly understood, extremely complex or costly, rare medical conditions, where are no existing process models to follow (integration of various specialized contributions)
5. *Elective procedures*, based on precise diagnoses and schedulable interventions (integration at the diagnostic and care planning phases)
6. *Cure process*, cases where a complete diagnosis cannot be made at the onset and the process can be planned accurately only a few steps at a time (integration of emerging understanding and adjusted care plan)
7. *Care processes*, dealing with chronic or terminal conditions (integration of several aspects into a comprehensive understanding of patient needs and the development of a continuous care scheme)

The model uses five classificatory variables to define the seven operating modes based on demand and supply

1. Urgency
2. Severity
3. Clarity
4. Continuity
5. Risk

The algorithm asks different questions and based on the answers the case ends up in one of the seven process models. The algorithm is presented in Figure 3. (Lillrank et al., 2010; Lillrank, 2012)

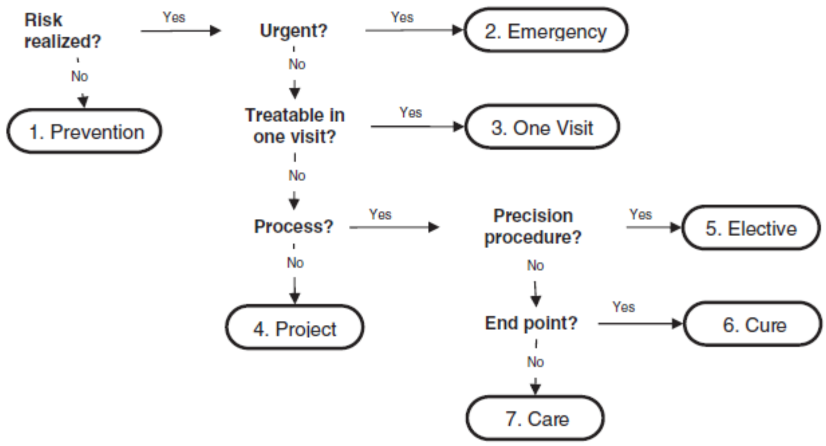


Figure 3. Demand and Supply-Based Operating Mode Flowchart

2.3.3 Preventive Interventions and Services

As discussed earlier, "it is essential to be able to identify in advance patients who are likely to have high costs in the future" (Billings & Mijanovich, 2007). In addition to identifying high-risk people, it is also critical to understand who might benefit from intervention (Cohen, Flaks-Manov, Low, Balicer & Shadmi, 2015), as people with some diseases (e.g. terminal illness) won't benefit from preventive services. Early intervention and providing the services people need improve the health and quality-of life for these patients, and potentially make net savings by reducing future social and health care costs in the same time by preventing future high-cost use (Bardsley et al., 2011; Murphy et al., 2011).

2.3.4 Disease Management

Coons (1996) states that disease management "targets populations with a specific condition and involves the implementation of coordinated, comprehensive interventions that will improve the clinical, humanistic, and economic outcomes associated with the management of that condition". Charlson et al. (2007) define disease management as "a systematic, population-based approach to patient care that aims to curb utilization by optimizing the process of care, increasing efficiency, and managing the total disease". Disease management often includes a process of selecting a group of patients with a specific, often chronic, condition and a substantial share of total costs, and then implementing programs targeting cost drivers within that group. The aim is to manage a large share of social and health care costs by allocating resources to a relatively small segment of population, i.e. HCUs (Cousins et al., 2002).

Disease management care is based on evidence based guidelines and patient pathways, and it is provided by multi-disciplinary teams with a special knowledge in the specific condition (Goodwin, 2006). Disease management program may include high level of intervention services, such as telephonic and home health visits from nurses, or low level intervention services, such as newsletter mailings with directions for self-care and a list of appropriate doctors (Cousins et al., 2002).

However, disease management is criticized for being a condition-specific program, managing each condition as a discrete problem (Cohen et al., 2015). Coons (1996) presents critical issues to be considered in regard to disease management. First, conflicting or overlapping disease management programs when a patient has multiple diseases. Second, appropriate management of some long term chronic conditions doesn't provide such dramatic impact or return on investment. Third, information systems allowing sufficient monitoring and reporting of disease management program.

2.3.5 Case Management

Whereas disease management regards one disease, case management, or care management, on the other hand, coordinates the effective management of numerous social and health conditions comprehensively for one person (Cohen et al., 2015). "Case management is intensive, individualized and involves enduring care that evaluates medical and nursing needs as they rapidly change" (Goodwin, 2006). The idea is to take a holistic view of the individual's situation and provide social and health care accordingly. According to the Pyramid of Care, the top level of the pyramid is case management and the population in this small group have multiple and complex long-term conditions (Goodwin, 2006).

From MH perspective, it is suggested that case managers have expertise in MH, but due to comorbidity it may be the best if case managers are general practitioners who can also manage patients' physical health care (Belnap, Kuebler, Upshur, Kerber, Mockrin, Kilbourne & Rollman, 2006).

2.3.6 Chronic Care Model

The Chronic Care Model is a guide to improve chronic illness management. The goal is to provide care mainly in the primary health care and use specialized health care only when necessary. (Bodenheimer, Wagner & Grumbach, 2002; Goodwin, 2006)

Chronic care model includes six components of care

1. Self-management
2. Decision support
3. Delivery system design
4. Clinical information systems
5. Health care organization
6. Community resources

2.3.7 Challenges in Finland

As HCUs tend to use more various services, the co-operation between organizations is extremely important, and the current complex service system is not a good fit for them, because their information is scattered around the system and there is no one taking the full responsibility with the big picture in mind (Leskelä et al., 2013). This is a serious problem especially for MH/SA HCUs, as their service use is scattered around the system to many different organizations, as they can be treated at specialized health care, primary health care or social care.

Additionally, the current service structure and information systems in Finland don't support HCU management, even though the skewed distribution of social and health care expenditures has been known for a long time (Leskelä et al., 2013). To manage coordination of services and customer records, Leskelä et al. (2013) suggest a Service Coordinator Model, which means that HCUs have an own coordinator who knows them by person and has the control over budget and their care plan. Callahan, Shepard, Beinecke, Larson & Cavanaugh

(1995) suggest that Service Coordinator Model decreases the total expenditures in MH/SA rehabilitation.

3 Methods

In this section, the methods used in the empirical part are shortly presented. First, two-step cluster analysis is discussed. Second, a predictive modeling method, logistic regression, is presented.

3.1 Two-Step Cluster Analysis

"Cluster analysis is the art of finding groups in data" (Kaufman & Rousseeuw, 2009). It aims at creating segments, i.e. clusters, where the objects in the same group are as similar and homogeneous as possible, and objects in different groups are as dissimilar as possible, based on a set of variables. If cluster analysis is successful groups obtained by cluster analysis can be described as meaningful, useful, intellectually satisfying and profitable (Chaudhary & Sharma, 2013).

Traditionally, clustering procedures are divided into hierarchical and non-hierarchical clustering. Two-step clustering is a newer method integrating hierarchical and partitioning clustering algorithms (Shih, Jheng & Lai, 2010). In this study, the focus is on two-step cluster analysis, which was invented by Banfield and Raftery (1993). One of the benefits compared to classical methods, i.e. hierarchical and non-hierarchical clustering, is that two-step clustering can be used when the variable set includes both continuous as well as categorical variables (Verma, 2013). Another benefit is that two-step clustering works well with larger datasets and requires a shorter processing time than other methods (Bacher, Wenzig & Vogler, 2004).

All clustering procedures require a choice of similarity measure, which is a measure assessing how close or similar the observations are. The log-likelihood distance measure, or natural logarithm of the likelihood function in other words, can handle both continuous and categorical variables, exactly like two-step clustering, making them a good pair.

Cluster quality can be measured in many ways. Silhouette coefficient is a popular measure and it measures both cohesion, i.e. the within cluster sum of squares, and separation, i.e. the between cluster sum of squares, of the clustering solution. It calculates the average distance of an element to all other elements in its cluster, and the average distance to all elements in each of the other clusters. Then, "the silhouette measure is the difference between the smallest average between cluster distance and the average within cluster distance, divided by the larger of the two distances" (Norušis, 2011). The value of the silhouette coefficient is always between -1 and +1, where higher value means better clustering solution in terms of cohesion and separation.

3.2 Predictive Modeling with Logistic Regression

Logistic regression, or logit regression, or logit model is a regression model where the dependent variable is binary. It was created by David Cox (1958) and it has been routinely available in statistical packages from the early 1980s. Binary logistic regression is used when the outcome for a dependent variable can have only two possible types, and over the last decade the logistic regression model has become a standard method in this situation. Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. (Hosmer & Lemeshow, 2000; Peng, Lee & Ingersoll, 2002)

When the response variable is binary, the shape of the response function is

$$y = \frac{e^{a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_k x_k}}{1 + e^{a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_k x_k}} \tag{1}$$

y is the response variable, that you are predicting (e.g. an order, a cancellation…)

$\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$, …, $\mathbf{x_k}$ are the multiple predictor variables (e.g. age, income…)

a is a constant numerical value

$\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{b_3}$, …, $\mathbf{b_k}$ are the numerical coefficients (weights) associated with each of the predictor variables

e represents the numerical value 2,71828 (Neper value)

Note that y never becomes strictly zero or one.

Though logistic regression belongs to general linear models we note that the response function is nonlinear. Logit is very commonly estimated by using maximum likelihood. Compared with linear regression the assumptions it requires are modest: e.g. the multivariate normal distribution is not required for the explanatory variables. The binomial distribution is the assumed distribution for the conditional mean of the outcome. The assumption can be tested by the normal z test or may be taken to be robust as long as the sample is random (observations are independent from each other). (Siegel & Castellan, 1988; Peng et al., 2002)

Summary

At this point, this study has explored all the relevant topics in order to continue to the case study. First part of the literature review covered what is known about the HCUs in social and health care, as well as what are the characteristics of mental health and substance abuse problems. Second part covered the topic of identifying future HCUs, including for example predictive risk modeling, risk segmentation and data requirements. Third part covered the topic of selected managerial models for managing HCUs, including for example Bridges to Health Model and Case Management. Previous section presented the methods used in the empirical part.

4 Data and Case Study

The goal of the empirical part of this study is to investigate mental health and substance abuse customers (MH/SA HCUs and also non-HCUs) in Tampere – what kind of services do they use and how can they be described. The goal is also to form segments of MH/SA HCUs, as well as recognize possible indicators for predicting future high-cost use. Of course, the cost distribution is also analyzed. This section first discusses case study settings, followed by presentations of data source, data description, data quality and processing methods, and service categorization.

4.1 Study Settings

This study is done for Tampere, located in Pirkanmaa Region, which is responsible by the law for organizing social and health care services for its residents. Tampere is participating in this study because they want to understand better their HCUs, especially MH/SA customers. Tampere has several ongoing projects focusing on leading with knowledge and enhancing operations. Goals of these projects include better services, effectiveness, productivity and better quality. This study supports these goals by gathering understanding of Tampere's MH/SA customers.

The empirical part consists of quantitative study of social and health care usage data in Tampere from years 2013 and 2014. The data also includes service usage in the health care district (PSHP). The following services are included in the data

- Primary health care
- Specialized health care
- Home care
- Substance abuse services
- Mental health services

The study population consists of the population of Tampere. By population size Tampere is Finland's third largest municipality (city) after Helsinki and Espoo, and it is the most populous inland city in any of the Nordic countries. In 2013 and 2014, the population of Tampere at the end of the year was 220 446 (avg. 218 934) and 223 005 (avg. 221 726) respectively. The number of 25-59-year-olds, the age group this study focuses on, was 106 328 and 108 723. See Table 2 for detailed population information. (THL, 2015)

Table 2: At the end of year population of Tampere (THL, 2015)

	2011	2012	2013	2014	E2020	E2030	E2040
Under 25-year-olds	62 270 29 %	62 730 29 %	63 286 29 %	63 682 29 %	62 412 27 %	65 761 27 %	66 503 26 %
25 - 59-year-olds	102 902 48 %	103 600 48 %	105 097 48 %	106 328 48 %	108 723 47 %	110 163 46 %	114 930 46 %
Over 60-year-olds	49 996 23 %	51 091 23 %	52 063 24 %	52 995 24 %	58 474 25 %	66 054 27 %	70 254 28 %
Total	**215 168**	**217 421**	**220 446**	**223 005**	**229 609**	**241 978**	**251 687**
Annual growth	0,9 %	1,0 %	1,4 %	1,2 %	1,1 %	0,5 %	0,4 %
Average population	214 193	216 295	218 934	221 726	N/A	N/A	N/A

4.2 Data Source and Description

The data was collected and submitted by Tampere as part of their ASKOP project. The longitudinal register data consists of publicly funded social and health care visits in 2013 and 2014. The data covers all residents of Tampere and all their visits to services that are both offered and purchased for them at Pirkanmaa Hospital District level (PSHP). Service use in private and third sector are not included in the data. Moreover, the data does not cover prescription drugs usage, information about moving away from Tampere, moving to Tampere, and possible deaths.

As mentioned, the visit data is in two datasets, years 2013 and 2014. Both of the datasets include visit data from social and health care services. The unique key in both datasets is a visit ID, which is a random number. Social security numbers are replaced by random customer IDs in order to secure individuals anonymity. However, this enables individual level research over these two years, as the customer IDs remain the same in the both datasets.

In addition to customer IDs and visit IDs, the visit data consists of some information for each visit, including service category (e.g. a medical specialty), detailed service (e.g. type of visit, for example a group visit), timestamp (including date and time), unit, diagnoses 1 and 2, date of birth, language, postal code, gender and cost. Some of the costs were total costs and in these cases the cost was divided by the number of visits or care days. It is noteworthy that several days' inpatient care period has one visit for each day. For example, 2 weeks' inpatient care consists of 14 visits.

4.3 Data Quality and Processing Methods

The data was first cleaned before conducting any analysis. The overall data quality was good and removing incomplete, but necessary, records didn't cause significant problems. In total, 39 830 (2013) and 30 659 (2014) records were deleted due to incomplete or missing values, which is 1,7-2,3 % of the original dataset. The missing values causing deletion of the record, in the sequence of deletion, were in 2013 and 2014, respectively

- Missing customer ID (49 and 62 records deleted)
- Missing birthdate (71 and 52 records deleted)
- Negative cost of visit (685 and 390 records deleted)
- Cost of visit equal to 0 € (15 921 and 12 620 records deleted)
- Missing postal code (3 067 and 3 842 records deleted)
- Postal code outside Tampere (10 876 and 22 864 records deleted)

After deleting poor records, the total number of visits equaled to 1 778 222 (2013) and 1 716 046 (2014). After removing incomplete records, the next step was to limit the study to concern only 25-59-year-olds. It was decided to include people who are born between 1.1.1955 and 31.12.1989, because they turn 25-59 years in 2014 and 24-58 years in 2013. After deleting records from under 25-year-olds (birth year >1989) and from over 59-year-olds (birth year < 1955) there were 336 162 (2013) and 348 731 (2014) records. Figure 4 demonstrates the study population.

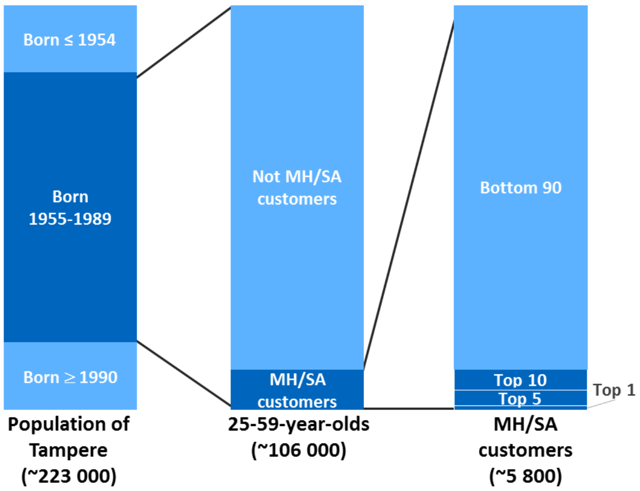


Figure 4. Description of the study population

After having a clean dataset consisting of visits, the next step was to process the data into a form where each line is a customer, i.e. transferring the visit data into customer data. This new dataset included all the relevant information on all the unique customers (n=7 973) who had used MH/SA services at least once during one of the years (2013 and/or 2014). The dataset included each customer's basic information, including customer ID, gender, birthdate, age group, postal code and city area. It also included more categorical and numerical information for both years, such as assigned segment (more on segmenting in the section 5.3), HCU status, number of service categories used, most used service category, most used detailed service, number of diagnosis groups, costs by service categories, and number of emergency room (ER) visits. Moreover, the new dataset also included binary information on what diagnosis groups each customer had received during a certain year. The data used ICD-10 diagnosis classification by World Health Organization. ICD-10 diagnosis set allows more than 14 000 codes for different diagnoses and due to the purposes of this study, it was decided to use the 26 main groups. Microsoft Access 2013 and Excel 2013 tools were used to process the data.

4.4 Service Categorization

The original data contained 100 different services or operations, which were further divided into 41 service categories (see Appendix A for detailed information on the groupings). The next step was to decide the service categories that are included in the MH/SA services. To do this, every visit was assigned a service category based on the service, not based on the service unit (expect in some cases due to additional information based on the unit). For example, if a regular visit to a psychiatry service was registered at neuropsychological outpatient department the visit was marked as a psychiatry visit, not as a neuropsychological visit. See Appendix B for the list of MH/SA units in Tampere.

In Tampere, MH services include four service categories

- *Mental health services*, including individual and group visits offered by psychologies at health care centers (primary health care)
    - Newly diagnosed mild to moderate mental health problems
    - Aims to provide structured and brief therapies
- *Geriatric psychiatry services*, offered either at the long term geriatric hospital or as an outpatient service (specialized health care)
    - Mental disorders, dementia and neuropsychiatric patients of the elderly

- *Neuropsychological services,* offered as an outpatient service (specialized health care)
    - Examinations, rehabilitation, counseling and guidance
    - Outpatient activities and hospital work
- *Psychiatric services*, including the widest range of services (specialized health care)
    - Psychiatrist visits (urgent/non-urgent/re-visit)
    - Family, home and group visits
    - Enhanced psychiatric hospital care (Päiväsairaala 1, 2 and 3)
    - Telephone reception
    - Consulting, supervision and specialist visit

SA services include one service category, *substance abuse services* (primary health care or social care depending on the service type), which include

- Detoxification
    - Alcohol intoxicated persons in the need of medical follow-up (max. 24h)
    - Arrival by police or ambulance
- Rehab
    - Alcohol and drugs
    - Inpatient and outpatient
- Replacement therapy (mainly at health care centers, also at Päiväperho)
- Outpatient activities (Huoltsu)
    - Spending time, relaxing, eating and taking care of hygiene
    - Guidance, support and advice
- Substance abuse residential care (Palhoniemi service home)
    - Arrival from hospitals, outpatient care or home
    - For people with severe substance abuse problems as a result of prolonged drug abuse

5 Results and Analysis

This section is reserved for results and analysis. First, distribution of MH/SA costs is presented. Second, demographics (gender and age), service use and cost analysis are presented. Third part is reserved for population segmentation. Fourth, MH/SA customers' diagnoses are analyzed followed by postal code area analysis and persistence of being a MH/SA HCU analysis. Seventh part is reserved for predictive models. All of the numbers and analysis in this section apply to age group born in 1955-1889, and all the HCUs discussed are MH/SA HCUs if not mentioned otherwise.

The analysis began by arranging the customers, i.e. the people who used MH/SA services at least once, in the order of annual MH/SA costs from the lowest to highest. The following percentages (cost groups) were used in analyzing the distribution of MH/SA costs

- Most expensive 1 %, 5 % and 10 % (as Top 1, Top 5 and Top 10)
    - HCUs are defined as the most expensive 10 %
- Least expensive 90 % (as Bottom 90)
    - Non-HCUs are defined as the least expensive 90 %

Before going to the analysis and results of MH/SA expenditures, the results of the distribution of total social and healthcare costs are also presented briefly, even though being outside of the main scope of study. In 2013, total costs were of 85,1 M€ and in 2014 they were 79,1 M€ (-7,2 %). The amount of customers remained stable over the two years, as in 2013 there were 47 484 unique customers (45,2 % of the total population), and in 2014 the amount was 48 142 (45,3 %, +1,4 %). This means that the amount of visits per customer also remained stable and grew from 7,1 to 7,2 (+2,3 %). However, the amount of visit contains some error as some episodes include one visit for each ward day, while some include only one visit for the first day of the episode.

The distribution of total social and health care costs seems to follow the results of the earlier studies, as the majority of costs were caused by a small number of expensive customers. In the other words, the distribution of total social and health care costs is highly skewed. The lower limits of annual total costs for each group in 2013 and 2014 are: 13 768 € and 12 852 € (Top 1), 3 600 € and 3 266 € (Top 5), and 1 543 € and 1 389 € (Top 10). Tables 3 and 4 summarize the results.

Table 3: Distribution of all social and health care costs per percentile, 2013

Cost group	n	Total costs (M€)	Avg. cost (€)	% of costs
Top 1	1 051	31,8	30 275	37 %
Top 5	5 254	58,9	11 213	69 %
Top 10	10 510	71,6	6 811	84 %
Bottom 90*	94 587	13,6	143	16 %
Total	**105 097**	**85,1**	**810**	**100 %**

* Of which 57 613 didn't use any services.

Table 4: Distribution of all social and health care costs per percentile, 2014

Cost group	n	Total costs (M€)	Avg. cost (€)	% of costs
Top 1	1 063	28,8	27 059	36 %
Top 5	5 316	54,5	10 252	69 %
Top 10	10 633	65,8	6 190	83 %
Bottom 90*	95 693	13,2	138	17 %
Total	**106 328**	**79,1**	**744**	**100 %**

* Of which 58 186 didn't use any services.

5.1 Distribution of Mental Health and Substance Abuse Costs

In 2013, total MH/SA expenditures were 19,4 M€ and in 2014 they were 17,4 M€. During both years, the majority of the costs came from psychiatry (72 % and 68 %), and in more detail from acute psychiatry ward (specialized psychiatric care). Moreover, substance abuse services caused a large share of costs (22 % and 24 %), in which detoxification was the largest expenditure. See Figure 5 for detailed results.

The main driver causing the decrease in total MH/SA costs by 2 M€ are psychiatric costs with a decrease of 2,1 M€. The average price per visit decreased from 280 € to 227 €, while the number of visits increased by 5 %. The costs of substance abuse services, geriatric psychiatry, mental health services and neuropsychology remained rather stable.

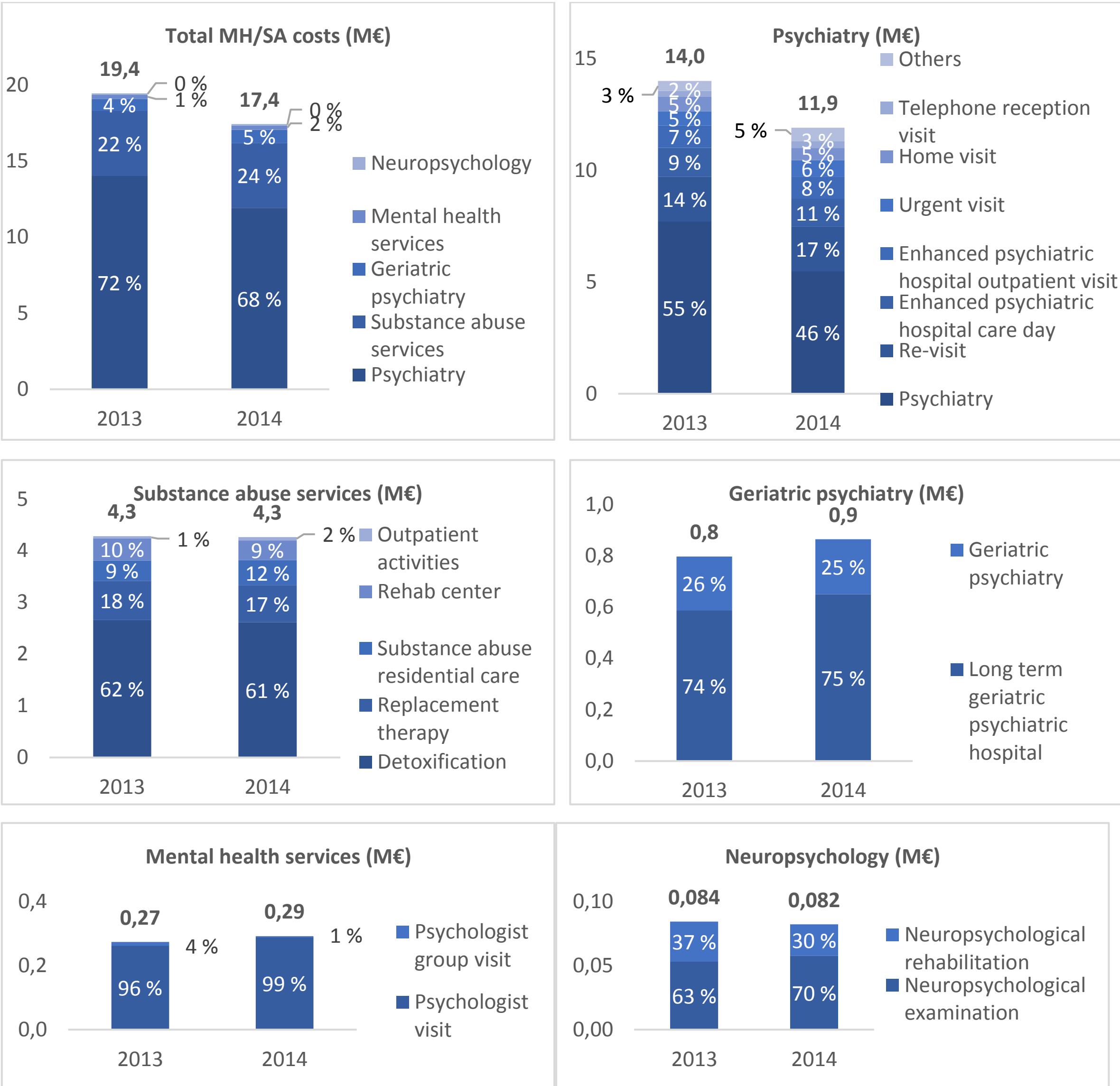


Figure 5. MH/SA costs in detail, 2013 and 2014

The distribution of MH/SA costs is also clearly skewed (see Figure 6). At first look it seems that the distribution of MH/SA costs is not as skewed as it is in total social and health care costs. However, in this population there are no people who don't use services, whereas in the distribution of total social and health care costs includes all the people who don't use any services. As this study focuses on MH/SA customers (5 771 customers in 2013 and 5 820 in 2014), the distribution is calculated based on these people. The distribution of MH/SA costs is very skewed, as the most expensive 10 % caused 62-64 % of all MH/SA costs, while Top 5 and Top 1 caused 46-48 % and 18-19 % of MH/SA costs respectively. Actually, 5,5 % of

Tampere's population caused 100 % of MH/SA costs. In 2014, average MH/SA cost was about 3 000 €, and as Figure 6 shows, high annual costs exist.

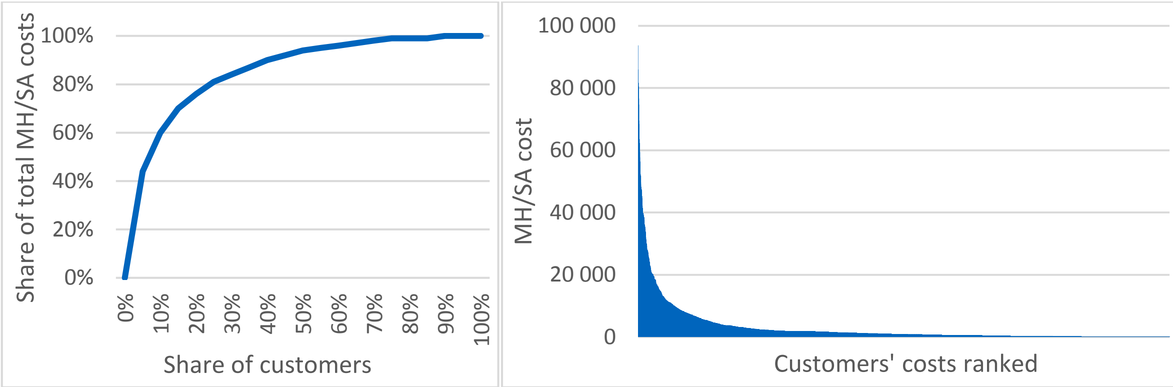


Figure 6. Cumulative MH/SA costs and individual customers' costs ranked from highest to lowest, 2014

The lower limits for MH/SA costs for each group in 2013 and 2014 are 43 516 € and 39 139 € (Top 1), 14 920 € and 12 423 € (Top 5), and 8 136 € and 7 250 € (Top 10). Moreover, other costs and total costs also seem to be higher in more expensive MH/SA HCU groups. Table 5 summarizes all MH/SA customers in 2014 (see appendix D for similar table for year 2013).

Table 5: Distribution of MH/SA customers' MH/SA and other costs, 2014

Cost group	n	Total MH/SA costs (M€)	Avg. MH/SA cost (€)	% of costs	Avg. other costs (€)	Avg. total costs (€)
Top 1	58	3,1	53 737	18 %	4 448	58 185
Top 5	291	8,0	27 436	46 %	4 003	31 440
Top 10	582	10,7	18 407	62 %	3 700	22 107
Bottom 90	5 238	6,7	1 280	38 %	1 897	3 177
Total	**5 820**	**17,4**	**2 992**	**100 %**	**2 078**	**5 070**

The next step was to analyze MH/SA service categories and detailed services in more detail. This revealed numerous differences among the service categories and detailed services. For example, in psychiatry, enhanced psychiatric hospital outpatient visit, detoxification, and geriatric psychiatry, HCUs caused the majority of costs but they represent only a small share of total customers, i.e. the distribution of a specific service's expenditures is skewed. Moreover, in inpatient care, for example in enhanced psychiatric hospital care, substance abuse residential care, and long term geriatric psychiatric hospital care, majority of customers are HCUs, and thus they also cause majority of the costs. It is also interesting, that in some detailed services, for example in psychologist treatment at health care center, neuropsychological examination, psychiatry re-visit, telephone reception and urgent visit, the share of HCUs is quite small. This means that these services are mainly used by non-HCUs. See Table 6 for detailed results (see appendix D for similar table for year 2013).

Table 6: Service categories and detailed services, share and costs of HCUs, and average and total costs, 2014

Service	Total n	Share of HCUs	Costs of HCUs	Avg. cost per customer (€)	Avg. cost per HCU (€)	Total cost (€)
Psychiatry	**4 680**	**10 %**	**61 %**	**2 546**	**15 483**	**11 913 618**
Psychiatry	909	36 %	84 %	6 026	14 019	5 477 725
Re-visit	2 844	9 %	12 %	702	856	1 995 916
Enhanced psychiatric hosp. care day	165	84 %	95 %	7 604	8 637	1 254 695
Enhanced psychiatric hosp. outpatient visit	318	48 %	84 %	3 097	5 388	984 924
Urgent visit	2 344	12 %	16 %	314	438	736 931
Home visit	369	28 %	33 %	1 501	1 813	553 898
Telephone reception visit	2 369	11 %	19 %	126	216	299 110
Consulting, supervision and specialist visit	1 343	14 %	16 %	187	215	251 186
Group visit	575	14 %	11 %	420	327	241 587
Family visit	409	14 %	16 %	288	334	117 646
Substance abuse services	**970**	**21 %**	**60 %**	**4 396**	**12 453**	**4 264 221**
Detoxification	612	22 %	59 %	4 273	11 391	2 615 086
Replacement therapy	134	44 %	57 %	5 313	6 857	711 995
Substance abuse residential care	44	82 %	95 %	11 179	12 993	491 868
Rehab center	217	31 %	38 %	1 757	2 148	381 264
Outpatient activities	424	25 %	34 %	151	201	64 008
Geriatric psychiatry	**63**	**49 %**	**96 %**	**13 717**	**26 666**	**864 167**
Long term geriatric psychiatric hospital	27	85 %	97 %	24 059	27 381	649 592
Geriatric psychiatry	38	32 %	92 %	5 647	16 407	214 575
Mental health services	**477**	**1 %**	**3 %**	**611**	**1 351**	**291 436**
Psychologist treatment/Health care center psychologist services	477	1 %	3 %	604	1 112	288 185
Psychologist group visit/Health care center psychologist services	6	50 %	10 %	542	108	3 251
Neuropsychology	**145**	**5 %**	**9 %**	**566**	**1 109**	**82 012**
Neuropsychological examination	137	6 %	5 %	419	394	57 411
Neuropsychological rehabilitation	19	16 %	19 %	1 295	1 538	24 601
Total	**5 820**	**10 %**	**62 %**	**2 992**	**18 407**	**17 415 455**

Earlier research has shown that HCUs use more service categories than non-HCUs (see e.g. Leskelä et al., 2013). This was also confirmed to be the case with MH/SA customers, as HCUs tend to use more service categories with an average of 3,9 service categories per year, while non-HCUs' average number is 3,1 service categories (see Figure 7).

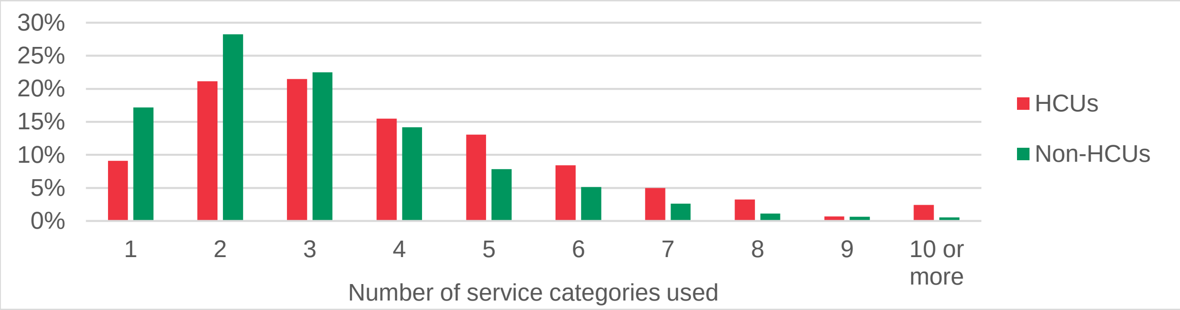


Figure 7. Distribution of service categories used, 2014

5.2 Demographics and Service Use

The analysis of the demographics includes the analysis of MH/SA HCUs' gender and age, as postal code areas are analyzed in an own section (5.5). The results show that all the age groups are relatively equally represented in the MH/SA HCU population, while younger customers are slightly more represented. Age groups' shares from the youngest to the oldest are 21 %, 17 %, 13 %, 13 %, 14 %, 10 % and 12 %. Gender distribution is also quite even, as 47 % of HCUs are females and 53 % are males.

When studying gender distribution by age groups, it is an interesting finding that female MH/SA HCUs are more often young than old, which means that MH/SA problems are more common among younger females. However, older female customers are more expensive after being 50 years old (avg. > 25 000 € / year) mainly due to increased geriatric psychiatry and other costs. Females use clearly more MH services than SA services, and the use of SA services is divided evenly across the age groups. See Figure 8.

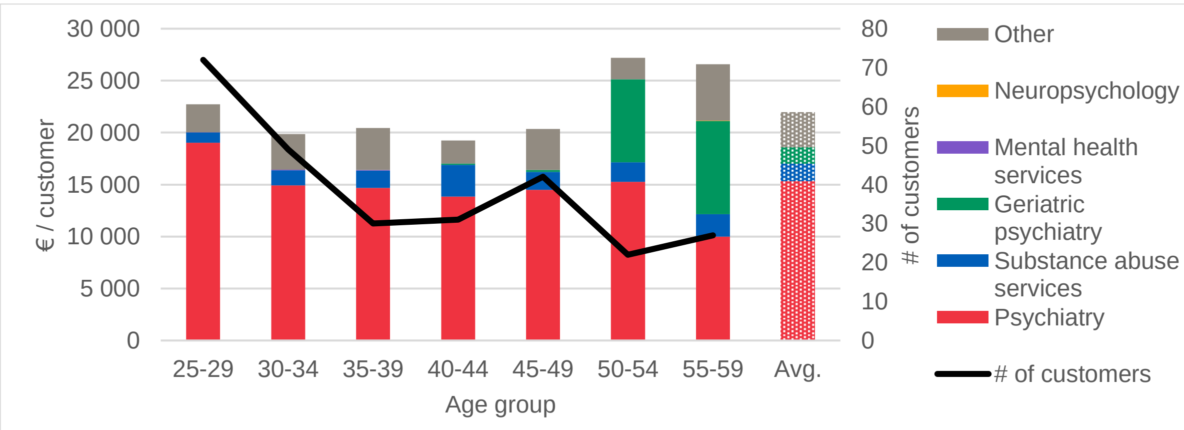


Figure 8. Female HCUs' service categories by age groups, 2014

Unlike females, males are divided more evenly across the age groups and there are only slightly more young customers. Also as in females, over 50-year-old males are more expensive (avg. > 25 000 € / year) mainly due to increased geriatric psychiatry costs. However, interestingly, males are using a lot more SA services than females, and as the age increases, even more and more SA services occur. In contrast to females, males' other costs have increased also among middle-aged customers. See Figure 9.

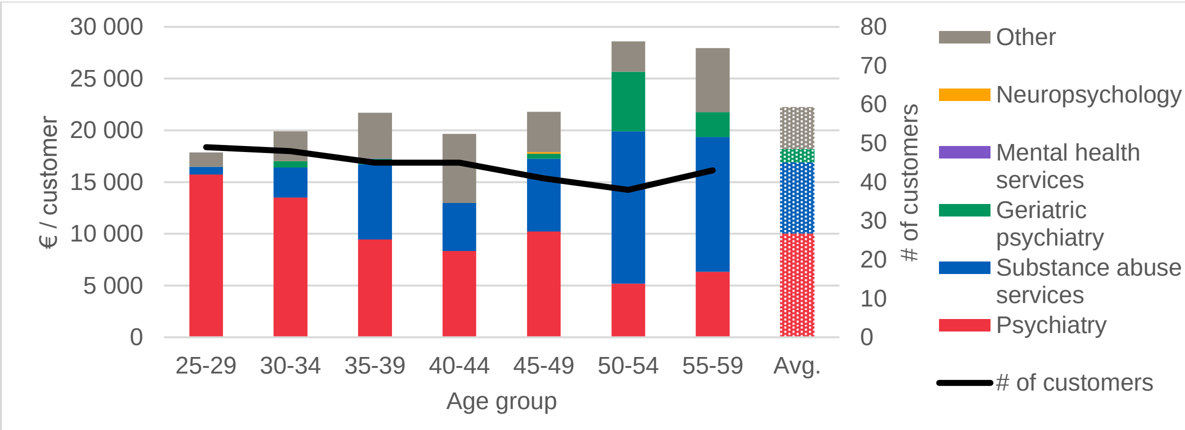


Figure 9. Male HCUs' service categories by age groups, 2014

Females' and males' average MH/SA costs are quite similar, ~18 600 € and ~18 200 € respectively. Other costs differ relatively more between the genders, as females' average other costs are ~3 300 € and males' ~4 000 €. When comparing genders by age groups, average costs seem to vary more. For example, young females have higher costs than young males. Increasing total costs are similar in both genders. This can be explained by other health issues related to older age. Moreover, MH/SA problems might have caused serious illnesses if the problems have existed for a long time. See Appendix D for more detailed tables.

The next step was to further investigate what are the other costs these customers have caused (the small grey share in previous figures). Among female, neurology and internal medicine are common other services, while geriatrics and variety of other services increase after the age of 55. With a few exceptions, other costs remain rather stable across the age groups. See Figure 10.

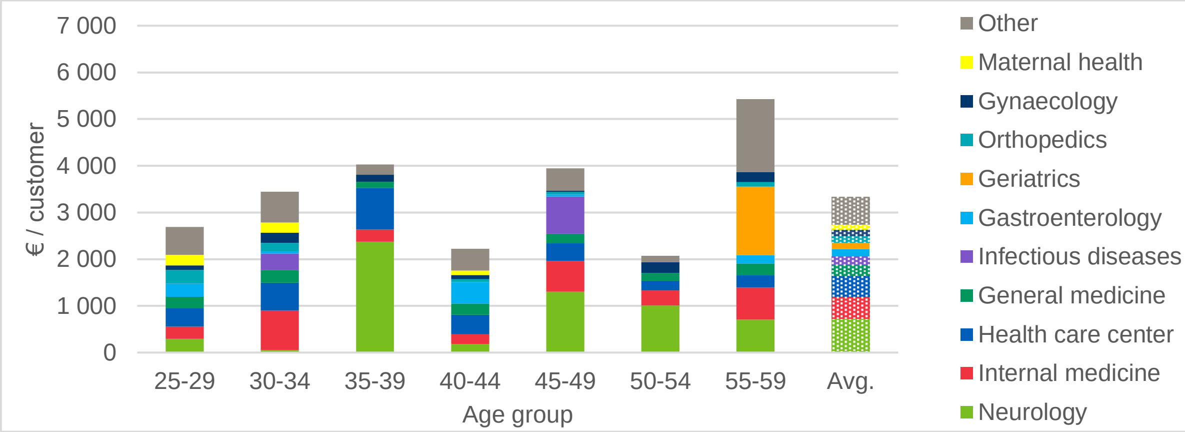


Figure 10. Breakdown of female HCUs' other costs, 2014

For males, internal medicine and health care center activities are common other services (see Figure 11). Also, geriatrics and neurology services increase after the age of 55. Similarly, other costs increase after the age of 55. However, there is also a peak in other costs among the 40-44-year-olds, whereas that age group has relatively low other costs among females. Altogether, males have higher other costs than females.

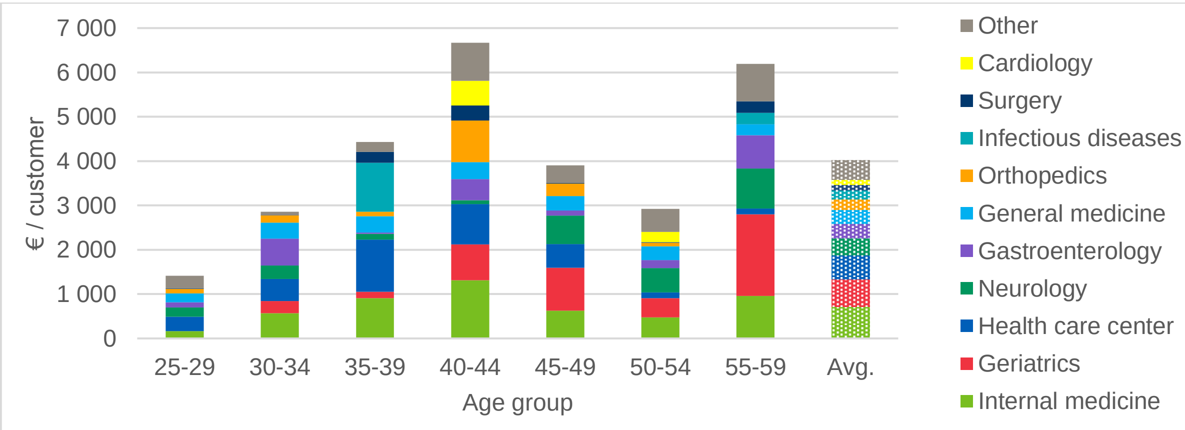


Figure 11. Breakdown of male HCUs' other costs, 2014

Comparison to non-HCUs

Similar cost structure analysis was conducted for non-HCUs in order to investigate the differences in the service use between HCUs and non-HCUs. Recall, non-HCUs are also MH/SA customers, but they don't belong to the costliest 10 % (=HCUs), i.e. they are the least expensive 90 % (=Bottom 90).

As in HCUs, there are more young than old females. Moreover, other costs, and thus total costs, increase with age, and females use mainly MH services. Psychiatry costs are almost twice as high among 25-29-year-olds as 55-59-year-olds. See Figure 12.

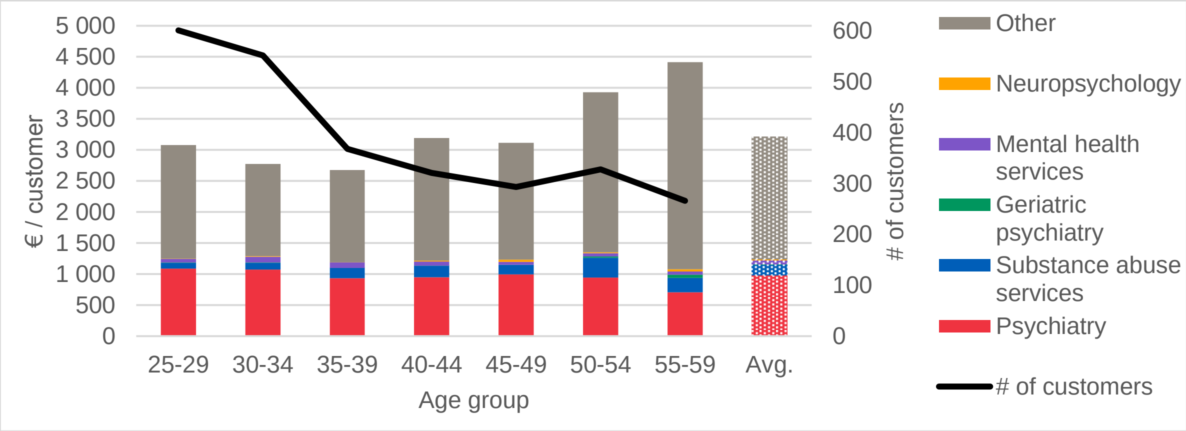


Figure 12. Female NON-HCUs' (bottom 90) service categories by age groups, 2014

For males, other costs and total costs, increase with age. Also SA services' share is again higher and increases with age. There are also slightly more young males. Female and male non-HCUs' MH/SA costs are quite similar, ~1 200 € and ~1 300 € respectively. See Figure 13.

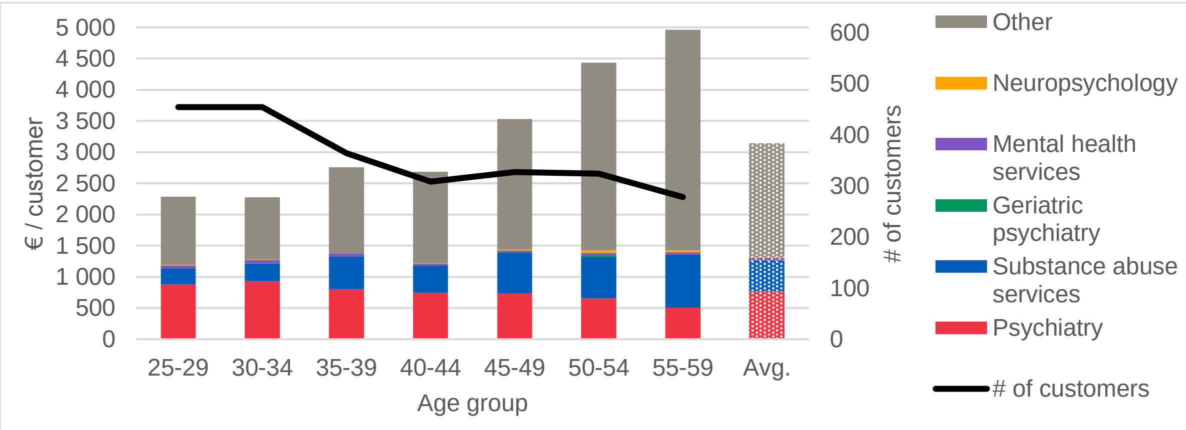


Figure 13. Male NON-HCUs' (bottom 90) service categories by age groups, 2014

The analysis continued also in the non-HCUs' case by exploring their other costs (the grey share in previous figures). Compared to HCUs, female non-HCUs have caused less other costs, about 2 000 € (HCUs ~3 300 €). Also, compared to HCUs, non-HCUs have used less general medicine and infectious diseases services. See Figure 14.

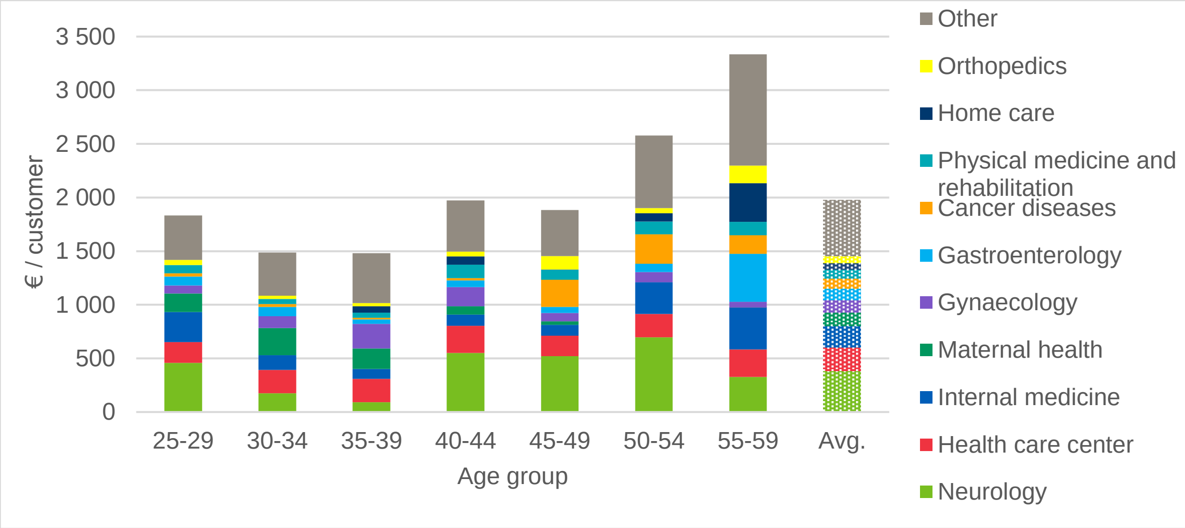


Figure 14. Breakdown of female NON-HCUs' (bottom 90) other costs by age groups, 2014

Interestingly, male non-HCUs have significantly less other costs, about 1 800 € (HCUs ~4 000 €). Non-HCUs also use less geriatrics and general medicine services. See Figure 15.

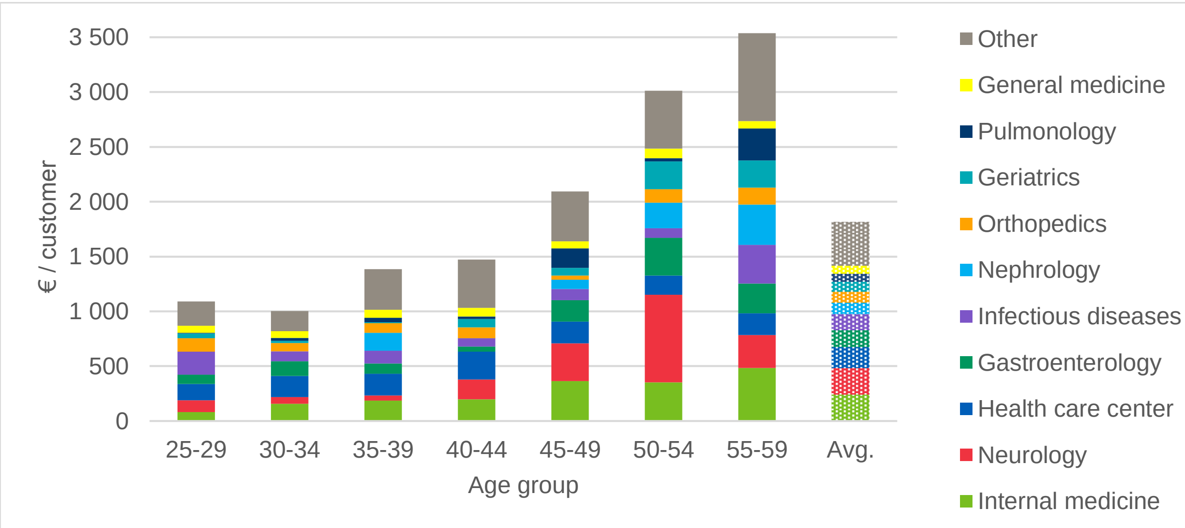


Figure 15. Breakdown of male NON-HCUs' (bottom 90) other costs by age groups, 2014

Figure 16 concludes the gender and age analysis. It is clear, that in addition to significantly higher MH/SA costs, HCUs have almost twice as high other costs as non-HCUs.

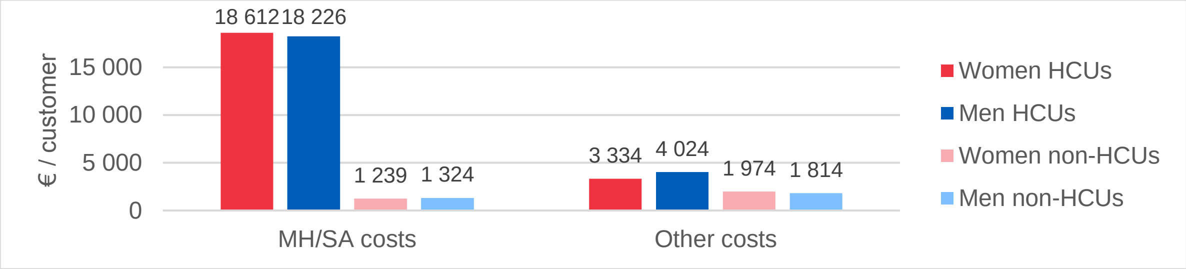


Figure 16. Average MH/SA and other costs by gender and HCU status, 2014

5.3 Segmentation of the Population

It is clear, that even though there are not that many MH/SA HCUs, they are actually a very diverse group of people with different service needs. Therefore, it is necessary to segment the people into homogeneous groups, so that the differences between the groups are clear in some term. As discussed, the purpose of using segmenting in social and health care is to divide individuals into large enough groups by some important aspects (Lillrank et al., 2010), in order to achieve cost reductions by leveraging economies of scale and mass production. It is important that the people in each segment have similar service needs, rhythm of needs, and priorities, in order to make the segment useful for planning (Lynn et al., 2007). As this study focuses mainly on HCUs the segmenting was also done for these people, and more specifically HCUs of year 2014 (n=582).

5.3.1 Segmentation Process and Results

The first two segments were created solely based on diagnoses, because the people with those diagnoses already form extraordinary groups with special needs. The first segment includes all the customers with a schizophrenia diagnosis (ICD-10: F20-F29) and the second segment includes all the customers with a mood disorder diagnosis (ICD-10: F30-F39). Rest of the HCUs with neither of the diagnoses (n=230) were divided into segments by using SPSS software's Two-Step Cluster Analysis and by using the customer data discussed earlier.

As there is no single method on deciding the suitable amount of clusters, the decision was made based on exploring various cluster solutions. In addition, hierarchical cluster analysis, with Ward Linkage and Squared Euclidean Distance, was utilized to find the most suitable number of clusters. The challenge was to keep all the clusters similar in terms of sizes without sacrificing cluster quality in terms of cohesion and separation.

As mentioned, the final clustering solution was executed by using Two-Step Clustering algorithm, Log-likelihood distance measure and Schwarz's Bayesian Criterion (BIC). Based on exploring possible cluster solutions, the number of clusters was set to be three. Used inputs (all describing year 2014) include

- Number of service categories used (continuous)
- Most used detailed service based on annual costs (categorical)
- MH costs (continuous)
- SA costs (continuous)
- Other costs (continuous)
- Number of other diagnosis groups (excl. mental health diagnoses) (continuous)
- Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified diagnosis (binary)
- Number of ER visits (continuous)

Clustering analysis was ran by using SPSS Modeler. After running the analysis, SPSS Modeler determines predictor importance, indicating that variables with large importance are more appropriate as predictors. Thus, predictor importance indicates the relative importance of each predictor and does not relate to model accuracy. Figure 17 presents the list of the selected input variables including the predictor importance, and Figure 18 presents the cluster sizes (63, 75, and 92), and cluster quality measure chart, using silhouette coefficient, which indicates that the overall model quality is "Fair".

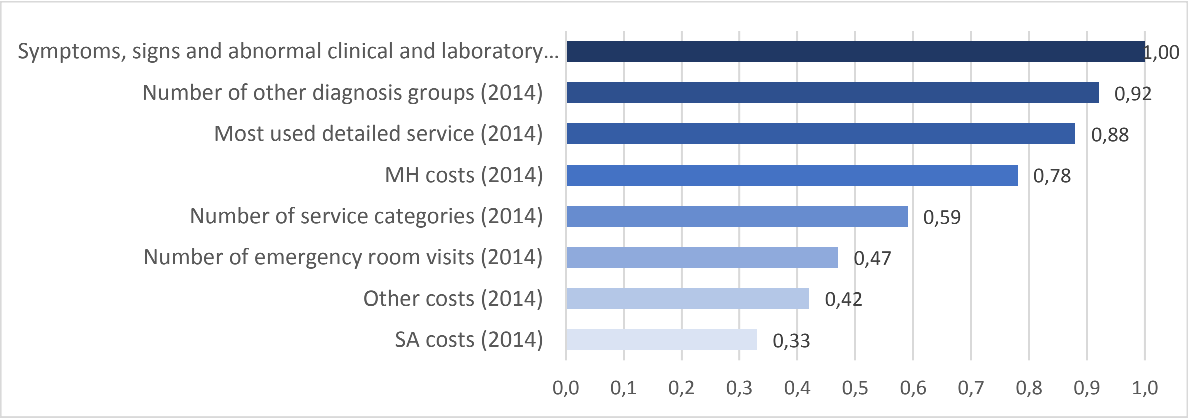


Figure 17. Input variables with predictor importances

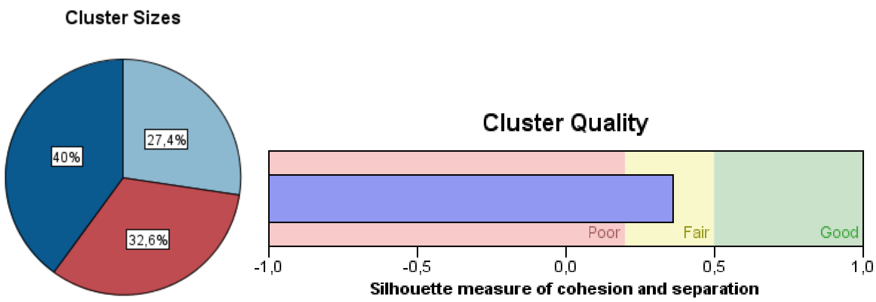


Figure 18. Cluster sizes and cluster quality (silhouette coefficient)

After forming the clusters and assigning each customer into the clusters, continuous variables were analyzed by using Oneway Descriptives. See Table 7.

Table 7: Oneway descriptives of continuous variables

	Cluster	N	Mean	Std. Deviation	Std. Error
Number of ER visits	1	63	0,67	1,470	0,185
	2	75	6,09	5,957	0,688
	3	92	1,52	2,438	0,254
	Total	230	2,78	4,458	0,294
Number of service categories used	1	63	2,79	1,557	0,196
	2	75	5,95	2,283	0,264
	3	92	3,42	1,718	0,179
	Total	230	4,07	2,298	0,152
Number of other diagnosis groups	1	63	0,41	0,754	0,095
	2	75	2,77	1,467	0,169
	3	92	0,84	0,855	0,089
	Total	230	1,35	1,466	0,097
MH costs	1	63	20 485	17 666	2 226
	2	75	4 483	6 707	774
	3	92	262	972	101
	Total	230	7 178	13 027	859
SA costs	1	63	830	1 963	247
	2	75	10 847	12 850	1 484
	3	92	16 237	15 769	1 644
	Total	230	10 259	13 862	914
Other costs	1	63	1 297	2 223	280
	2	75	11 655	13 311	1 537
	3	92	2 654	2 876	300
	Total	230	5 217	9 073	598

Continuous variables were also analyzed by using Oneway Anova. All variables' significance levels were below 0,05 and, therefore, there are statistically significant differences in the means of these variables.

The categorical variable, detailed service, was analyzed by using crosstab (see Table 8). Cluster 1's common detailed services are psychiatry, enhanced psychiatric hospital care day and long term geriatric psychiatric hospital. Cluster 2's most common detailed services are detoxification and psychiatry. However, many of them also have other detailed services, as there are altogether 19 different detailed services. Cluster 3's services more often used are detoxification, replacement therapy and substance abuse residential care, with no other detailed services.

Table 8: Detailed service crosstab

Detailed service	Cluster			
	1	2	3	Total
Cardiology		1,3 %		0,4 %
Detoxification		34,7 %	52,2 %	32,2 %
Ear, nose and throat diseases		1,3 %		0,4 %
Enhanced psychiatric hospital care day	28,6 %	1,3 %		8,3 %
Enhanced psychiatric hospital outpatient visit	6,3 %			1,7 %
General medicine and geriatrics service line		9,3 %		3,0 %
Geriatric psychiatry		1,3 %		0,4 %
Health care center activity	1,6 %			0,4 %
Infectious diseases		5,3 %		1,7 %
Internal medicine service line		1,3 %		0,4 %
Long term geriatric psychiatric hospital	17,5 %	1,3 %		5,2 %
Lung diseases		1,3 %		0,4 %
Nephrology		1,3 %		0,4 %
Neurology		1,3 %		0,4 %
Neuropsychological rehabilitation		1,3 %		0,4 %
Neurosurgery		1,3 %		0,4 %
Orthopedics		1,3 %		0,4 %
Psychiatry	42,9 %	21,3 %		18,7 %
Rehab center	3,2 %	1,3 %		1,3 %
Replacement therapy		9,3 %	32,6 %	16,1 %
Substance abuse residential care		2,7 %	15,2 %	7,0 %
Total	100,0 %	100,0 %	100,0 %	100,0 %

The binary variable, "symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified diagnosis" (ICD-10: R00-R99), was also analyzed by using crosstab (see Table 9). In this binary variable, 0 means that the customer has not received the diagnosis, while 1 means that the customer has received the diagnosis at least once during the year. Cluster 3's customers have all received the diagnosis, and out of cluster 1's population 98,4 % have also received the diagnosis. However, out of cluster 2's population only 22,7 % have received the diagnosis.

Table 9: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified diagnosis crosstab

		Cluster			
		1	2	3	Total
Symptoms, signs and abnormal clinical and laboratory findings	0	98,4 %	22,7 %	100,0 %	74,3 %
	1	1,6 %	77,3 %		25,7 %
Total		100,0 %	100,0 %	100,0 %	100,0 %

The next section is reserved for presenting detailed cluster profiles, including the two segments based on schizophrenia and mood disorder diagnoses.

5.3.2 Segment Profiles' Characteristics and Service Use

As discussed in the previous section, there are five segments, out of which two are based on diagnosis (schizophrenia and mood disorder) and three are based on cluster analysis. The segments are named as follows

1. Mood disorder (diagnosis based)
2. Schizophrenia (diagnosis based)
3. Other mental health (cluster 1)
4. Substance abuse (cluster 3)
5. Super service user (cluster 2)

As seen in Table 10, segment sizes vary from 63 to 185, and total segment costs from 1,4 M€ to 4,1 M€.

Table 10: Summary of segments

Segment	n	Total segment costs
Mood disorder	185	3,6 M€
Schizophrenia	167	4,1 M€
Other mental health	63	1,4 M€
Substance abuse	92	1,8 M€
Super service user	75	2,0 M€

First, segments are analyzed based on the input variables describing service use (see Figure 19). As the results show, super service users use more service categories than others with an average of 5,9. They have also dramatically more ER visits (6,1) than others. Logically, they also have more different diagnoses (2,8). Note that other diagnosis groups measure doesn't include mental health diagnoses. When analyzing the number of service categories used, diagnosis groups and ER visits, other segments don't differ that much from each other. Their profiles also are quite similar to non-HCUs' profile. However, an average HCU has more service categories used, more other diagnosis groups and more ER visits than an average non-HCU.

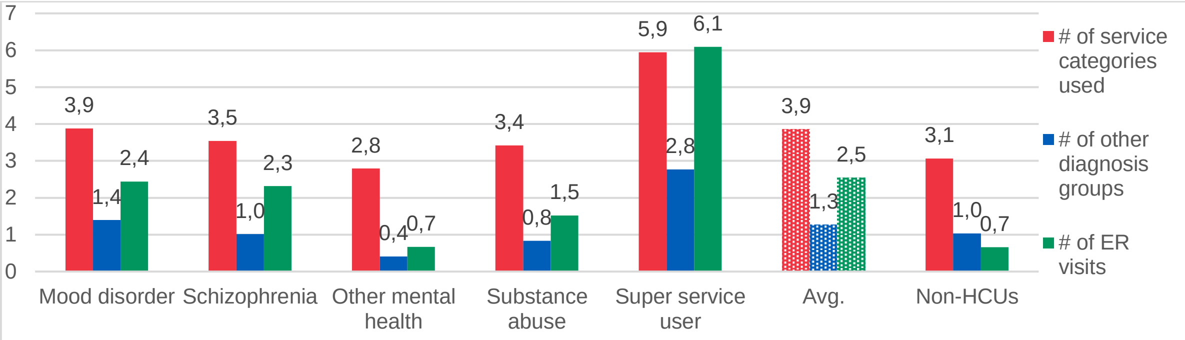


Figure 19. HCUs' service use and diagnosis characteristics by segment

Second, segments are analyzed based on their service categories used and related costs (Figure 20). Mood disorder and schizophrenia segments have similar cost structure, even though schizophrenics have higher average total cost (~24 500 € vs. ~19 300 €). Other mental health segment differs by having a lot more geriatric psychiatry services. All these three segments don't use that much substance abuse services. However, substance abuse segment uses almost solely substance abuse services. Probably the most interesting group, super service users, use some MH services and a bit more SA services. However, their average total cost is a lot higher and they use also very much other services (~11 700 €), while other groups' other costs vary between ~1 300 € and ~3 330 €.

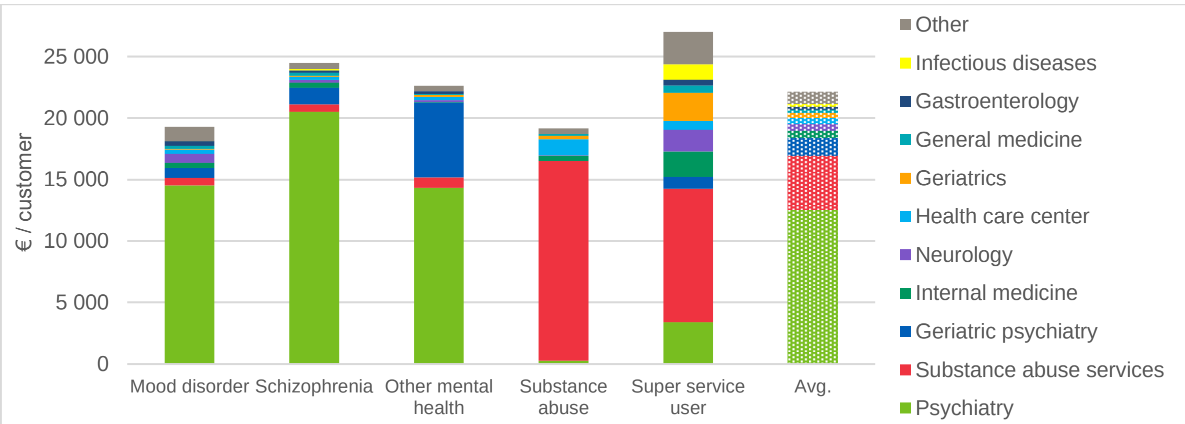


Figure 20. HCUs' service categories by segment

As segments seem to differ by their other services use, it is analyzed in more detail in Figure 21. The most interesting findings include the substance abuse segment's high use of health care center services, other mental health segment's lack of use of internal medicine services and high gastroenterology services, and most importantly, super service user segment's extremely high other cost due to high service use of several service categories, e.g. geriatrics, internal medicine, neurology and infectious diseases.

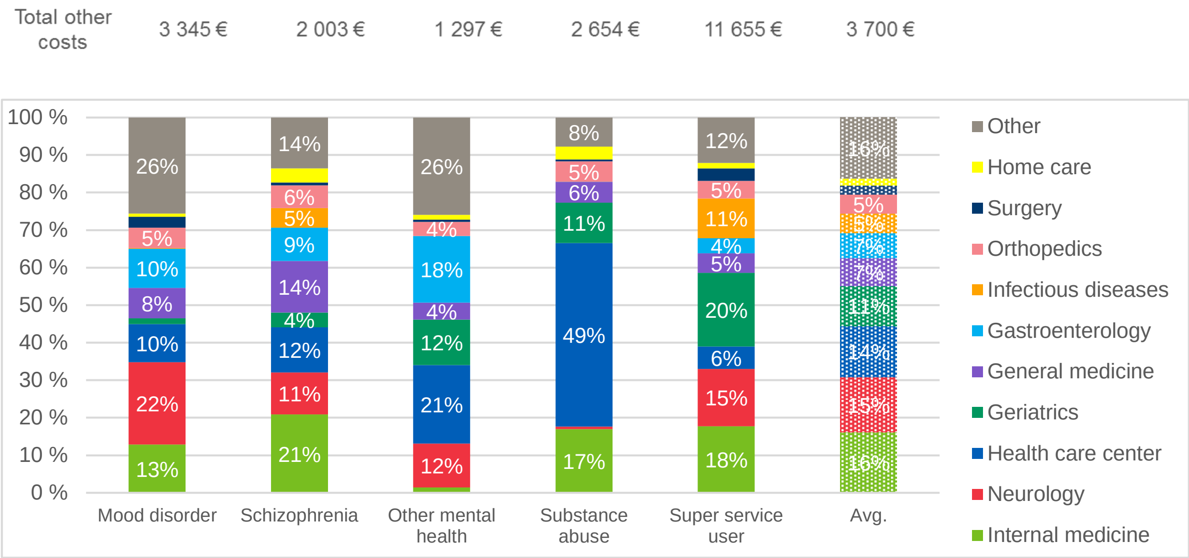


Figure 21. Top 10 other service categories (not MH/SA services) by segment

Next, the segments' demographics are analyzed in Figure 22. Gender differences are quite clear, as females belong most likely to the mood disorder and schizophrenia segments, while other mental health, substance abuse and super service user segments are rare for them. Males belong most likely to the mood disorder, substance abuse and schizophrenia segments. It is interesting that 17 % of the females belong to the super service user and substance abuse segments, while the same percentage for males is 39 %.

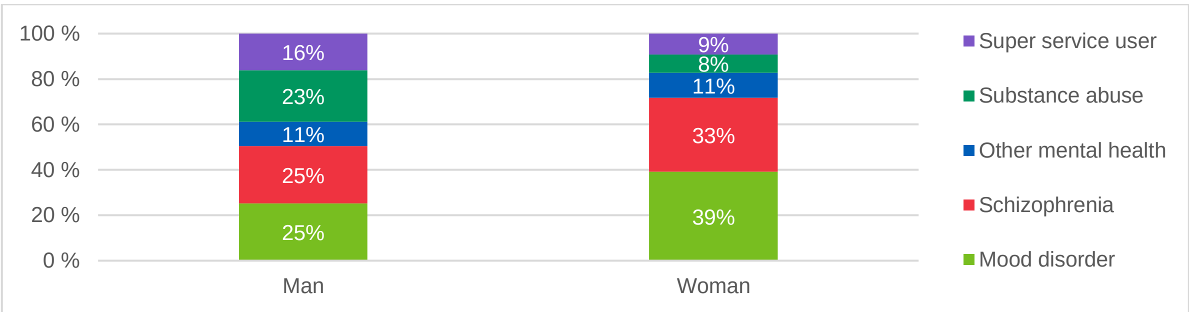


Figure 22. HCUs' segments by gender

Age group differences are also clear (Figure 23). Super service user and substance abuse segments become more common as the age increases, as mood disorder segment becomes rarer. Other mental health and schizophrenia segments seem to remain rather stable across the age groups.

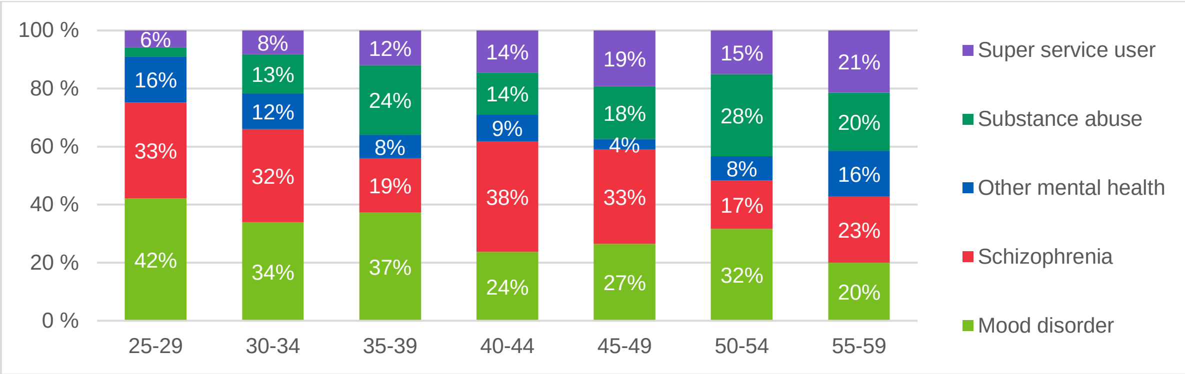


Figure 23. HCUs' segments by age groups

5.4 Diagnoses

This section is reserved for diagnostic analysis, which was done in the following manner. If a customer had received a certain diagnosis even once during the year, he or she was marked as having the diagnosis, no matter how many times the diagnosis was received during the year. It is important to notice that diagnosis data was possibly incomplete, because some visits may not have yielded any diagnosis or may not have been documented in the system at all. In 2013, out of 336 162 visits, 32,9 % included a diagnosis mark, and in 2014, out of 348 731 visits, 39,7 % included a diagnosis mark.

Analysis of the diagnostic differences between HCUs and non-HCUs revealed expected results (see Table 11). The HCUs have more diagnoses in most of the diagnosis groups. The biggest differences are in schizophrenia and delusional disorders; injury, poisoning and certain other consequences of external causes; other mental and behavioral disorders; infectious diseases; heart disease; blood and blood-forming organs diseases; and HIV disease.

On the other hand, some diagnosis groups are more common among non-HCUs, such as pregnancy, childbirth and the puerperium; respiratory system diseases; tumors; and genitourinary system diseases. However, some diagnoses, for example congenital malformations, deformations and chromosomal abnormalities, are extremely rare among all people, making the comparison unreliable with such a small population.

Table 11: Occurrence of diagnoses among HCUs and non-HCUs

Diagnosis group	HCUs	Non-HCUs	Difference
Other mental and behavioral disorders	54 %	31 %	+ 77 %
Mood disorders	37 %	35 %	+ 5 %
Schizophrenia and delusional disorders	29 %	14 %	+ 102 %
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	23 %	17 %	+ 34 %
Injury, poisoning and certain other consequences of external causes	22 %	12 %	+ 89 %
Musculoskeletal system and connective tissue diseases	15 %	13 %	+ 13 %
Respiratory system diseases	9 %	11 %	- 16 %
Nervous system diseases	8 %	8 %	+ 2 %
Infectious diseases	7 %	4 %	+ 73 %
Skin and subcutaneous tissue diseases	7 %	5 %	+ 26 %
Endocrine, nutritional and metabolic diseases	6 %	5 %	+ 16 %
Digestive system diseases	6 %	5 %	+ 11 %
Circulatory system diseases	6 %	4 %	+ 43 %
Genitourinary system diseases	5 %	5 %	- 2 %
Eye and adnexa diseases	4 %	4 %	+ 11 %
Ear and mastoid process diseases	3 %	3 %	+ 26 %
Tumors	3 %	3 %	- 16 %
Diabetes	3 %	3 %	+ 2 %
Heart disease	2 %	1 %	+ 73 %
Pregnancy, childbirth and the puerperium	2 %	3 %	- 41 %
Liver disease	1 %	1 %	+ 38 %
Blood and blood-forming organs diseases	1 %	1 %	+ 50 %
Congenital malformations, deformations and chromosomal abnormalities	1 %	0 %	+ 17 %
HIV disease	0 %	0 %	+ 50 %

Diagnosis analysis of the differences between the segments also revealed some interesting findings (Table 12). As expected and discussed earlier, super service users are using a lot of other services and have more diagnoses than other segments in most of the diagnosis groups. The biggest difference is in symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified diagnosis group, which consists of, for example, cases for which no specific diagnosis can be made, and certain symptoms that represent important problems in medical care in their own right. Super service users also have significantly more injuries, poisonings and certain other consequences of external causes; respiratory system diseases; and infectious diseases.

Mood disorder and schizophrenia segments are quite similar in terms of diagnoses. However, for some reason, mood disorder segment has double the amount of musculoskeletal system and connective tissue diseases, and nervous system diseases compared to the schizophrenia segment.

Other mental health segment has the least diagnoses and they seem to be quite healthy, except having other mental and behavioral disorders. Substance abuse segment, in turn, has some special characteristics compared to mood disorder, schizophrenia and other mental health segments. For example, they have significantly more injuries, poisonings and certain other consequences of external causes and infectious diseases.

Table 12: Occurrence of diagnoses among segments

Diagnosis group	Mood disorder	Schizo-phrenia	Other mental health	Substance abuse	Super service user
Other mental and behavioral disorders	50,8 %	40,7 %	55,6 %	59,8 %	84,0 %
Mood disorders	100,0 %	17,4 %			
Schizophrenia and delusional disorders		100,0 %			
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	24,9 %	16,2 %	1,6 %		77,3 %
Injury, poisoning and certain other consequences of external causes	19,5 %	19,2 %	9,5 %	25,0 %	42,7 %
Musculoskeletal system and connective tissue diseases	21,6 %	11,4 %	4,8 %	12,0 %	21,3 %
Respiratory system diseases	8,6 %	7,8 %	3,2 %	9,8 %	20,0 %
Nervous system diseases	11,4 %	4,2 %	3,2 %	3,3 %	16,0 %
Infectious diseases	3,8 %	2,4 %	3,2 %	10,9 %	21,3 %
Skin and subcutaneous tissue diseases	7,6 %	6,0 %		6,5 %	10,7 %
Endocrine, nutritional and metabolic diseases	5,9 %	7,2 %	3,2 %	2,2 %	12,0 %
Digestive system diseases	6,5 %	4,8 %	4,8 %	1,1 %	14,7 %
Circulatory system diseases	4,3 %	4,2 %		5,4 %	17,3 %
Genitourinary system diseases	5,4 %	7,2 %	4,8 %	1,1 %	5,3 %
Eye and adnexa diseases	5,9 %	3,6 %		3,3 %	6,7 %
Ear and mastoid process diseases	6,5 %	2,4 %	1,6 %	2,2 %	1,3 %
Tumors	2,7 %	3,0 %		1,1 %	5,3 %
Diabetes	1,1 %	3,6 %	3,2 %		6,7 %
Heart disease	1,1 %	0,6 %		2,2 %	6,7 %
Pregnancy, childbirth and the puerperium	3,2 %		1,6 %		2,7 %
Liver disease		1,2 %		4,3 %	2,7 %
Blood and blood-forming organs diseases	1,1 %	1,8 %			1,3 %
Congenital malformations, deformations and chromosomal abnormalities	0,5 %	0,6 %			1,3 %
HIV disease					2,7 %

5.5 Postal Code Area

Analyzing MH/SA HCUs based on their postal code area revealed very interesting findings (Figure 24). The central area has relatively most HCUs (number of HCUs divided by total amount of customers), while southwest and northwest areas have relatively least HCUs. Note that the baseline is at 10 %, because that is the set limit for high-cost use. See Appendix C for area and postal code grouping.

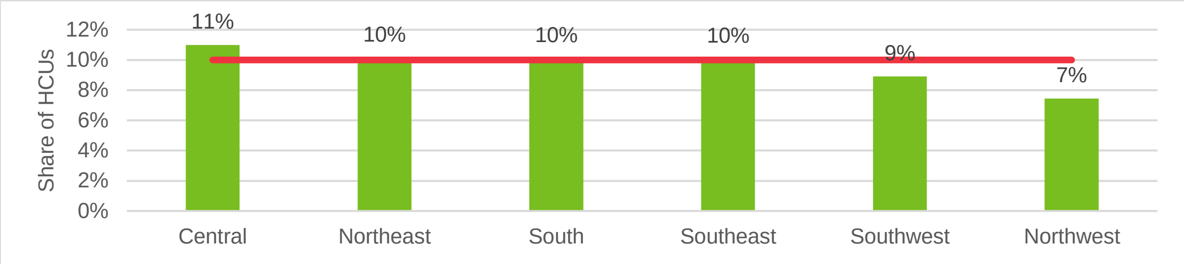


Figure 24. Number of HCUs divided by the total amount of customers in the area, 2014

Analyzing MH/SA and other costs in the different areas also revealed some differences. First of all, SA services are used relatively most in the south and southwest areas, while MH services are used relatively most in the northwest and central areas. Total average costs are highest in the central, south and northeast areas. Over half of the MH/SA HCUs come from the central and southeast areas, which is mostly explained by higher populations. Figure 25 summarizes the results and Table 13 provides more detailed information (see appendix D for similar table for year 2013).

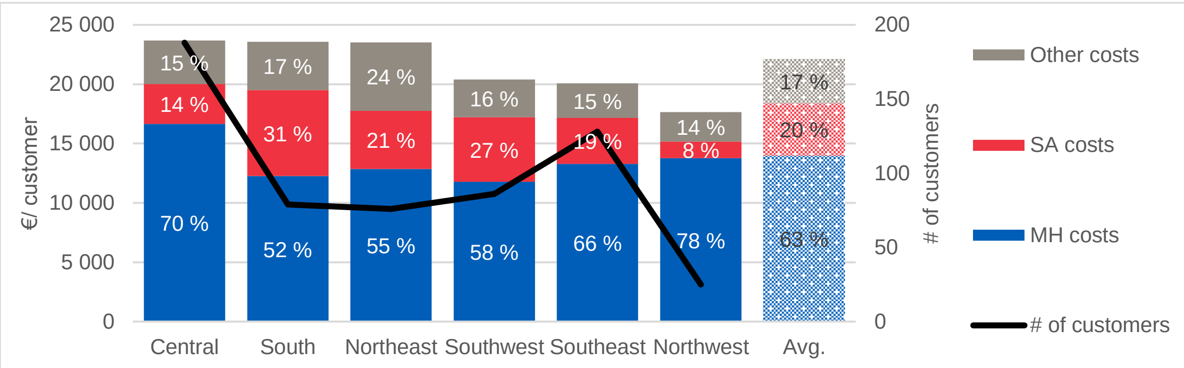


Figure 25. HCUs' service categories by area, 2014

Table 13: Distribution of social and health care costs among HCUs by area, 2014

Area	n	% of HCUs	Total MH/SA costs (€)	Avg. MH/SA costs (€)	Avg. other costs (€)	Avg. total costs (€)
Central	183	31 %	3 700 363	20 221	3 634	23 676
Southeast	130	22 %	2 214 566	17 035	3 498	20 345
Southwest	86	15 %	1 480 023	17 210	3 440	20 409
South	83	14 %	1 612 285	19 425	4 181	23 354
Northeast	76	13 %	1 333 458	17 545	5 850	23 241
Northwest	24	4 %	372 078	15 503	2 355	17 859
Grand Total	**582**	**100 %**	**10 712 773**	**18 407**	**3 894**	**22 107**

The analysis on the areas of Tampere doesn't reveal severe differences in segment distribution between the areas (Figure 26). However, customers from south and southwest seem to belong to substance abuser segment more often than others.

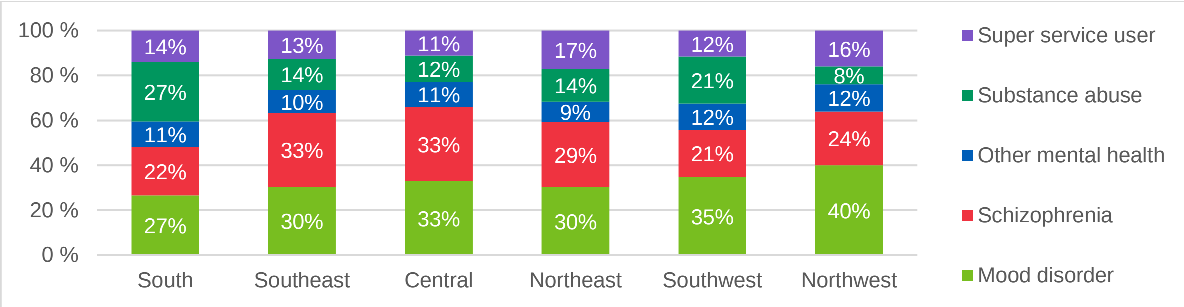
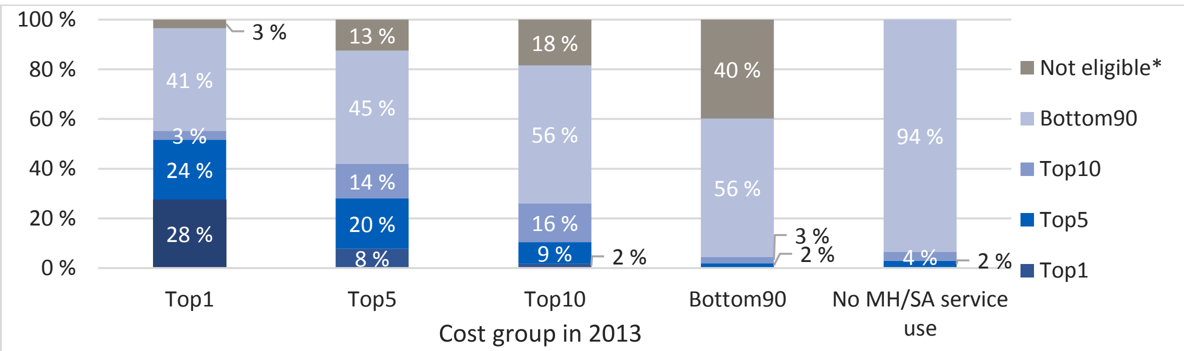


Figure 26. HCUs' segments by areas

5.6 Persistence of Being a High-Cost User

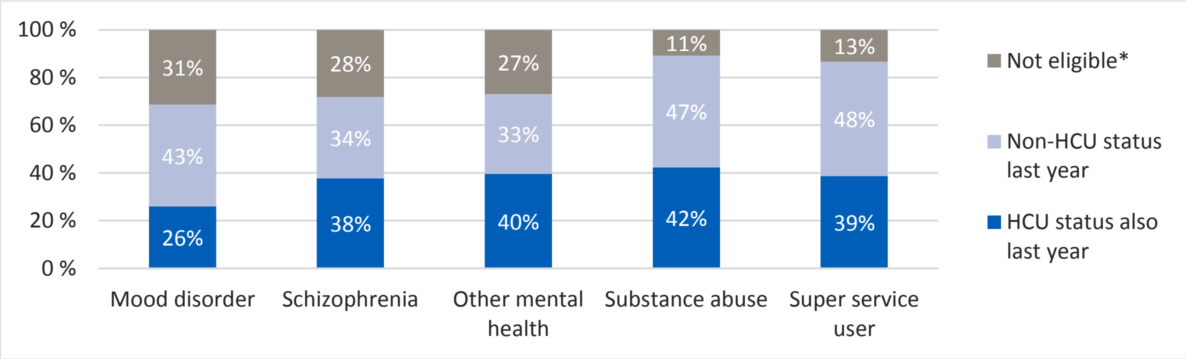
In order to see the short term persistence of being a HCU, all customers observed in 2013 were divided according to their HCU status in 2014 (Figure 27). Note that the X axis represents customers' HCU status in 2013, while series indicate their 2014 HCU status. The results revealed that being a MH/SA HCU is a quite stable status in the highest cost groups, at least in the two-year time period. The more expensive the customer, the more likely he or she stays expensive in the following year. Out of the Top 1 customers, 55 % were HCUs also in the following year, from Top 5 customers, 42 % remained HCUs, and from Top 10 customers 25 % remained HCUs. Out of the Bottom 90 customers only 5 % became HCUs in the following year. The customers not eligible either moved away from Tampere, didn't use any MH/SA services or died. A further investigation is needed to understand why and when MH/SA customers stop being MH/SA customers.



* Died, moved or did not use MH/SA services

Figure 27. Customers observed in 2013 (X axis) according to their 2014 expenditure status

The persistence of being a HCU by segment was also analyzed. Figure 28 differs from the previous figure (Figure 27) so that the X axis represents 2014 customers by their segment, and the series indicate their previous year's HCU status. The results indicate that mood disorder patients are the least persistent HCUs, as only 26 % of them were HCUs also previous year. Other segments' same share is around 40 %, while substance abuse segment is most likely to stay HCUs, with 42 % being HCUs also during the previous year.



* Did not live in Tampere or did not use MH/SA services

Figure 28. HCUs' previous year's HCU status by segment (X axis)

5.7 Predicting High-Cost Users

This section is reserved for predictive modeling with the aim to identify future HCS of year 2014 based on the 2013 data. As discussed earlier, predictive models differ in four principal ways: predicted event, set of predictor variables, time period and statistical technique. In this case, predicted event is the move from the non-HCU group to the HCU group, the time period is two years, the statistical technique is logistic regression, and the predictor variables (all from year 2013) include background variables, other costs, MH/SA costs (and total costs), diagnoses, number of ER visits and number of service categories used (see Appendix E).

Predicting future MH/SA HCUs, meets the criteria (Lewis et al., 2011) of a predicted event. First, becoming a MH/SA HCU is an undesirable status, and by offering preventive services the quality of life and health status can be improved. Second, MH/SA HCUs are costly in terms of service resources, thus preventive services can generate future cost savings. Third, becoming a MH/SA HCU is preventable, at least in most of the cases. Fourth, these predictive risk models are built on routine administrative data, including a range of potential explanatory variables, for example age, deprivation, patterns of health service use, and a range of different diagnoses.

5.7.1 Logistic Regression

A predictive model was built and estimated by the logistic regression and SAS Enterprise Miner Workstation 13.1. First the data was partitioned into training (60 %) and validation (40%) sets. Then the interval or ratio-scaled explanatory variables underwent log-transformations to deal with highly skewed variables. The selection model used for the explanatory variables was the stepwise method (for entering and staying in the model p = 0,05).

The sample data included 6 631 customers (n = 6 631), who were non-HCUs in 2013 but had used at least some services in 2013 (total social and health care cost ≠ 0 €). In other words, the sample included all the people who had used MH/SA services in 2013 and/or 2014 excluding, however, those people who were already HCUs in 2013 and people with no costs in 2013. This exclusion was made as there was no information if the reason for the no service use was that they only recently had moved to Tampere, or if there was some other reason. Out of these 6 631 customers 313 (~4,7 %) became new HCUs in 2014, while 6 318 remained non-HCUs.

The explanatory variables used in the predictive model are from 2013, because the goal is to predict the following year's new HCUs in advance. This means that the predicted event is becoming a HCU after being a non-HCU; this is a binary variable where value 1 means that the customer became a HCU and 0 means that the customer remained a non-HCU. The explanatory variables included background variables (gender, area, age group), other costs, MH/SA costs (and total costs), number of emergency room visits, number of service categories used, and diagnosis groups as binary variables (list in Appendix E).

In predictive modeling it is possible to use decision weights which is a flexible way to take into account e.g. costs and benefits in applications. If no decision weights are used, then the model estimated emphasizes the identification of true positives (ones) and true negatives (zeros). However, in this case we may state that offering extra services for a person not becoming HCU (false positive) in 2014 causes costs. Moreover, not being able to identify a new HCU probably causes even more costs. Table 14 shows the benefit matrix used. True positives were set to be the most valued outcomes, because it means that a high-risk customer was identified correctly. Moreover, true negatives are also good outcomes, less valuable though, because in that case a non-high-risk customer was identified correctly. False positives, i.e. someone was identified as a high-risk without a reason, and false negatives, i.e. someone was identified as a non-high-risk with an actual high-risk, were assigned a negative value, because they are undesirable events. As there was no information available to assess the benefits and costs the figures used reflect no real situation.

Table 14: Benefit matrix for logistic regression

	True class	
Hypothesized class	True Positives 8,0	False Positives -0,5
	False Negatives -0,5	True Negatives 0,2

When running stepwise regression only one explanatory variable was introduced into the model, i.e. total social and health care costs (see Table 15). This means that higher total costs indicate higher risk of becoming a MH/SA HCU.

Table 15: Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5,32	0,42	157,16	< 0,0001
Total cost 2013 LOG	1	0,32	0,05	33,12	< 0,0001

The result of the prediction is displayed by the cumulative lift chart (Figure 29). By using the model and targeting the top 10 % of customers with the highest estimated probabilities to become a MH/SA HCU, the model is able to capture 10,2 % of the at-risk customers for intervention (instead of 4,7 % in the case of no model), meaning that the lift is 2,2. Thus the model brings some improvement but nothing really remarkable.

When carrying out sensitivity analysis of the weights one could find that often the logit produced the model with intercept only.

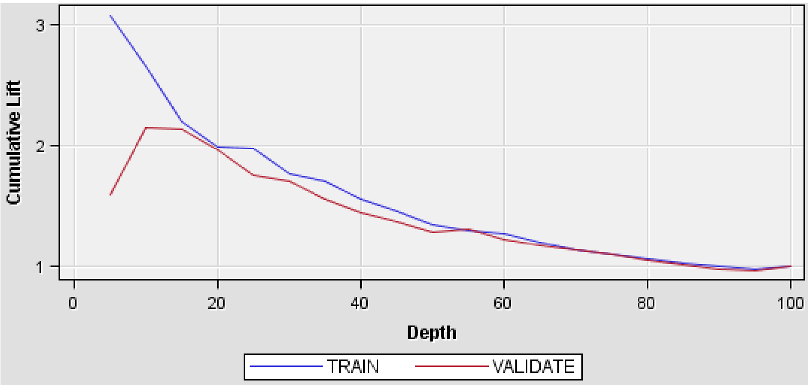


Figure 29. Cumulative lift of logistic regression

Also models with more specific cost data as explanatory variables were run with no improved success. Sensitivity analysis of the benefit matrix could bring some improvement. It is surprising that the model yielded only one significant explanatory variable even though n was high (6 631).

When considering the results, it should be remembered that the employed definition of the HCUs, the costliest 10 % of the population, is problematic. It cuts the population into positives and negatives and even slight differences in costs can separate observations into different categories. One could explore further the possibilities to predict future high-cost use using various limits, for example costliest 1 %, 5 % or even 20 %. As stated, now customers immediately below the HCU limit are treated as non-HCUs, which has an effect on the model.

5.7.2 Power of Predictive Modeling in Predicting Future High-Cost Users

To conclude, predicting future MH/SA HCUs by the available health and social care data is challenging. This study was not able to achieve a promising predictive model able to identify future high-cost users. On the other hand, the study revealed that total social and health care costs is an indicator that should be looked at when assessing a certain customer’s risk. Moreover, it is noteworthy that this identified indicator also includes MH/SA services, meaning that many of the customers are already using MH/SA services prior to becoming a MH/SA HCU. This can make identifying high risk patients easier, as they are already in contact with MH/SA professionals, and other professionals as well.

The customers analyzed in this study are all already using MH/SA services, or have at least used once during the year. It would be even more challenging to predict the people who are not using any MH/SA services yet. And as presented, there are people who have become MH/SA HCUs without using any MH/SA services previous year.

Curry et al. (2005) state that the most commonly used predictive variables are diagnoses and prior utilization combined with demographic data, and it is exactly what turned out to be the best option in this case as well. To enhance the performance of the logistic regression, as discussed earlier (Elissen et al, 2015), additional variables should be included in the data, for example income level, income source, household income, geographic disparities in terms of population and environment, available care services, type of household, size of household, housing circumstances, and degree of loneliness.

Another way of increasing the performance is to have longer time period of data. If 2012 data were available years 2012 and 2013 could be used to see the changes that have happened prior becoming a HCU in 2014. This study was only able to look at a static situation in 2013, without knowing possible changes in customer's life during the previous year. If year 2012 was included in the analysis, new variables indicating the change between 2012 and 2013 could be included and used as predictor variables. These predictive models were predicting the following year's HCUs, as a time horizon of greater than one year decreases the prediction power significantly (Panattoni et al., 2011).

6 Managing High-Cost Users

Finally, this section aims to understand what should be taken into account when planning services for HCUs, and to create a framework of potential methods to manage HCUs by each segment. Thus, suggestions for HCU management by segments is proposed. These management solutions and methods should enhance the care coordination process, take a person centered focus and facilitate management of complex conditions, problems and the felt needs. As MH/SA problems are somewhat lifestyle related problems and conditions, management should focus also on behavioral change and support, as well as personal contact.

As discussed, the purpose of using segmenting is to divide individuals into large enough groups by some important aspect (Lillrank et al., 2010), in order to achieve cost reductions by leveraging economies of scale and mass production. Logically, these suggestions are made for each of the five segments discussed earlier.

6.1 Mood Disorder Customers

A mood disorder customer, most likely a 25-39-year-old (61 %) female (58 %), uses mostly psychiatry services (75 % of total costs). She uses 4 different service categories, which are in addition to psychiatry, most likely neurology, internal medicine and gastroenterology. She has also 1-2 other diagnoses, in addition to a mood disorder (100 %), and other mental health and behavioral disease (51 %), which are most likely other unclassified symptoms and findings (25 %), musculoskeletal system disease (22 %) or injury, poisoning or other external cause (20 %). She is also rather likely to have a nervous system disease (11 %). She uses ER 2-3 times a year. Mood disorder customers' persistence of being a MH/SA HCU also previous year is only 26 %, what means that it is very shifty condition and status. Mood disorder customers' total annual costs are 3,6 M€ and there are 185 customers in this segment in total.

Mood disorder segment is based on mood disorder diagnosis, including manic episodes, bipolar disorders, depressive disorders, persistent mood disorders and other unspecified mood affective disorders. Their care should be organized based on disease management by organizing the care practices, especially hospital ward, accordingly based on medical guidelines. However, mood disorder customers might benefit from predictive risk modelling as the high-cost use is not very stable and the condition is somewhat preventable by efficient proactive measures, early stage support and self-management.

As mood disorder segment uses mainly psychiatry services, their care coordinator should be an assigned psychiatrist, who is also aware of other services the customer is using and physical problems he or she might have.

6.2 Schizophrenia Customers

A schizophrenia customer, most likely a 25-34 (43 %) or 40-49-year-old (34 %) female (53 %), uses a lot of psychiatry services (84 % of total costs). In addition to having a schizophrenia (100 %), a mood disorder (17 %) and other mental and behavioral disorders (41 %), she is likely to have 1 other diagnosis, which is probably an injury, poisoning or other external cause (19%), or other unclassified symptoms and findings (16 %). She goes to ER 2 times a year and uses 3-4 service categories, which are likely, in addition to psychiatry and geriatric psychiatry, internal and general medicines. There is a 38 % chance that she was a MH/SA HCU also previous year. Schizophrenia customers' total annual costs are 4,1 M€ and there are 167 customers in this segment in total.

Due to the nature of the disease, customers having a schizophrenia might not benefit from preventive services, but better coordination of care might be beneficial as they have very distinctive needs. As with mood disorder customers, their care should be organized based on disease management by organizing the care practices, especially hospital ward, accordingly based on medical guidelines.

6.3 Other Mental Health Customers

Other mental health customer, likely a 25-34 (49 %) or 55-59-year-old (17 %) male (52 %), doesn't have a mood disorder or schizophrenia diagnoses, but likely another mental and behavioral disorder (56 %). He is rather healthy and has only 0-1 other diagnoses, which is most likely an injury, poisoning or other external cause (10 %). He goes to ER less than 1 time a year, and he uses 3 service categories, including psychiatry, geriatric psychiatry and substance abuse services. He is unlikely to use other services, but if he does, it is likely health care center or gastroenterology services. There is a 40 % chance that he was a MH/SA HCU also previous year. Other mental health customers' total annual costs are 1,4 M€ and there are 63 customers in this segment in total.

Even though this segment is the smallest (63 customers), it could be helpful to divide this population further. As the likely age groups suggests, it is possible that this segment includes both older people using geriatric psychiatry, and younger individuals with other mental health and substance abuse problems. As other mental health customers are persistent and stable MH/SA HCUs, focus should be on supporting existing HCUs, at least if the case is the older geriatric customers. If the case is the young (possibly) symptomatic mental health customer, the focus should be on helping them to get the needed services. This might prevent problems and facilitate their wellbeing. This segment clearly should receive focus in the future.

6.4 Substance Abuse Customers

A substance abuse customer, most likely a 35-39-year-old (20 %) male (76 %), uses almost solely substance abuse services. He uses 3-4 service categories, and in addition to other mental and behavioral disorder diagnosis (60 %) he has only 1 other diagnosis, most likely an injury, poisoning or other external cause. He goes to ER 1-2 times a year, but he is a very active user of health care center services. He also uses some internal medicine specialty. Substance abuse segment is the most persistent segment (42 %), thus he might have been a MH/SA HCU also previous year. Substance abuse customers' total annual costs are 1,8 M€ and there are 92 customers in this segment in total.

Substance abuse problems are related to lifestyle choices that increase the likelihood to develop "lifestyle" diseases. These problems can develop slowly, so there is possibly enough time to offer preventive services for people at high risk before the situation gets out of control. Substance abuse patients who have already got severe health problems are in the super service user group discussed in the following section. Beneficial methods to manage substance abuse customers include self-management capabilities and lifestyle change support and management.

These customers' mainly use substance abuse services and their main diagnosis include almost solely injuries and poisonings taken care of at the health care centers. Due to this, their care should be managed my using disease management framework, which is a systematic, population-based approach to patient care that aims to curb utilization by optimizing the process of care, increasing efficiency, and managing the total disease by implementing programs targeting cost drivers within the group (Charlson et al., 2007).

Substance abuse customers should have a care coordinator either at the health care center or at the most visited substance abuse service unit, who knows him or her in person and managers the care and helps the customer to follow the assigned care plan. Their care should be looked at two levels, which are health care center and substance abuse services, as health care center visits are causes of injuries and poisonings due to substance abuse.

6.5 Super Service Users

A super service user, most likely a male (67 %) aged over 40 years (68 %), has very complex and obscure needs, he is using a lot of different services (6) and he has several other diagnoses (3) in addition to likely other mental and behavioral disorder diagnosis (84 %). He goes to the ER every second month likely due to symptoms, signs, and abnormal clinical and laboratory findings (77 %), as well as injuries, poisonings and other causes (77 %), but he is not seen at the health care centers that often. He is also likely to have for example musculoskeletal system and connective tissue diseases (21 %), infectious diseases (21 %), and respiratory system diseases (20 %). He receives some psychiatric services, but the largest share of costs come from substance abuse services. He is also likely to receive treatment from geriatrics, internal medicine or neurology, among others. As super service users have mediatory MH/SA HCU status persistence (39 %), he might have been a MH/SA HCU also previous year. However, this persistence takes into account only MH/SA persistence, so his severe health problems can also be more long-term problems. Super service users' total annual costs are 2 M€ and there are 75 customers in this segment in total.

Super service users require attention on personal care and service planning. Their care takers must synchronize treatment and share information in order to organize the services accordingly. As super service users have also other health problems in addition to substance abuse and mental health problems, their care should be managed by using case (or care) management framework, which coordinates the effective management of numerous social and health conditions comprehensively for one person (Cohen et al., 2015). It is suggested that these people are assigned a case manager, who is a general practitioner who can also manage patients' physical health care (Belnap et al., 2006).

Super service users could benefit from improved predictive risk modeling and better information sharing. They are in contact with the system very often, so possible interventions are possible if targeted and implemented correctly. Many visits to ER suggests that super service users have urgent and unplanned needs, which may make predicting challenging. Integration of care should be vertical integration of different levels of services, as well as integrated policy-making and management by working across sectors.

Super service users' care is mostly arranged by targeting them as patients with stable chronic conditions or serious but stable disabilities. Depending on their condition, their care can be described as a project, a cure process, or as a care process.

7 Discussion and Conclusions

This study aimed to understand the characteristics, service use and costs of MH/SA HCUs, form relevant segments from the customer population (~5 800 customers), as well as understand how to manage different MH/SA HCU segments. These three goals enable each other, as the understanding of high-cost use enables segmenting, which in turn enables management. All these three goals are relevant for customers who have already become HCUs. Moreover, this study aimed to explore the possibilities of predicting future HCUs and finding risk indicators, which in turn focuses on the future and on the people who have not, and hopefully won't, become MH/SA HCUs. In this section, the results and findings are evaluated and discussed.

7.1 Contribution to the Literature

MH/SA customers have not been studied earlier from this point of view. This study fills the gap by exploring the demographics of this increasing and rather costly population, their service use, costs and indicators of becoming a HCU in the future. This study tackles these topics to achieve better understanding, and thus management, of HCUs, that ultimately leads to better health and cost savings.

Cost Distribution

It is widely acknowledged that a small share of population causes a large share of total social and health care costs in general. Studies in Finland have shown that, for example, in Uusimaa region the costliest 15 % caused 70 % of specialized health care costs (Leskelä et al., 2015), in Oulu the costliest 10 % caused 81 % of social and health care costs (Leskelä et al., 2013), and in Capital Region the costliest 0,1 % caused 4 % of health and elderly care costs (Kapiainen et al., 2010). In this study, the costliest 10 % caused 83 % of all social and health care costs, while the costliest 10 % of MH/SA customers caused 62 % of mental health and substance abuse costs. Actually, 5,5 % of the population of Tampere caused 100 % of MH/SA costs. This study confirms the trend of skewed social and health care costs also in Tampere, as well as confirms the hypothesis that the distribution MH/SA costs are also highly skewed.

Cost distribution was also analyzed across service categories. This revealed that in psychiatry, enhanced psychiatric hospital outpatient visit, detoxification, and geriatric psychiatry HCUs caused the majority of costs but represented only a small share of total customers, i.e. the distribution of the expenditures is skewed. However, in services like enhanced psychiatric hospital care, substance abuse residential care, and long term geriatric psychiatric hospital care majority of customers were HCUs, and thus caused also majority of the costs. This means that some services were burdened by HCU customers who also caused majority of costs, in some services HCUs were only a small share of customer base, but they still caused a majority of costs, and in some services the share of HCUs and their costs were quite equal or nonexistent.

Characteristics of HCUs

Common finding in earlier research is that HCUs are often more sick than rest of the population, for example in terms of number of service categories used (see e.g. Garfinkel, Riley & Iannacchione, 1988; Berk & Monheit, 2001; Leskelä et al., 2013). This study revealed that MH/SA HCUs used 3,9 service categories on average, while non-HCUs used 3,1 on average. HCUs also had 1,3 other diagnosis groups on average (excl. mental health diagnosis), while non-HCUs had 1,0.

Another common finding is that in general, there is no significant differences among females and males in terms of average cost per customer and HCU share (see e.g. Rais et al., 2013). The same finding is clear in this study, as female HCUs' average cost was 21 950 € and males' 22 250 €, while 47 % of HCUs were females. However, when comparing genders by age groups, it is interesting that in the younger age groups females were more represented, after which the amount of HCUs decreased steadily. When it comes to the average cost, it seemed to increase quite dramatically in both genders after the age of 50 years.

MH problems can also cause somatic symptoms, for example headache, fatigue, dizziness, and pain, which in turn may lead to increased outpatient medical visits (Koenke, 2003), i.e. ER visits. This study revealed that on average MH/SA HCUs visited ER 2,5 times, while non-HCUs visited ER only 0,7 times annually.

Service Use

A study from Uusimaa (Leskelä et al., 2015) revealed that mental health problems go often together especially with heavy service use from surgery specialty and internal medicine. In this study, mental health customers' (mood disorder, schizophrenia and other mental health segments) most common other services were internal medicine, neurology and health care center services.

Another study from Oulu (Leskelä et al., 2013) showed that substance abuse service users' annual costs were mainly from elderly care and social care housing (47 %), mental health and psychiatric care (30 %) and specialized somatic health care (14 %). In this case, comparison is challenging as service categorization is done differently, but in this study 86 % of substance abuse segment's annual cost came from psychiatry and substance abuse services (incl. housing), which is comparable with the earlier study's share of 77 % (47 % + 30 %).

The same study (Leskelä et al., 2015) also revealed that psychiatric care patients' annual costs came from mental health and psychiatric care (63 %) and specialized somatic health care (26 %). In this study, psychiatry care patients' (mood disorder, schizophrenia and other mental health segments) annual costs came from psychiatry care (87 %), which is comparable with the earlier study's share of 63 %.

Persistence

Another research indicates that specialized psychiatric care patients are very likely to be HCUs two years in a row (Leskelä et al., 2015). This study revealed that substance abuse, other mental health and schizophrenia segments were most likely to stay HCUs two years in a row, with 42 %, 40 % and 38 % respectively, while mood disorder segment's HCU status seemed to be most varying. Additionally, this study revealed that the more expensive the customer, more likely he or she stayed expensive also during the following year.

Diagnoses

Hypothesis from earlier research (Katon, 2003) was that MH/SA problems are also often accompanied with chronic general illnesses, such as diabetes, heart disease, neurological illnesses and cancer. This study compared these findings between HCUs and non-HCUs and found out that diabetes (+2 %), heart disease (+73 %) and liver disease (+38 %) were more common among HCUs. Neurological illnesses were not studied due to ICD-10 diagnosis grouping.

Another hypothesis (Institute of Medicine US, 2006) suggested that substance abuse problems (alcohol and drugs) are associated with liver disease, immune system disorders, cardiovascular diseases, diabetes, hepatitis C, HIV and hepatitis B. This study compared various segments' diagnosis and found out that liver disease (4,3 % vs. 1,4 %), cardiovascular diseases, or heart disease in detail (2,2 % vs. 1,7 %), and infectious diseases, including hepatitis C and B (10,9 % vs. 6,7 %) were more common among substance abuse segment than HCUs on average. Immune system disorders were not studied due to ICD-10 diagnosis grouping.

Areas

This research analyzed the differences among the areas of Tampere. The analysis revealed that the service use and costs varied between the areas, while the number of HCUs divided by the number of total customers was quite stable. The findings include that customers in south and southwest used relatively most substance abuse services and they also belonged to substance abuse segment more often than the others. On the other hand, customers in central area and northwest used relatively most mental health services. Moreover, customers in northeast used relatively most other services.

Segmentation

The customers were divided into five segments. These segments varied by some important variables and thus formed the basis for HCU management framework. Mood disorder segment and schizophrenia segment were formed based on diagnoses, while other mental health, substance abuse and super service user segments were formed by using cluster analysis and the following variables: number of service categories, detailed service, number of other diagnoses, number of ER visits, MH costs, SA costs, other costs and symptoms, signs and abnormal clinical and laboratory findings diagnosis.

Mood disorder, schizophrenia and substance abuse segments were quite similar in terms of number of service categories used (3,4-3,9), number of other diagnoses (0,8-1,4) and number of ER visits (1,5-2,4). Other mental health segment seemed to be a little less intense with same figures 2,8, 0,4 and 0,7 respectively. Super service user segment was revealed to be the most different segment, which used heavily services with an average of 5,9 service categories, 6,1 ER visits and 2,8 other diagnosis groups.

In terms of costs, mood disorder and schizophrenia segments had similar cost structure with majority of costs coming from psychiatry and only a small share of costs from other services, while schizophrenia had a bit higher total cost. They also had similar other cost structure, while mood disorder segment had more other costs.

Other mental health segment, in turn, differed by having more geriatric psychiatry costs. The segment also didn’t have internal medicine costs like others, and their other costs were by far the lowest.

Substance abuse segment had lowest total cost with 85 % of costs coming from substance abuse services. Their other costs come mainly from health care center services.

Finally, super service user segment is the most interesting, as its average total costs came from many services and 43 % of total costs came from many other services than MH/SA services.

Management of HCUs

Based on the formed segments, managerial suggestions were conducted based on service use and other characteristics of these segments. As MIELI - National Plan for Mental Health and Substance Abuse Work emphasizes, all mental health and substance abuse work should be managed by focusing on client's status, promotion, and prevention. Moreover, the focus is ought to be on basic and outpatient services instead of increasing inpatient and institutional care, and thus substitute them to outpatient services. Also, integrated community care is seen as a way to avoid costly hospitalizations.

Mood disorder and schizophrenia customers' care should be organized based on disease management by organizing the care practices, especially hospital ward, accordingly based on medical guidelines. Additionally, mood disorder condition is somewhat preventable by efficient proactive measures, early stage support and self-management. It can also be cured, and thus the focus of integration is on emerging understanding and adjusting care plan. On the contrary, schizophrenia is a condition, which is somewhat a condition without a cure and possibility of prevention.

Other mental health segment's needs require the focus to be on supporting existing HCUs, at least if the case is on the older geriatric customers, whose condition will probably remain. If the case is the young symptomatic mental health customers, the focus should be on helping them to get the needed services. This might prevent problems and facilitate their wellbeing. This segment clearly should receive focus in the future.

Substance abuse problems are somewhat related to lifestyle choices that increase the likelihood to develop "lifestyle" diseases. Suggested methods to manage this segment include self-management capabilities and lifestyle change support and management. Their care should be managed my using disease management framework. Substance abuse customers could also benefit from predictive measures offered before the problems become worse.

Super service users require a lot of personal and enduring care and service planning, due to fast-changing medical needs. Their care should be managed by using intensive case management framework led and evaluated by a case manager. They could also benefit from improved predictive risk modeling and better information sharing across organizations.

Predicting Future HCUs

This study aimed to explore the possibilities of predicting future HCUs using logistic regression. The result from the model was minor – the model yielded only one variable with significance (total social and health care costs) and a lift of 2,2 for top 10 % of customers with highest predicted probabilities to become a HCU.

To conclude, using this data, predicting future HCUs turned out to be challenging. Nevertheless, if improved predictions for MH/SA problems can be made (by e.g. including more explanatory variables) preventive measures, as interventions can lead to reduced resource utilization and better wellbeing. It is important to highlight the fact that someone becoming a MH/SA HCU depends also on decisions made by the professionals, availability of services, and local, regional and national policies, and not only on their social and health care needs.

7.2 Strengths and Limitations

This study has some limitations as well as strengths. First of all, a clear advantage is the fact that this study includes almost whole social and health care. The data consists of publicly funded social and health care visits in 2013 and 2014, and covers all residents of Tampere and all their visits to services that are both offered and purchased for them at the hospital district level.

However, this study excludes prescription drugs, dental care and social benefits, as well as the service use in private and third sector. Additionally, the data doesn’t include information on deaths and moving to or away from Tampere, what may affect for example persistency findings. Moreover, some service costs had to be normalized which may have caused some cost allocation errors, for example in neurology service line, which on the other hand is a very small service category.

In addition, overall number of MH/SA customers in Tampere is rather small, making the total study population quite small. Additionally, in some cases (e.g. services, areas or diagnoses) the number of customers was very small, which can make the reliability questionable. Statistical significance needs to be assessed in more detail in future research.

In terms of reliability and validity this study does not have any issues. HCUs are estimated based on reliable data, and the data measures the exact things it is purposed to measure. This study can be repeated by other researchers by using the same conditions and generate the same results.

7.3 Suggestions for Future Research

This study sets the base for studying mental and health care customers. However, more profound studies are needed in order to gain even more insight into the topic.

First, customers' service use and episodes should be analyzed in terms time and place, by using for example process mining in order to fully understand the service offering and how often and where various segments are using services. For example, some MH/SA HCUs might get treatment only at one specific place, while someone with similar condition might be receiving same treatment from many different places without care professionals knowing it. Analyzing processes could reveal overlap and inefficiencies in MH/SA customers' service offering.

Second, by including social benefits a lot of important insight could be gained. In terms of data availability, including social benefits should be rather easy. As discussed, MH/SA problems can be very equivocal, and employment, income level and social status can have a serious impact on customer's risk of becoming a MH/SA HCU. Predictive modeling could yield much better results if the data included these kinds of variables.

Third, a further research is needed on the topic of interventions and preventive services, as well as their return on investment. This study included benefit matrix for each outcome, but if these benefits could be replaced by actual monetary values of risk realization and successful intervention, the results could reveal more interesting insight.

Fourth, HCU managerial models need a further investigation taking into account the times and places of service use. Also, other mental health and super service user segments need to be further analyzed.

References

Abidi, S.S.R. (1999) "Applying Data Mining in Healthcare: An Info-Structure for Delivering 'Data-Driven' Strategic Services", *Studies in Health Technology and Informatics*, pp. 453-456.

Allaudeen, N., Schnipper, J.L., Orav, E.J., Wachter, R.M. & Vidyarthi, A.R. (2011) "Inability of providers to predict unplanned readmissions", *Journal of General Internal Medicine*, Vol. 26, No. 7, pp. 771-776.

Ash, A.S, Zhao, Y., Ellis, R.P. & Kramer, M.S. (2001) "Finding Future High-Cost Cases: Comparing Prior Cost versus Diagnosis-Based Methods" *Health Services Research*, Vol. 36, No. 6, pp. 194-206.

Bacher, J., Wenzig, K., & Vogler, M. (2004) "SPSS TwoStep Cluster - A First Evaluation", *Nürnberg, Germany: Lehrstuhl für Soziologie* pp. 578-588.

Banfield, J.D. & Raftery, A.E. (1993) "Model-Based Gaussian and Non-Gaussian Clustering", *Biometrics*, Vol. 49, No. 3, pp. 803-821.

Bardsley, M., Billings, J., Dixon, J., Georghiou, T., Lewis, G.H. & Steventon, A. (2011) "Predicting Who Will Use Intensive Social Care: Case Finding Tools Based on Linked Health and Social Care Data" *Age Ageing*, Vol. 40, No. 2, pp. 265-270.

Belnap, B.H., Kuebler, J., Upshur, C., Kerber, K., Mockrin, D.R., Kilbourne, A.M. & Rollman, B.L. (2006) "Challenges of Implementing Depression Care Management in the Primary Care Setting", *Administration and Policy in Mental Health and Mental Health Services Research*, Vol. 33, No. 1, pp. 65-75.

Berk, M.L. & Monheit, A.C. (2001) "The Concentration of Health Care Expenditures, Revisited" *Health Affairs*, Vol. 20, No. 2, pp. 9-18.

Billings, J. & Mijanovich, T. (2007) "Improving the Management of Care for High-Cost Medicaid Patients" *Health Affairs*, Vol. 26, No. 6, pp. 1643-1655.

Bodenheimer, T., Wagner, E.H. & Grumbach, K. (2002) "Improving Primary Care for Patients with Chronic Illness: the Chronic Care Model, Part 2." *Jama*, Vol. 288, No. 15, pp. 1909-1914.

Callahan, J.J., Shepard, D.S., Beinecke, R.H., Larson, M.J. & Cavanaugh, D. (1995) “Mental Health/Substance Abuse Treatment in Managed Care: The Massachusetts Medicaid Experience”, *Health Affairs,* Vol. 14, No. 3, pp. 173-184.

Calver, J., Brameld, K.J., Preen, D.B., Alexia, S.J., Boldy, D.P. & McCaul, K.A. (2006) “High-Cost Users of Hospital Beds in Western Australia”, *The Medical Journal of Australia*, Vol. 184, No. 8, pp. 393-397.

Charlson, M., Charlson, R.E., Briggs, W. & Hollenberg, J. (2007) “Can Disease Management Target Patients Most Likely to Generate High Costs?”, *Journal of General Internal Medicine,* Vol. 22, No. 4, pp. 464-469.

Chaudhary, K. & Sharma, A. (2013) “Study on Two Step Clustering”, *IJITKM*, Vol. 7, No. 1, pp. 137-139

Cohen, C.J., Flaks-Manov, N., Low, M., Balicer, R.D. & Shadmi, E. (2015) “High-Risk Case Identification for Use in Comprehensive Complex Care Management”, *Population Health Management,* Vol. 18, No. 1, pp. 15-22.

Coons, S.J. (1996) “Disease Management: Definitions and Exploration of Issues”, *Clinical Therapeutics*, Vol. 18, No. 6, pp. 1321-1326.

Cousins, M.S., Shickle. L.M. & Bander, J.A. (2002) ”An Introduction to Predictive Modeling for Disease Management Risk Stratification” *Disease Management*, Vol. 5, No. 3, pp. 157-167.

Cox, D.R. (1958) "The regression analysis of binary sequences (with discussion)”, *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 20, No. 2, pp. 215-242.

Curry, N., Billings, J., Darin, B., Dixon, J., Williams M. & Wennberg, D. (2005) Predictive Risk Project: Literature Review, The King’s Fund, London, 24 p.

Deber, R. & Lam, K.C.K. (2009) “Handling the High Spenders: Implications of the Distribution of Health Expenditures for Financing Health Care”, paper presented at the American Political Science Association Annual Meeting.

Ehrlich, E., Kofke-Egger, H. & Udow-Phillips, M. (2010) “Health Care Cost Drivers: Chronic Disease, Comorbidity, and Health Risk Factors in the U.S. and Michigan”, *Center for Healthcare Research & Transformation – Issue Brief*, No. July

Elissen, A.M.J., Struijs, J.N., Baanc, C.A. & Ruwaarda, D. (2015) “Estimating Community Health Needs against a Triple Aim Background: What Can We Learn from Current Predictive Risk Models?”, *Health Policy*, Vol. 119, No. 5, pp. 672-679.

Fleishman, J.A. & Cohen, J.W. (2010) “Using Information on Clinical Conditions to Predict High-Cost Patients”, *Health Services Research*, Vol. 45, No. 2, pp. 532-552.

Fryers, T., Jenkins, R. & Melzer, D. (2004) Social Inequalities and the Distribution of the Common Mental Disorders, Psychology Press, Hove and New York, 252 p.

Garfinkel, S.A., Riley, G.F., & Iannacchione, V.G. (1988) ”High-Cost Users of Medical Care”, *Health Care Financing Review*, Vol. 9, No. 4, pp. 41-52.

Goodwin, J. (2006) ”Long-Term Conditions: How to Manage Them?”, *Geriatric Medicine-Conventry*, Vol. 36, No. 1, pp. 17-30.

Hosmer Jr, D.W. & Lemeshow, S. (2004) Applied logistic regression, John Wiley & Sons, New Jersey, 392 p.

Institute of Medicine (US). Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders (2006) Improving the quality of health care for mental and substance-use conditions, National Academy Press, Washington, 528 p.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M. & Kripalani, S. (2011) “Risk Prediction Models for Hospital Readmission: A Systematic Review” *Jama*, Vol. 306, No. 15, pp. 1688-1698.

Kapiainen, S., Seppälä, T.T., Häkkinen, U., Lauharanta, J., Roine, R.P. & Korppi-Tommola, M. (2010) Pääkaupunkiseudun erittäin kalliit potilaat, THL Avauksia 3/2010

Katon, W.J. (2003) “Clinical and Health Services Relationships between Major Depression, Depressive Symptoms, and General Medical Illness”, *Biological psychiatry*, Vol. 54, No. 3, pp. 216-226.

Kaufman, L. & Rousseeuw, P.J. (2009) Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, Inc., Hoboken, New Jersey, 349 p.

Kessler, R.C. (2004) “The Epidemiology of Dual Diagnosis”, *Biological Psychiatry*, Vol. 56, No. 10, pp. 730-737.

Kessler, R.C., Nelson, C.B., McGonagle, K.A., Edlund, M.J., Frank, R.G. & Leaf, P.J. (1996) “The Epidemiology of Co-Occurring Addictive and Mental Disorders: Implications for Prevention and Service Utilization”, *American Journal of Orthopsychiatry*, Vol. 66, No. 1, pp. 17-31.

Kozyrskyj, A., Liz, L., Dahl, M. & Soodeen, R-A. (2005) High-Cost Users of Pharmaceuticals: Who Are They?, Manitoba Center for Health Policy, Winnipeg, 53 p.

Kuntaliitto (2015). Online. Available at: www.kunnat.net, [14.12.2015].

Leskelä, R.-L., Komssi, V., Sandström, S., Pikkujämsä, S., Haverinen, A., Olli, S.-L. & Ylitalo-Katajisto, K. (2013) ”Heavy Users of Social and Health Care Services in the City of Oulu”, *Suomen Lääkärilehti*, Vol. 48, No. 68, pp. 3136-3169

Leskelä, R.-L., Silander, K., Komssi, V., Koukkula, L., Soppela, J. & Lehtonen, L. (2015) ”Patients Using Most Specialized Care Services”, *Suomen Lääkärilehti*, Vol. 43, No. 70, pp. 2865-2872.

Lewis, G., Curry, N. & Bardsley, M. (2011) “Choosing a Predictive Risk Model: A Guide for Commissioners in England” *London: The Nuffield Trust*

Lillrank, P. (2012) “Integration and Coordination in Healthcare: An Operations Management View”, *Journal of Integrated Care*, Vol. 20, No. 1, pp. 6-12.

Lillrank, P.P., Groop, J. & Malmström, T.J. (2010) “Demand and Supply–Based Operating Modes — A Framework for Analyzing Health Care Service Production” *Milbank Quarterly*, Vol. 88, No. 4, pp. 595-615.

Lynn, J., Straube, B.M., Bell, K.M., Jencks, S.F. & Kambic, R.T. (2007) *The Milbank Quarterly*, Vol. 85, No. 2, pp. 185-208.

Mukamel, D.B., Chou, C.C., Zimmer, J.G. & Rothenberg, B.M. (1997) “The Effect of Accurate Patient Screening on the Cost-Effectiveness of Case Management Programs”, *The Gerontologist*, Vol. 37, No. 6, pp. 777-784.

Murphy, S.M., Castro, H.K. & Sylvia, M. (2011) “Predictive Modeling in Practice: Improving the Participant Identification Process for Care Management Programs Using Condition-Specific Cut Points." *Population Health Management*, Vol. 14, No. 4, pp. 205-210.

The National Institute for Health and Welfare (2015). Online. Available at: www.thl.fi, [14.12.2015]

Norušis, M.J. (2011) IBM SPSS Statistics 19 Statistical Procedures Companion, Prentice Hall, 464 p.

Panattoni, L.E., Vaithianathan, R., Ashton, T. & Lewis, G.H. (2011) ”Predictive Risk Modelling in Health: Options for New Zealand and Australia”, *Australian Health Review*, Vol. 35, No. 1, pp. 45-51.

Peng, C.Y.J., Lee, K.L. & Ingersoll, G.M. (2002) “An introduction to logistic regression analysis and reporting”, *The journal of educational research*, Vol. 96, No. 1, pp. 3-14.

Rais, S., Nazerian, A., Ardal, S., Chechulin, Y., Bains, N. & Malikov, K. (2013) “High-Cost Users of Ontario’s Healthcare Services”, *Healthcare Policy*, Vol. 9, No. 1, pp. 44-51.

Shenas, S.A.I., Raahemi, B., Tekieh, M.H. & Kuziemsky, C. (2014) “Identifying High-Cost Patients Using Data Mining Techniques and a Small Set of Non-Trivial Attributes” *Computers in biology and medicine*, Vol. 53, pp. 9-18.

Shih, M.Y., Jheng, J.W. & Lai, L.F. (2010) “A two-step method for clustering mixed categroical and numeric data” *淡江理工學刊*, Vol. 13, No. 1, pp. 11-19.

Siegel, S. & Castellan, N.J. (1988) Nonparametric statistics for the behavioral science (2nd ed.), McGraw-Hill, New York, 399 p.

Unützer, J., Harbin, H., Schoenbaum, M. & Druss, B. (2013) ”The Collaborative Care Model: An Approach for Integrating Physical and Mental Health Care in Medicaid Health Homes”, *HEALTH HOME, Information Resource Center*, No. May.

Verma, J.P. (2013) Data Analysis in Management with SPSS Software, Springer, New Delhi, 480 p.

World Health Organization (2008) “Integrated health services – What and Why?”, *WHO Technical Brief*, No. 1.

Appendix A: Service Categories

Some of the medical specialties were assigned based on service unit, which can't be seen in this table.

Original Finnish name	Translated English name	Service category	MH/SA
ALLERGOLOGIA	Allergology	Allergology	
AUDIOLOGIA	Audiology	Audiology	
SYÖPÄTAUDIT JA SÄDEHOITO	Cancer diseases and radiotherapy	Cancer diseases	
SYDÄN- JA RINTAELINKIRURGIA	Cardiac and chest surgery	Cardiology	
KARDIOLOGIA	Cardiology	Cardiology	
Kliininen rasituskoe	Clinical exercise test	Cardiology	
IHO- JA SUKUPUOLITAUDIT	Dermatology and venerel diseases	Dermato-Venereology	
AMMATTI-IHOTAUDIT	Occupational skin diseases	Dermato-Venereology	
SUKUPUOLITAUTIEN VASTAANOTTO	Venerel disease reception	Dermato-Venereology	
Jatkuvatoimintoinen Glukoosin mittaus ho	Continuous glucose measurement	Endocrinology	
Dehko-hoitajat	Diabetes nurse	Endocrinology	
Lääkärikäynti	Doctor visit	Endocrinology	
ENDOKRINOLOGIA	Endocrinology	Endocrinology	
Käynti jalkaterapeutilla	Foot therapist visit	Endocrinology	
Diabetespumput	Insulin pump	Endocrinology	
Hoitajakäynti	Nurse visit	Endocrinology	
Silmä- ja korvayksikkö	Eye and ear unit	Eye and ear diseases	
Barronin ligiatuura	Barron ligature	Gastroenterology	
GASTROENTEROLOGIA	Gastroenterology	Gastroenterology	
Gastroskopia	Gastroscopy	Gastroenterology	
YLEISLÄÄKETIEDE	General medicine	General medicine	
GERIATRINEN PSYKIATRIA	Geriatric psychiatry	Geriatric psychiatry	MH
PITKÄAIK. SH., PSYKOGERIATRINEN SAIRAALA	Long term geriatric psychiatric hospital	Geriatric psychiatry	MH
Yleislääketieteen ja geriatrian palv.lin	General medicine and geriatrics service line	Geriatrics	
NAISTENTAUTIEN OSASTO	Gynaecologic ward	Gynaecology	
NAISTENTAUDIT JA	Gynaecology	Gynaecology	
Gynekologian poliklinikka	Gynaecology outpatient clinic	Gynaecology	
TERVEYSASEMATOIMINTA	Health care center activity	Health care center activity	
AIKUISNEUVONTA	Adult assistance	Health counselling	
Sternaalipunktiot	Bone marrow examination	Hematology	
HEMATOLOGIA	Hematology	Hematology	
VERISUONIKIRURGIA	Vascular surgery	Hematology	
PERUS KOTIHOITO	Basic home care	Home care	
TEHOSTETTU KOTIHOITO	Enhanced home care	Home care	
TILAPÄINEN KOTIHOITO	Temporary home care	Home care	
Kotisairaala	Home hospital	Home hospital	
INFEKTIOSAIRAUDET	Infectious diseases	Infectious diseases	
Kolonoskopia	Colonoscopy	Internal medicine	
SISÄTAUDIT	Internal medicine	Internal medicine	
Sisätautien poliklinikka	Internal medicine outpatient clinic	Internal medicine	

Sisätautien palvelulinja	Internal medicine service line	Internal medicine	
SYNNYTYSVASTAANOTTO	Childbirth reception	Maternal health	
ÄITIYSPOLIKLINIKKA	Maternity clinic	Maternal health	
ÄITIYSULTRA	Maternity ultrasound	Maternal health	
SYNNYTYSOSASTO	Maternity ward	Maternal health	
PERINNÖLLISYYSLÄÄKETIEDE	Medical genetics	Medical genetics	
RYHMÄKÄYNTI PSYK./TERV.AS. PSYK.PALVELUT	Psychologist group visit/Health care center psychologist services	Mental health services	MH
PSYKOL. HOITO/TERVEYSAS. PSYKOL.PALVELUT	Psychologist treatment/Health care center psychologist services	Mental health services	MH
NEFROLOGIA	Nephrology	Nephrology	
Rannekanavan oireyhtymän tutkimus	Carpal tunnel syndrome examination	Neurology	
Neurologinen poliklinikka	Neurologic outpatient clinic	Neurology	
NEUROLOGIA	Neurology	Neurology	
Neurologinen palvelulinja	Neurological service line	Neurology	
NEUROPSYKOLOGINEN TUTKIMUS	Neuropsychologic examination	Neuropsychology	MH
NEUROPSYKOLOGINEN KUNTOUTUS	Neuropsychologic rehabilitation	Neuropsychology	MH
NEUROKIRURGIA	Neurosurgery	Neurosurgery	
RYHMÄKÄYNTI/AIK.VÄESTÖN RAVITSEMUSTERAPI	Nutrition therapist group visit	Nutritional Therapy	
Perus- ja laajakäynti/aikuisten ravi	Nutrition therapist visit	Nutritional Therapy	
TYÖLÄÄKETIEDE JA TYÖTERVEYSHUOLTO	Occupational medicine and occupational health care	Occupational health care	
SILMÄTAUDIT	Eye diseases	Ophthalmology	
LASTENTAUDIT	Children's diseases	Ophthalmology	
KLIININEN HAMMASHOITO	Clinical oral health	Oral and Maxillofacial Diseases	
SUU- JA LEUKAKIRURGIA	Oral and maxillofacial surgery	Oral and Maxillofacial Diseases	
HAMMAS-, SUU JA LEUKASAIRAUDET	Oral, mouth and jaw diseases	Oral and Maxillofacial Diseases	
OIKOMISHOITO	Orthodontics	Oral and Maxillofacial Diseases	
Murtumapoliklinikkatoiminta	Fracture clinic activities	Orthopedics	
ORTOPEDIA	Orthopedics	Orthopedics	
Ortopedian poliklinikka	Orthopedics outpatient clinic	Orthopedics	
KORVA-, NENÄ- JA KURKKUTAUDIT	Ear, nose and throat diseases	Otorhinolaryngology	
AIKUISVÄESTÖN PUHETERAPIA	Adults' speech therapy	Phoniatrics	
FONIATRIA	Phoniatrics	Phoniatrics	
HUOLTO, KORJAUS JA KULJETUS -KÄYNTI	Maintenance, repair and transportation visit	Physical medicine and rehabilitation	
FYSIATRIA	Physiatry	Physical medicine and rehabilitation	
FYSIO&TOIM.TERAPIA 60 MIN SIS. AS.T.PALV	Physical and occupational therapy 60min	Physical medicine and rehabilitation	
APUVÄL. LAINAUS JA PALAUTUSKÄYNTI 30 MIN	Prosthetics borrowing and return visit 30min	Physical medicine and rehabilitation	
APUVÄL. HANK. JA KÄYTÖN OHJAUSK. 90 MIN	Prosthetics introduction training 90min	Physical medicine and rehabilitation	
KUNTOUTUSRYHMÄT	Rehabilitation groups	Physical medicine and rehabilitation	

KONSULT.-, TYÖNOHJ.KÄYNTI JA AS.TUNT.TYÖ	Consulting, supervision and specialist visit	Psychiatry	MH
PÄIVÄSAIRAALAN HOITOPÄIVÄ	Day hospital care day	Psychiatry	MH
Päiväsairaalan poliklinikkakäynti	Day hospital outpatient visit	Psychiatry	MH
PERHEKÄYNTI	Family visit	Psychiatry	MH
RYHMÄKÄYNTI	Group visit	Psychiatry	MH
KOTIKÄYNTI	Home visit	Psychiatry	MH
OIKEUSPSYKIATRIA	Forensic psychiatry	Psychiatry	
PSYKIATRIA	Psychiatry	Psychiatry	MH
UUSINTAKÄYNTI	Re-visit	Psychiatry	MH
PUHELINVASTAANOTTOKÄYNTI	Telephone reception visit	Psychiatry	MH
VAATIVA KÄYNTI	Urgent visit	Psychiatry	MH
KEUHKOSAIRAUDET	Lung diseases	Pulmonology	
Reumapoliklinikka	Rheumatologic outpatient clinic	Rheumatology	
REUMATOLOGIA	Rheumatology	Rheumatology	
LIIKUNTALÄÄKETIEDE	Sports medicine	Sports medicine	
SELVIÄMISHOITOASEMA	Detoxification	Substance abuse services	SA
Avotoiminta	Outpatient activities	Substance abuse services	SA
KATKAISUHOITOASEMA	Rehab center	Substance abuse services	SA
KORVAUSHOITOASIAKAS	Replacement therapy	Substance abuse services	SA
PÄIHDEHUOLLON LAITOSPALVELUT	Substance abuse residential care	Substance abuse services	SA
YLEISKIRURGIA	General surgery	Surgery	
KÄSIKIRURGIA	Hand surgery	Surgery	
PLASTIIKKAKIRURGIA	Plastic surgery	Surgery	
KIRURGIA	Surgery	Surgery	
Kirurginen poliklinikka	Surgical outpatient clinic	Surgery	
Työttömät	Uneployed	Unemployed health check	
UROLOGIA	Urology	Urology	
Urologian poliklinikka	Urology outpatient clinic	Urology	

Appendix B: Mental Health and Substance Abuse Service Units in Tampere

Mental health services	
	ALL HEALTH CARE CENTERS
Geriatric psychiatry	
	AKUUTTIPSYKIATRIAN OSASTO
	HATANPÄÄN PUISTOSAIRAALA K03, K04, K12
	PSYKOGERIAT.PKL, PITKÄNIEMI
	PSYKOGERIATRIAN OSASTO
	PSYKOGERIATRIAN TOIMENPIDEPKL
Neuropsychology	
	HATANPÄÄN PUISTOSAIRAALA
Psychiatry	
	AKUUTTIPSYKIATRIAN OSASTO
	AKUUTTIPSYKIATRIAN PKL
	HALLITUSKADUN PSYKIATRIA
	HÄMEENKYRÖN PSYKIATRIAN PKL
	IKAALISTEN PSYKIATRIAN PKL
	KAIVANNON SAIR/PSYKIATRIAN OS
	LIIKKUVA PSYKIATRINEN TYÖRYHMÄ
	MIELENTERVEYSPALVELUT
	NEUROPSYKIATRIAN PKL
	ORIVEDEN PSYKIATRIAN PKL
	PARKANON PSYKIATRINEN PKL
	PITKÄNIEMEN SAIRAALA
	PSYKIATRIAN PKL MAAHANMUUTTAJI
	PSYKIATRIAN PKL, AIKUISPSYKIAT
	PSYKOGERIAT.PKL, PITKÄNIEMI
	PSYKOGERIATRIAN OSASTO
	PSYKOGERIATRIAN TOIMENPIDEPKL
	PÄIHDEPSYKIATRIAN PKL
	PÄIHDERASKAUS, VAUVAPERHETYÖR.
	PÄIVÄSAIRAALA 1, 2, 3
	SARVIKSEN PSYKIATRIA
	TIPOTIEN PSYKIATRIA
	TRANS-POLIKLINIKKA
	TRE.KAUP.AKUUTTIPSYKIATRIANPKL
	YL.SAIRAALAPSYKIATRIAN OSASTO
	YLEISSAIRAALAPSYKIATRINEN PKL
Detoxification	
	SELVIÄMISHOITOASEMA
Rehab	
	KATKAISUHOITOASEMA
Replacement therapy	
	HATANPÄÄN TERVEYSASEMA
	HERVANNAN TERVEYSASEMA
	KAUKAJÄRVEN TERVEYSASEMA
	LIELAHDEN TERVEYSASEMA
	LINNAINMAAN TERVEYSASEMA
	PÄIVÄPERHON KRIISI-JA KATKO, AVOTYÖ KOHT.PAIKKA, KORVAUSHOITO, PERHEKUNTOUTUSOS, ÄITIYS-LASTENNLA
	TAMMELAKESKUKSEN TERVEYSASEMA

	TESOMAN TERVEYSASEMA
	TIPOTIEN TERVEYSASEMA
Substance abuse residential care	
	PALHOHIEMEN HUOLTOKOTI

Appendix C: Grouping of postal codes

Postal code	Area
33100	Central
33180	Central
33200	Central
33210	Central
33230	Central
33500	Central
33520	Central
33530	Central
33540	Central
33800	Central
33560	Northeast
33580	Northeast
33610	Northeast
33700	Northeast
33730	Northeast
33400	Northwest
33410	Northwest
33680	Northwest
34240	Northwest
34260	Northwest
34270	Northwest
33820	South
33840	South
33850	South
33870	South
33900	South
33710	Southeast
33720	Southeast
33721	Southeast
33240	Southwest
33250	Southwest
33270	Southwest
33300	Southwest
33310	Southwest
33320	Southwest
33330	Southwest
33340	Southwest
33420	Southwest

Appendix D: Distribution of MH/SA Customers' Costs by Cost Groups, Service Categories and Detailed Services, 2013

Distribution of MH/SA customers' MH/SA and other costs, 2013

Cost groups	n	Total MH/SA costs (M€)	Avg. MH/SA cost (€)	% of costs	Avg. other costs (€)	Avg. total costs (€)
Top 1 (costliest 1 %)	58	3,8	64 834	19 %	5 742	70 576
Top 5 (costliest 5 %)	289	9,3	32 119	48 %	4 196	36 315
Top 10 (costliest 10 %)	577	12,4	21 559	64 %	3 892	25 450
Bottom 90 (least costly 90 %)	5 194	7,0	1 348	36 %	2 016	3 364
Total	**5 771**	**19,4**	**3 369**	**100 %**	**2 204**	**5 573**

Service categories and detailed services, share and costs of HCUs, and average and total costs, 2013

Service	n	Share of HCUs	Cost of HCUs	Avg. Cost per customer (€)	Avg. Cost per HCU (€)	Total cost (€)
Psychiatry	**4 738**	**11 %**	**67 %**	**2 957**	**18 346**	**14 008 563**
Psychiatry	1 074	35 %	85 %	7 179	17 181	7 710 427
Re-visit	3 065	9 %	11 %	656	805	2 010 987
Enhanced psychiatric hosp. care day	154	82 %	95 %	8 370	9 660	1 289 004
Enhanced psychiatric hosp. outpatient visit	311	46 %	86 %	3 218	6 027	1 000 692
Urgent visit	2 232	12 %	15 %	285	372	637 188
Home visit	382	24 %	29 %	1 688	2 110	645 003
Telephone reception visit	2 373	12 %	22 %	108	195	256 298
Consulting, supervision and specialist visit	611	18 %	21 %	165	185	100 979
Group visit	399	13 %	10 %	381	298	151 947
Family visit	391	17 %	17 %	284	280	110 887
Substance abuse services	**885**	**17 %**	**55 %**	**4 834**	**15 676**	**4 278 472**
Detoxification	584	16 %	57 %	4 557	16 046	2 661 126
Replacement therapy	141	25 %	29 %	5 373	6 329	757 639
Substance abuse residential care	32	84 %	97 %	12 278	14 059	392 905
Rehab center	202	36 %	50 %	2 111	2 900	426 490
Outpatient activities	331	27 %	36 %	122	163	40 312
Geriatric psychiatry	**66**	**48 %**	**95 %**	**12 063**	**23 641**	**796 180**
Long term geriatric psychiatric hospital	15	87 %	99 %	39 090	44 477	586 351
Geriatric psychiatry	52	38 %	85 %	4 035	8 915	209 829
Mental health services	**436**	**2 %**	**2 %**	**627**	**687**	**273 371**
Psychologist treatment/Health care center psychologist services	434	2 %	2 %	604	687	262 319
Psychologist group visit/Health care center psychologist services	8	0 %	0 %	1 382	0	11 052
Neuropsychology	**119**	**5 %**	**4 %**	**707**	**494**	**84 124**
Neuropsychological examination	111	5 %	5 %	476	408	52 861
Neuropsychological rehabilitation	16	6 %	2 %	1 954	513	31 264
Total	**5 771**	**10 %**	**64 %**	**3 369**	**21 559**	**19 440 712**

Appendix E: Input Variables for Logistic Regression

Demographics and other:
Sex
Age group
Area
Number of service categories used
Number of other diagnosis received
Number of ER visits

Diagnoses:
Schizophrenia
Mood disorder
Other mental and behavioral…
Congenital malformation and…
Respiratory system disease
Nervous system disease
Skin disease
Tumor
Ear disease
Symptoms, signs and…
Pregnancy, childbirth and…
Digestive system disease
Eye disease
Infectious disease
Musculoskeletal system and…
Endocrine, nutritional and…
Injury, poisoning and…
Circulatory system disease
Blood and blood-forming…
Genitourinary system disease
HIV disease
Diabetes
Heart disease
Liver disease

Costs:
Total cost
MH/SA cost
Other cost