BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Climbing the ladder of knowledge about drug shortages through business analytics

Navn:	Mustafe Hussein Jama, Stian Taalesen

Start:	15.01.2021 09.00
Finish:	01.07.2021 12.00

Climbing the ladder of knowledge about drug shortages through business analytics

Hand-in date:
01.07.2021

Campus:
BI Oslo

Supervisor:
Kim van Oorschot

Examination code:
GRA 19703 – Master Thesis

Program:
Master of Science in Business Analytics

Acknowledgements

This thesis marks the conclusion of our five years as students at BI Norwegian Business school and the Master of Science program in Business Analytics. As such, we would like to direct our praise to the many professors who have taught us so much througout our education.

First and foremost, we would like to express our gratidude towards our supervisor, Kim van Oorschot. Her knowledge, support, availability and continous feedback has been extremely valuable. We would like to thank her for introducing us to the field of System Dynamics and the topic of drug shortages as a potential subject for the thesis. Her treasured insights and time invested in us is greatly appreciated.

We would also like to thank our families and friends for their their support and encouragement throughout the process. A special thanks to the guys of WT64 for your challenging discussions, motivation and feedback.

Executive summary

Despite techonological tranformations and innovations, the global supply of drugs has faced increasing difficulties to meet its demand. The resulting deviations have lead to an exponential increase in the number of annualy reported drug shortages. In the absense of sufficient global coordination, national policymakers are currently leading the fight against these deficits through domestic stategies and decision-making.

The purpose of this thesis is to examine the information currently available to Norwegian health policymakers with regards to drug shortages, and research how practises from the field of business analytics can be applied to assist in their decision-making. Bohn's (1994) framework for measurement of technological knowledge is applied to evaluate the current knowlegde level with respect to drug shortages, as well as the projected progress following the introduction of analytical methods and perspectives. The source of data for our analysis is the annualy collected and published reports from Statens Legemiddelverk (SLV), which is the official platform for reporting of drug shortages in Norway. Through comprehensive data cleansing and in-depth evaluation of the data quality, we have identified key characteriztics and shortcomings with regards to standarization in the reporting process, which prevents further analysis and subsequent advancements of knowledge.

Our research suggests that the knowledge level regarding drug shortages in Norway remains at quite a low level, thus possessing great potential for progression. Our findings imply that higher knowledge levels can be reached through application of analytical methods, as the first steps towards realization is the appropriate facilitation of such practice. We propose changes to the current reporting system and subsequent data management to counteract the lack of standardization. As such, we believe that the true value of business analytical practices, to aid decision-making regarding drug shortages, will be unlocked when the required foundation is in place.

Table of Contents

I. Introduction

In this section we will present the background information on our research topic, establish and justify our research question and define the structure of the thesis.

1.1 Background

The past decades have seen incredible advancements in the fields of science and technology. The business environment is rapidly adjusting to the technological progress, and thus the pace of change is greater than it has ever been before (Todnem By, 2005). The increased availability of data, combined with scientifical advancements, has resulted in more relevant and powerful solutions in many important areas, including health and wellness. In fact, healthcare, biotechnology, and pharmaceuticals rank among the industries facing most digital disruptions (Accenture, 2017). Despite revolutionizing improvements in related fields, drug shortages have become a global area of interest due to the steep increase in reported instances since the turn of the twenty-first century.

Drug shortages can be defined as situations in which a current or projected demand of a medicine is inadequately met (Bogaert et al., 2015). As a result of the increased number of shortages reported, the World Health Organization (WHO) has described drug shortages as a complex global challenge (Gray & Manasse, 2012). Drug shortages pose a significant threat to public health and safety. Shortages may delay or prevent necessary treatment to patients, resulting in a potential loss in medical care. As a result of shortages, the proceeding efforts could lead to increased risk of medication errors or to prescribers using medicinal options, which can be less effective or poses additional risk (FDA, 2019).

As endorsed by the WHO, member states have begun to implement legal frameworks and national guidelines as mitigation strategies regarding shortages (Bocquet et al., 2017). Since 2016, the WHO has urged member states to establish a best practice for procurement, distribution and contract management for medicines and vaccines (WHO, 2016). While national tracking of drug shortages has been ongoing in most industrialized countries for the better part of the past decades, there currently exists no international standard in which such shortages are reported. This has subsequently made international comparisons challenging, which in return has led to limited literature in the field (De Weerdt et al., 2017).

Research providing evidence regarding the success of counter measures is also scarce. The lack of a uniform definition for drug shortages in Europe is hampering the process of identifying the preferable way of reporting them (De Weerdt et al., 2015).

Among the emerging technologies from the Information Age is the use of analytics in decision support systems. The phenomenon of business analytics refers to the process of leveraging value from collected data, in which the resulting analysis enables decision-making (Acito & Khatri, 2014). Raghupathi and Raghupathi (2014) discuss the challenges of implementing data analytics in health care and conclude that establishing standards and governance rank among the most pressing issues. International standards are a way of dealing with externalities, and externalities occurs whenever one actor's conduct affects the well-being of another (Abbott & Snidal, 2001). As such, the lack of an international standard, with regards to the reporting and definition of drug shortages, may therefore be seen as a disadvantage in the decision-making in the global fight against it.

Braa & Sahay (2012) discuss standardization with respect to health information architecture. They suggest that the use of technical standards is fundamental to integration and interoperability, and that alternative solutions easily get too complex. Furthermore, they propose that in terms of the levels above the technical standards, at the level of the data standards, the only alternative to shared standards is chaos. In the context of drug shortages, this observation is validated when attempting to compare statistics from one country to another. While direct observations are possible, such as making the comparison *"country A's total number of reported drug shortages is X, which is twice as many as country B's number of Y"*, the disparity in terms of definitions, perspective, interpretations, and nuances invalidates most comparisons.

The interest regarding the situation of drug shortages has increased exponentially since it was declared a global challenge in 2012. However, due to the recency of the concern, there is still a scarcity of international literature on the topic. The majority of the literature on the subject of drug shortages focuses on the alleged causes, and how these causes can be mitigated (e.g., Ventola, 2011; Gatesman &

Smith, 2011; De Weerdt et al., 2015). Although alternative options and mitigation measures have been proposed, the annual number of recorded drug shortages has been steadily increasing. This thesis seeks to contribute to the existing body of literature by reviewing the current system in place for reporting of drug shortages, and suggesting changes where weaknesses are detected. From the perspective of business analytics, we will evaluate whether the introduction of digital tools to facilitate analysis may support future decision-making with regards to strategies for mitigation.

1.2 Research question and aim

The lack of an international standard regarding the reporting of drug shortages has led to a massive variety in terms of structure, quality, and thoroughness in the collected data. There is currently no globally accepted definition for the term "drug shortages", which further complicates the efforts to unite international reporting. Hence, the required consistency of data for relevant international comparisons is not present. Inadequate analysis of reported drug shortages indicates that policymakers could be operating at a suboptimal level with regards to their decision-making. The very nature of reporting and record-keeping is constantly generating data, of which we believe reported drug shortages may contain valuable information that is yet to be extracted. To evaluate the potential of this data and the value it could offer health policymakers, we will apply business analytics methods on reported drug shortages.

In the effort to legitimize our analysis, we see it as relevant to conduct our research on a country-specific scale, in line with the existing literature. The lack of an international consensus with regards to key definitions and standardization further validates the reasoning of restraining the research scope to a specific country. Hence, the choice for the scope of this thesis is reported drug shortages in Norway. As such, the overall objective of our thesis is not to generalize our findings on behalf of the global industry, but rather to analyze the reported cases of drug shortages in Norway.

The Norwegian Medicines Agency (*Statens Legemiddelverk)* publishes annual reports regarding drug shortages in Norway. Thus, these reports will form the basis for our thesis. The intent of the study is to research the current level of

knowledge regarding drug shortages in Norway and explore the application of standardization and digital tools in the data collection and subsequent data analysis. Our research will examine whether use of business analytical methods to the data will uncover information that can be of use in health policymaking and in the process of establishing the required standards. This led us the following research question for this thesis, of which we will seek to answer:

"How can business analytics methods be applied to help health policymakers in Norway in the fight against drug shortages?"

In the field of business analytics, the technological advancements made in recent time have brought an abundance of digital tools to ease the process between data collection and application. The digitalization has resulted in significant improvements of the insights for organizations in all types of industries. However, one of the major gaps that remain between relevant analytics and an organization's strategic needs, is the proper collection and transforming of the appropriate data (Kohavi et al., 2002). As such, we will examine the existing lists of drug shortages in Norway, how they are compiled and the quality of the collected data. The objective of the thesis is to evaluate this process, and how application of business analytics methods would impact the current system. To address the scientific gaps regarding reporting of drug shortages, the paper offers two main contributions: How optimization of reporting will increase knowledge, and the link between knowledge-growth and decision-making.

1.3 Thesis structure

This master thesis consists of a total of seven chapters. Following this introduction, a literature review is provided, examining the literary evolution of the main topics of the study. The ensuing chapter discusses the methodology of the thesis, assessing the research design, the data collection, and subsequent data cleansing process. Following the data preparation is the data analysis, executed through the application of a cluster analysis. The fifth chapter consists of a discussion regarding some of the main findings, including a review of the existing system applied for reporting of drug shortages in Norway and the subsequent administration of the reports. Concepts from the field of system dynamics is applied to illustrate the causal relationships between the many variables in the

process. The findings from our research are subsequently applied in the process of recommending changes to the current system. Following the discussion, the next chapter evaluates the potential implications of our recommendations. For the suggested changes to the current reporting system, we consider the practical implications of introducing digital reporting tools and data management systems. Conclusively, we review the theoretical implications of our recommendations, followed by the limitations of our study and suggested future research, before we close the thesis with a conclusion.

II. Literature Review

2.1 Defining drug shortages

Among the academic literature, various definitions and criteria are used to define the term "drug shortages". A 2015 study of underlying problems regarding drug shortages in Belgium and France, revealed that great diversity exists regarding the definition, as well as in the opinions regarding at which level of the supply chain the shortages should be assessed (Bogaert et al., 2015). We believe that the discrepancies in the different interpretations may be part of the reason as to why there is scarce literature of advanced international comparisons. Acosta et al. (2019) claims that even the 2018 report regarding global medicine shortages from WHO's Director-General included a variety of terms such as "shortage", "scarcity" and "stock-outs" in the different translations of the report, reiterating the need for common terminologies.

A cross-sectional survey study of 28 European countries investigated the general characteristics of, and alertness to drug shortage, as well as the information systems in place to capture them and the associated national regulations (Bochenek, et al., 2018). The findings revealed significant variation with regards to the way drug shortages are defined, depending on whether they relate to supply problems or actual drug shortages, permanent or temporal discontinuations, affected disease classes and time frame. The results are consistent with the verdict of De Weerdt et al. (2015), in that drug shortages generally can be expressed in four different ways: demand focused, supply focused, delivery impact or patient availability. Even within the borders of a relatively small country such as Norway, similar disparities occur. Interviews with Norwegian stakeholders reveal great

variation with regards to the definition of drug shortages, depending on who you are asking (Jahre et al., 2021).

Countries such as Belgium and France have introduced legal definitions of drug shortages, of which both refer to a specific number of days of unavailability, for a medicine to be reported as in shortage (De Weerdt et al., 2015). The verdict is however slightly different across the two countries, whereas the law in France refer to a supply disruption of 72 hours or less, and the Belgian law to 96 hours. Other definitions, such as the one proposed by the EMA in 2014, is more all-encompassing, suggesting that a drug shortage is defined as: *"When the delivery of a medicine cannot comply to the need of the patients, whether this is local, national or international"*. The EMA has since its publication removed the suggested definition, and is currently referring to a definition agreed upon by themselves and the Heads of Medicines Agencies: *"A shortage of a medicinal product for human or veterinary use occurs when supply does not meet demand at a national level"* (EMA-HMA, 2019).

2.2 About drug shortages (causes and consequences)

Causes and consequences of drug shortages are closely related and often difficult to distinguish. The complicated link between the two is part of the reason as to why the problem is so complex and hard to tackle. While previously reported as a concern in countries such as the USA and Australia, most literature regarding drug shortages refer to Gray & Manasse's (2012) bulletin of the World Health Organization as one of the first mentions of it reported as a global challenge. They suggested that there is a wide range of causes for medicine supply shortages. One of the main reasons discussed in this report are manufacturing and quality problems. The arguments are supported with studies from the USA, pointing to an investigation conducted by the Food and Drugs administration. This report indicated that 43% of the shortages investigated were attributed to manufacturing quality problems (FDA, 2011). Weerdt et al. (2015) suggest that manufacturing issues or compliance problems have resulted in several public health crises in Europe. Quality assurance for pharmaceutical distribution in the EU is regulated by the rules of good manufacturer practices (GMP) and other strict requirements. In cases in which audits reveal a violation of GMP, production processes may be ceased until the issue has been resolved (De Weerdt et al., 2015).

Gatesman & Smith (2011) suggest that the main cause of drug shortages is economic, referring to how decreasing profits for a manufacturer will lead to a stop in their production of generic drugs. This coincides with the finding of Tucker et al. (2020) scoping review of literature from "The Drug Shortage Era" of 2001-2019, which found that 64 of the total 112 papers regarding causes of drug shortages reported economics as one of the causes.

Some of the most susceptible forms of medicine to experience drug shortages are the so-called generic medications (Johnson, 2011). Generic drugs are medications that are created to be the same as an already existing and marketed brand-name drug, to provide the same clinical benefits (FDA, 2021). Due to the monopolistic economic incentives for manufacturers to produce patent-protected medications, only a few manufacturers will likely be producing an off-patent drug at any time (The Lancet, 2011). The corresponding loss of incentive to continue production once a medication is off-patented, may lead to companies discontinuing production of trade-named drugs (Jenks, 2011). The resulting stand-off between manufacturers chasing profit margins creates a system that is vulnerable to potential drug shortages.

Ventola (2011) discusses the need for an advanced warning system in place to prepare for impending drug shortages and refers to the lack of such a system as one of the main causes of associated problems. Ventola's research of drug shortages in the United States suggests that, while most reported shortages fall into one of five categories, more than 50% of all reported shortages in 2011 were classified as being due to "unknown" causes. While accurate prediction and sufficient preparation for every drug shortage would in practice be impossible, Ventola suggests that careful planning could prevent the consequential problems, from turning into a crisis. The elements required for the necessary planning involves appropriate information-gathering and timely communication. Similarly, Fox, et al. (2009) examines guidelines for management of drug shortages and concludes that proper planning is optimal for minimization of consequences. Among the critical success factors is the effectiveness of the information gathering. This research emphasizes the importance of the collection of information with regards to drug shortages, suggesting that the quality of the

collected data plays an important role in the planning process for prevention and mitigation.

Increased interest in countries which historically have not been exposed to the global market of pharmaceuticals, has resulted in increased demand for active pharmaceutical ingredients (API). The resulting changes in the API consumption may be affecting the relationship between production and medicine availability (WHO, 2015). In a report regarding medicine shortages from 2015, the WHO stated that some high-income countries were reporting shortages for situations in which hospitals or purchasers were unable or unwilling to pay, although the medicine in question was technically available. While these situations are no indication of supply shortages for the drug, the procurement-related shortcomings as a result of financial problems are still reported as drug shortages. Khan (2019) discuss causes of drug shortages in low- and middle-income countries (LMIC) and similarly concludes that many reports come as a result of inadequate financing.

The International Federation of Pharmacists' summit in 2013 summarized the causes for drug shortages as *"... several and multidimensional, in the context of a complex global supply chain"* (Besancon & Chaar, 2013). Among the unfortunate side-effects of drug shortages, is the fact that the increased demand for substitute products may result in shortages for the alternative products as well. A fundamental principle in Europe, specifically for members of the European Union or EEA, is the free circulation of goods and services. In accordance with this principle, medicines can be sold or purchased across national borders, resulting in the phenomena of parallel export and import (EU Pharma Ltd., u.d.). Parallel exports are sales of medicines to other countries, which are distributed in the domestic market of the parallel importer. In many cases where a country has experienced a shortage for a medicine they have been exporting, parallel trading has been reported as causes of the shortage. De Weerdt et al. (2015) and Forrester & Dawes (2008) both mention parallel trade as a potential cause of drug shortages, while (Aguiar & Ernest, 2020) argues that this link is not proven and based on unreliable sources. Regardless of the actual link between the two phenomena, parallel trade has been and is still a frequently reported cause of drug shortages.

2.3 Digital Reporting

Business analytics refers to the process of supporting decision-making using available data. The process is enabled through a complex composition of various applications, techniques, technologies, and systems (Chen, Chiang, & Storey, 2012). The technological revolution, resulting from increased levels of automation, new systems and increased digitalization, has enabled the use of new technologies for a number of industries (Lasi et al., 2014). Valentinetti & Muñoz (2021) and Madakam et al. (2015) discuss the emerging technological concept of Internet of Things (IoT), in which the aim of the concept is to *"... unify everything in our world under a common infrastructure, giving us not only control of things around us, but also keeping us informed of the state of the things"*. Concept such as Big Data and IoT represent exceptional opportunities for insight, resulting in organizations reshaping management and business strategies through digitalization (Muljani & Ellitan, 2019). Büyüközkan & Göçer (2018) contribute to the discussion regarding the potential benefits of digitalization, including automation to improve collaboration and enablement of analytical technologies.

The literature regarding the process of reporting has consistently emphasized the importance of the format of the reporting. Rohrmann (1986) argues that the format in which information is presented and reported, is viewed as a technology that can assist decision-makers. More recent literature is predominantly focused on the shift from physical to digital reporting, and the subsequent requirements. In their article regarding the progress on digital reporting, ICAEW (2004) identified two levels of digital reporting: The first level refers to how digital publication of existing reports increase accessibility, as well as the efficiency with regards to disseminating reports. The second level refers to the required standardization of the format in which the information is registered, to facilitate the process of analysis and exchange of information (Bonsón & Escobar, 2006).

In 2019, the Norwegian Directorate of Health published a report of *'assessments and recommendations'* regarding the National drug preparedness. Among the recommended measures for improvement was the process of further digitalization (Helsedirektoratet, 2019). The report evaluates the current system in place for reporting of drug shortages and the considered weaknesses of the current process

that could be mitigated through the suggested measures. The current reporting system does not include any consideration of the severity of reported shortages, or recommendations with regards to replacements for the drugs subject to shortages. Apart from the yearly reports published by SLV[1], information about on-going drug shortages is to a great extent inaccessible. The proposed measure for this problem is to improve the functionality with regards to reporting and publication of drug shortages for SLV. The report briefly discusses the possibilities of creating a portal for reporting of drug shortages, in which the reports will automatically become part of a database but fails to include any details regarding the functionality or format of such a portal. Consistent to the potential benefits of digitalization discussed by Büyüközkan & Göçer (2018), the findings of the report suggest that improved digital reporting would facilitate partial automatic analysis and categorization, as well as establish a foundation that in the future could be used for some sort of automatic 'warning-system' for affected end users, similar to the solution suggested by Ventola (2011).

2.4 The need for standardization

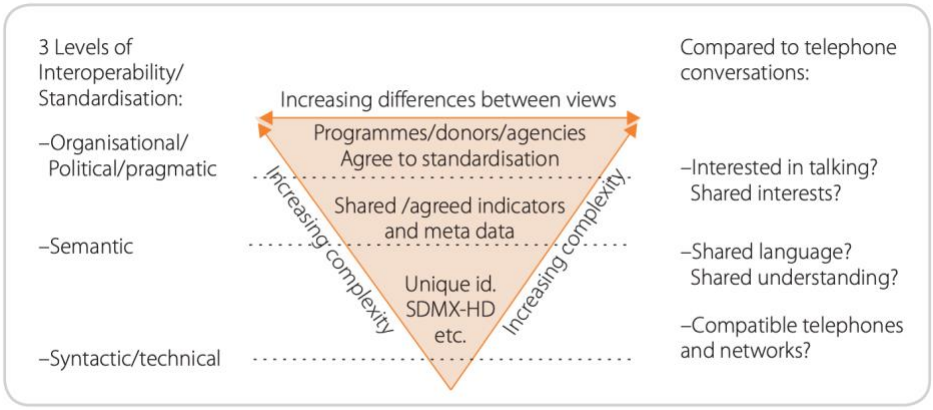
A common theme in the literature regarding drug shortages in Europe is the lack of current standardization. Pauwel et al. (2014) suggest that in Europe, the origins of the occurring drug shortages are underreported by the national health authorities. They conclude that a general reporting template could contribute to better insight into the causes of the shortages and provide fundamental mitigation solutions. The article argues that while drug shortages have been extensively studied in the USA, the issue is understudied in Europe, and suggests that Europe's lack of standardized reporting system is one of the main reasons.

Weerdt et al. (2015) refer to how standardized reporting templates could potentially be implemented in the European Medicines Agency (EMA)'s existing centralized database, but that currently "*…only drugs which are in shortage at the same time in several European Member States are included in this database".*

[1] *The Norwegian Medicines Agency (NOMA),* often referred to as *Legemiddelverket or SLV*, is an agency under the Ministry of Health and Care Services. Their mission includes "safeguarding public and animal health by ensuring the efficacy, quality and safety of medicines and to administer and enforce the medical devices regulation."

With regards to how a standardization is implemented, Braa & Sahay (2012) discuss the core concepts of standards, as well as the process of standardization, and describes three levels of standardization as depicted in Figure 1.

Figure 1 – Three levels of standardisation of the increasing differences and complexities (Braa & Sahay, 2012)



The lowest and most comprehendible of the levels is the "syntactic/technical" level. This is the process of agreeing on a shared "grammar" or terminology. The second level is the "semantic". This is the level of data, data dictionaries and metadata. The objective of this level is to reach a shared meaning and understanding among the users that will by applying the standard. The final level is the "organizational/political". This is the level of decision-making, with the authority to decide the data and indicator standards. In the process of potentially implementing a common standard for reporting of drug shortages in Europe, the organizational/political level would likely be an organization such as the EMA. A centralized body with the authority to decide the semantic level of a standard would allow European member states to begin an essential part of Braa & Sahay's suggested framework: the need for iterative cycling through the levels, through prototyping. They propose that for standards, prototyping is essential to uncover consequences, allow integration, reveal needed adjustments, and gradually solve differences in understanding. One of the most important aspects of the framework is the rising complexity of the levels, due to increasing differences between the views of the decision-makers. An international organ such as the EMA provides an illustration of the intricacy of deciding a unified standard. The agency's Management Board consists of 37 representatives, including one from each of the 27 member states and, observers from Iceland, Liechtenstein, and Norway (EMA, 2021). Thus, the potential deciding organ of the standards consists of individuals

that may have defined drug shortages quite differently in their respective countries. As such, discussions with the interest of unifying a standard will likely include views self-interest and international politics. The absent of this international coordination highlights the challenge we are facing. In the current state of the European market, most countries have started to develop their own standard in terms of drug shortage reporting (Bochenek, et al., 2018). In the process of making the fight against drug shortages international or continental, Europe and one of its organizational bodies should be the driver for standardization of the reporting.

Many scholars have attempted to measure knowledge from different perspectives (Roos & Roos, 1997). One such measure includes Bohn (1994) framework for measuring and understanding technological knowledge. Bohn presents a scale for measuring knowledge about a process that consists of eight stages. The stages are described as follows:

Stage One – Complete ignorance
Unaware of the existence of the phenomenon, or if aware, there is no knowledge of the relevancy between the phenomenon and your process.
Stage Two – Awareness
Awareness that the phenomenon exists and that it may have relevancy to your process, but there is no way to use the variables in the process.
Stage Three – Measure
Able to measure the variables through development and installation of specific instrumentation, but the variables cannot be controlled. If variables are of enough importance, the process may be altered in response to the variables.
Stage Four – Control of the mean
Knowledge of how to control the variables, but the control is not necessarily precise. Able to control the variables at their mean level, but there is some variation of that level.
Stage Five – Process capability
Variables can be controlled with precision across a range of values. Allows for consistent process, although quality variation may still occur.
Stage Six – Process characterization

Knowledge regarding how the variable will affect results. Enables fine-tuning of the process to reduce costs or change product characteristics.

Stage Seven – Know why

A scientific model of the process, including nonlinear and interaction effects of this variable with other variables, which allows for optimization.

Stage Eight – Complete Knowledge

The complete functional form and parameter values that determine the results is known. In practice never reached as it would require knowledge of all interactions among variables but can be approached asymptotically by studying the process in more and more detail.

The framework presented by Bohn was originally designed as a method of measuring technological knowledge with regards to the production processes. However, the framework translates well to other technological processes, such as the management of drug supply chains, as it encapsulates the benefits of increased knowledge. As such, the framework can be applied to measure the current and desired knowledge level about a process of interest. With respect to drug shortages, it appears that Norway and even the global knowledge level about the process is still at a relatively low level. Considering the extent to which countries are reporting annual drug shortages it appears that we are at least at stage two, **"Awareness"**. However, as there appears to be little or no control of the associated variables in the process, it would be erroneous to suggests a knowledge level at which we are able to **"control the mean"**, stage four. As such, this indicates that, according to the framework presented by Bohn, we are currently at knowledge level three, "**Measure"**, or lower. This correlates well with our initial assumption, that although data is being collected or measured, associated variables are currently treated as if they are incontrollable.

Awareness of the current knowledge level is perhaps not very significant in isolation. However, it is highly beneficial in the process of achieving the desired knowledge level, as it provides insight about the discrepancy between the levels. In terms of mitigation of drug shortages, the required planning discussed by Ventola (2011) and Fox, et al. (2009) suggests that the knowledge level should be sufficient to understand how variables will affect results. With regards to Bohn's framework, this would indicate that the required knowledge level should be at

stage six, although ideally even higher. Assuming that the current knowledge level is approximately at stage three, this would suggest that there is still a big gap between what is currently known and what should be known.

The research question for this thesis is: ***"How can business analytics methods be applied to help health policymakers in Norway in the fight against drug shortages?"***. In terms of the knowledge framework, this would suggest that business analytics methods should be applied to increase the knowledge level to the needed stage.

III. Methodology

Our methodological approach is rooted in the research question of the thesis. The overarching goal of the research will be to test if application of business analytics methods will result in valuable conclusions, that could assist policymakers in their decision-making process. The chosen mix of analytic methods will depend on the choice of data for the application. The required level of data preparation before a sufficient level of analysis can be completed will provide valuable indications of potential improvements, with regards to future data collection.

For this section we will clarify the methodological procedure of our research process, and explain the actions performed with regards to data collection and the subsequent data cleansing. Firstly, we will discuss the reasoning behind our quantitative research approach. Further, we will explain how and why the data was gathered. The ensuing chapter will present how the data was analyzed, before concluding with an evaluation of the research quality and a summary of the analysis.

3.1 Research design

To answer our research question, we will have to examine the data available to health policymakers. Furthermore, after a decision has been made with regards to the data source, an evaluation of the data quality follows. Our research with regards to the current creation, collection and application of data will provide us with valuable insights of how the existing data could potentially be optimized.

The research process is split into four main sections across the remaining chapters of the thesis, of which each section represents key findings in our pursuit to answer the research question. The first section revolves around the data collection process. For health policymakers to make use of the subsequent analysis, the applied data should ideally be readily available for continuous application of the analysis process. As such, the optimal source of data would be a source that possesses all, or most of the necessary information.

The second section refers to data preparation, namely the operations required to transform the raw data and make it ready for analysis. As the research question suggests that the analytics efforts should be supporting decision-making for health policymakers, it is essential that the findings from the analysis offer real value. Analysis of *'uncleansed'* data could potentially lead to erroneous conclusions. As the main objective of the analysis is to facilitate decision-making, the alternative operation to data cleansing is to cope with the consequences of unknown inaccuracy (Krishnan, et al., 2015). The resulting efforts of the two first sections will provide us with a better understanding of how the current system is working, thereby enabling identifications of where and how other methods could be applied to improve it.

The third section will be the application of data analysis, and the subsequent validation of the findings. Findings from this section will include results that provide immediate value to decision-makers, as well as required changes in data structure, collection, or quality to provide greater value.

The final section of the research will be a discussion about the findings of our study, including suggested changes and implications. For policymakers to be able to support their decision-making, the result of the analysis must provide some value that would otherwise not be accessible. The purpose of this section is to evaluate the value of current findings in light of the effort required, with regards to the data analysis, and how potential changes in current operations could facilitate better analysis and thereby improve this ratio.

3.2 Data collection

In order to be able to answer our research question and gain valuable insight into the nature of drug shortages, we recognized the need to acquire solid and relevant data. As we established that collecting primary data regarding drug shortages would likely require more time than the available timeframe for this thesis, we became dependent on reliable secondary data.

The Norwegian Medicines Agency, *Statens Legemiddelverk* (SLV), is the national administration and regulatory body in the field of medicines, both for humans and for animals (SLV, 2014). Since 2014 they have been publishing annual statistics regarding drug shortages and deregistration in Norway. All reports are publicly available, and the datasets are published as Excel spreadsheets through SLV's website at the beginning of each year. According to SLV's annual report of 2020, their strategic goal is to *"... collaborate across disciplines so that drug shortages affect public and animal health as little as possible"*. The data analysis for this thesis will rely on using the statistics collected by SLV. However, the scope of the thesis only includes the reported drug shortages of medicinal products for human use and not for animals. As the data is both collected by and administered through governmental bodies, the data applied in this thesis is the same data that is currently available for health policymakers today.

The decision to focus our analysis on drug shortages specifically on the human products was made to ease the process of data merger. The data structure of the reported shortages for veterinary medicinal preparations is however structured in the same format as the human medicines and is reported through the same channels. Thus, although not specifically included in the analysis of this thesis, it is assumed that all recommendations and claims made regarding the human drugs will apply to the veterinary drugs as well.

3.3 Data Cleansing and Preparation

The datasets retrieved from Statens Legemiddelverk (SLV) are annual reports of deregistration and drug shortages (SLV, 2020a). It is apparent from the reports that the structure of the datasets and the quantity of information included, has changed over the years. As we wanted to include observations from multiple years in our analysis to avoid possible externalities or one-off situations, the first part of the cleansing process was to merge the datasets. Changes made in terms of the

layout for the annual report for each year has led to increased complexity in term of assuring consistency for a merged dataset. Due to the gradually increased amount of information included in the reports, we found it infeasible to include all of the reports in the desired format. As such, the final decision regarding which reports to include was to merge the datasets from 2018, 2019 and 2020.

While referring to potential externalities affecting the reports of drug shortages, there has been much speculation regarding the potential impact of the global pandemic of Covid-19. In recent times, the number of reported drug shortages in Norway has almost doubled every year. As opposed to the expected effects of the pandemic, the growth has diminished in 2020. While the total number of shortages reported in 2019 was 1250, the number of shortages reported in 2020 was 1391, whereas 391 of them were continued from the year prior (SLV, 2020b). Thus, it appears to be the case that Covid-19 has shown little or insignificant impact on the situation of drug shortages.

3.3.1 Purpose of cleansing and preparation of the dataset

The acquired raw datasets were considerably disorganized and presented multiple challenges for conducting analysis. Our primary objective with the data was to explore what type of analytical methods could be applied in order to gain practical insights regarding drug shortages. In this regard, a common way to understand data is to interpret patterns and grouping, which can contribute to identify meaningful ideas of why drug shortages occur. Cluster analysis is a form of exploratory analysis which attempts to find such structures within the data. This method groups similar observations into a number of clusters based on the various input variables. The clusters will normally tend to differentiate on the variables which usually demonstrates traits that are common for each. We have formed two hypotheses for this analysis in terms of why this will benefit SLV.

1. Cluster analysis will contribute to increase the accuracy of the expected period of return for drug shortages.
2. Cluster analysis will assist to gain more practical insights that will assist for further investigation of why shortages occur.

Nevertheless, in order for cluster analysis to be conducted, cleansing and data preparations must be properly executed to acquire sensible output. Therefore, in the following sections, we will explain and demonstrate our process for the data cleansing.

3.3.2 Merging datasets

In order to merge the datasets, we had to make a decision with regards to the layout of the combined dataset. The format in which the published reports were structured is slightly different for each year. Although the amount of information included in each report is comparatively consistent, the presentation of the information displays great variation. As the layout of the report from 2020 appeared to be the most thorough and complementary, it was decided to adjust the other datasets to this standard.

The merger of the datasets required a significant number of manual operations, as information that would previously be registered in one column, had over the years been divided into two or three separate columns. For instance, the 'shortage period', the time between when the shortage was first reported and when it was reported as resolved, was previously reported as a combined value. However, since 2019, the two dates have been listed in separate columns, with a *'from'* date and a *'to'* date.

3.3.3 Acquiring the delta

One variable we identified as a possible way of yielding interesting results was unfortunately not specifically included in the original datasets. While most shortages were reported with one date for when the shortage was reported and another date for when it was reported as resolved, the datasets did not include a delta-value to represent the difference between the two dates. By subtracting the return date from the reported date, we were able to calculate the exact number of days for the shortage of each drug.

While most observations, especially from the later reports, were reported with specific dates in an appropriate format, several assumptions were made to adjust all observations to the same format:

- All observations reported with dates such as “Week 31 2019” was changed to the specific date of the first day of that week, such as “29.07.2019”.
- All dates reported as “middle of month *X*” was changed to the 15th of that respective month.
- All dates reported as “start of month *X*” was changed to the 1st of that respective month.
- All dates reported as “end of month *X*” was changed to the final date, 30th or 31st, of that respective month (28th for February).

3.3.4 Dummy variables to categorical

Due to the nature of time-based data observations, and how different analytic tools handle this information, four dummy variables were created to represent the different yearly quarters: Q1, Q2, Q3, Q4. The variables were created based on the reported start-date for the corresponding drug shortage, in which the variable Q1 is equal to 1 if the start-date occurred in January, February or March, and equal to 0 if not. The same logic was applied to all quarters, in which Q2 represents April, May and June; Q3 represents July, August and September; Q4 represent October, November and December. Once the four new variables were generated, a new column labeled “Quarter” was constructed as the four were merged into one categorical variable to replace the dummy variables. Consequently, each reported shortage was now categorized into a calendar quarter for when the shortage first occurred.

The annual reports from SLV are published in the very beginning of each year. As such, a variable that is included in all of the datasets, although in different forms, is the availability status for each of the mentioned drugs at the end of the respective calendar year. The later reports have included specific statements such as “status per 29.12.2020” marked as either available or unavailable, whereas earlier reports have applied color-coding in which green represents available, yellow as on-going shortage and red as unavailable.
In the effort to unify the information from all datasets, these columns have been replaced with a dummy variable *“Available at year-end”*. This variable is set to either 1 or 0 if reported as available or unavailable at year-end in accordance with the different standards in all the previous reports.

3.3.5 ATC-codes

A feature of the 2020-report from SLV, which has previously not been included in its reports, was the inclusion of the ATC-code for the active substance of each drug. The ATC-system is an international system for drug classification (SNL, 2020). The function of the ATC-system is to assign each active substance a unique code, according to which organ the drug acts on and the therapeutic effect it has. The codes are divided into five levels: one anatomical, two therapeutic and two for the chemical.

For example, the active substance Paracetamol has the ATC-code *N02B E01*, representing the following information:

- N - Nervous system (*anatomical*)
- N02 - Analgesics (*therapeutic*)
- N02B - Other analgesics and antipyretics (*therapeutic*)
- N02B E - Anilides (*chemical*)
- N02B E01 - paracetamol (the specific active substance)[2]

Due to the vast amount of information contained in the ATC-codes, we identified these as a key component for further analysis. While the 2020-report was the first to include ATC-code specifically, every other report included the active substance for each drug. After reaching out to Felleskatalogen, the encyclopedia of pharmaceutical preparations marketed in Norway, we received a comprehensive list of all active substances and its corresponding ATC-code. Through the use of Microsoft Excel's LOOKUP-function we were able to combine the two documents, adding ATC-code as a variable for all observations in the dataset.

When the "ATC-code"-feature was implemented for all observations, we were able to take advantage of the information contained in the codes. As the first letter of each code represents the anatomical level affected by the active substance, we created a new variable called "Anatomic Level". Based on the first letter in each ATC-code, each drug was categorized in one of the fourteen possible groups:

[2] *Example of ATC-code: Granås, Anne Gerd; Øye, Ivar: The ATC system in Store medisinske leksikon on snl.no. Retrieved March 21st, 2021, from https://sml.snl.no/ATC-ssystem*

A – DIGESTIVE ORGANS AND METABOLISM
B – BLOOD AND BLOOD-FORMING ORGANS
C – HEART AND CIRCUIT
D – DERMATOLOGICAL MEASURES
G – UROGENITAL SYSTEM AND GENDER HORMONES
H – HORMONES FOR SYSTEMIC USE, EXCL. SEX HORMONES AND INSULINS
J – ANTI-INFECTIVES FOR SYSTEMIC USE
L – ANTINEOPLASTIC AND IMMUNOME MODULE AGENTS
M – MUSCLES AND SKELETON
N – THE NERVOUS SYSTEM
P – ANTIPARASITIC, INSECTICIDES AND INSECTS
R – RESPIRATORY BODIES
S – SENSORY ORGANS
V – VARIA

3.3.6 Types of Dosage

Each reported drug shortage includes one column with detailed information regarding the specific drug's strength and concentration, package and pack size, type of dosage and the container. As an example, when a shortage for the drug Heparin was reported in December 2020, the name reported was: ***"Heparin 5000 IU / ml solution for injection, 10x5 ml vial".***

For the process of preparing the data for further analysis, this information was broken down and split into four separate variables: ***drug name***, ***number of doses***, ***amount per dosage***, and ***type of dosage.***

After successfully splitting the data into separate columns, the results showed that the total number of different package-type or container variants for all reported shortages was nineteen. For the sake of the impending analysis, the decision was made to group these variants into one of three categories: *Liquids*, *Powder/Mixture* and *Solids*. The resulting grouping is as follows:

Group 1, ***Liquids***: Drops, Vial, Cream, Balm, Spray, Liquid.

Group 2, ***Powder/Mixture***: Mixture, Solution, Powder.

Group 3, ***Solids***: Implants, Capsule, Pen, Pill, Patch, Pillow, Syringe, Tablet, Chewing gum.

3.3.7 Correcting for human errors in the input data

As we inspected the information from the reported drug shortages, it quickly became apparent that the inputs had been manually entered. Each individual entry contains a great deal of information, which provides many opportunities for human errors such as typos, hyphenation or information entered in the wrong area. As the authors of the reports have seemingly been able to input information freely, the authors are left with input-decisions that further complicates the structure of the data set. For instance, when listing the details regarding the different drugs, at times the authors vary between the use of Norwegian or English wording, the use of Latin for active substances, and full name or brand name of the manufacturers. Thus, a big part of the data cleansing for this thesis was to correct such inputs to make the data consistent.

Mass-correction of text-based data such as this is a challenging task, both with respect to the complexity of the task as well as the large quantity of data. With regards to corrections names such as the manufacturer, this was done through alphabetical sorting the names and removing all duplicates, before manually correcting the wrongfully entered inputs. The same procedure was subsequently repeated for the active substances and the name of the drugs.

3.3.8 Country and Continent of Origin

As a focus of this thesis is on examining underlying patters for the occurring drug shortages, we felt the need to expand on some of the information provided through the dataset. One of the areas we wanted to explore further was the relationship between the drug shortages and the respective manufacturers. The original datasets do not contain any information about the manufacturers apart from their name. To broaden this category for further analysis, we therefore included the country and continent of origin for each manufacturer.

3.3.9 Size of the manufacturer

For the purpose of expanding the information regarding the manufacturers further, a new feature was added to the dataset: “Size of the manufacturer”. Each manufacturer was manually assigned to one of three possible groups: ***Small, Medium*** or ***Large.*** A mix of logics was applied in the assignment process.

Manufacturers with less than 300 employees were initially categorized as small, more than 300 employees as medium and more than 1000 employees as large. The companies with a total revenue in 2020 of more than $10 billion or those who ranked among the most valuable companies in Torreya's different categories from their 2020 report regarding pharmaceutical companies, were automatically assigned to the "***Large"***-category (Torreya, 2020). Similarly, if any of the small or medium-sized companies outperformed the majority of its category, it was moved from small to medium, or from medium to large respectively.

3.3.10 Causes of drug shortages

All shortages reported by SLV is registered with a suspected cause of the shortage. While most reported causes fall into one of a number of categories, the nature in which the shortages are reported allows for slight or drastic differences due to the fact that most causes are reported with specific details. The lack of a uniform definition of shortages is fundamentally affecting the data in the sense that the suggested causes vary heavily, assumably as a result of different perspectives of the different submitters.

The process of implementing standardization as explained by Braa & Sahay (2012) allows for an evaluation of how far the process has come with regards to level standardization in Norway. At the top level of a standardization process is the decision-making authority of an organizational or political body. As Norway is currently not subject to any international standards from bodies such as WHO or the EU, the body in question will most likely be the Norwegian Government, presumably through a branch such as the Norwegian Medicines Agency. The body at this level has the power to decide the required standards, but as our research suggests and the data confirms, no such standard is yet universally agreed upon.

The second level in the standard implementation is the semantic level, indicating the level of standards for data and indicators. The deciding requirement at this level is shared understanding and meaning. The degree of variation in the collected data suggests that the Norwegian system for drug shortage reporting is far away from a common understanding and decided definitions. The data quality actually suggests that it is not yet at the lowest level, the *"Syntactic/technical"***.** At

this level it is expected that users, in this case the submitters of drug shortages, to share an agreement for the appropriate "grammar" or terminology. The absence of definitions or standards in both the context of drug shortages in general, as well as in the reporting process of these shortages, indicates that Norway is currently far from implementing standardization.

Subsequently, as the shortages are currently reported without strict standards or guidelines, the resulting reports lead to a lot of variation, making logical grouping of the results a lot more challenging. For instance, two separate shortages could be caused as a result of price changes but are not easily defined as a result of the same cause. If the first cause is simply reported as *"price changes"* while the other shortage is reported with the cause *"limited availability in this period due to unforeseen changes in the selling price"*, the desired grouping requires a lot of manual operations.

Application of automated sorting of the reported causes based on some simple duplicate-checks indicated that the reports include a total of more than 800 different variations of causes. Manual reviews of the causes enabled us to place all reported causes into one of 37 initial categories. From these categories we were then able to sort each category into one of five groups, eventually providing a classification for each individual shortage. The five final classifications and the included causes in these categories are as follows:

- **Expired/Deregistered:** Deregistered, discontinued, temporarily expired, expired.
- **Import/Export/Distribution:** Change in distribution, Export/Import, Modification of portfolio, Delivery, Parallel exports.
- **Administrative/Deficiencies/Problems:** Admin, Waiting, Deviations, Access, Fires, Changes in marketing-rights, API's, Delays, Quality problems, Missing, Problems, Technical problems, Recalls, Accidents.
- **Production/Price/Demand:** Price changes, Gasket changes, Change of production site, Demand, Incorrect calculations of sales, Capacity, Production-related, Commodity-related, Serialization, Contracts.
- **External factors/Others:** Brexit, Covid, Unknown, FMD.

3.3.11 Create nominal variables

When the data was transformed through the forementioned stages and the observations were categorized into the newly created groups, the final step of the data cleansing process was to prepare the data for further analysis. As most of the data from the reports, though transformed for consistency, mainly consisted of text, we created new variables to represent a nominal value for each of the categories.

The first nominal variable created was for each of the groups in the types of dosage. The three categories in this group were ***Liquids***, ***Powder/Mixture*** and ***Solids.*** The nominal values assigned to the groups were: ***Liquids = 1***, ***Powder/Mixture =2*** and ***Solids = 3.***

The second nominal variable created was for the "Cause"-feature. In this feature we have the five categories ***Expired/Deregistered, Import/Export/Distribution, Administrative/Deficiencies/Problems, Production/Price/Demand*** and ***External factors/Others.*** The nominal values assigned to these groups were: ***Expired/Deregistered = 1, Import/Export/Distribution = 2, Administrative/Deficiencies/Problems = 3, Production/Price/Demand = 4*** and ***External factors/Others = 5.***

The third nominal value created was to represent each of the continents of origin for the manufactures. The assigned nominal values for this category were: ***Europe = 1, Americas = 2, Asia = 3*** and ***Africa = 4.***

The final nominal feature added to the dataset, was done to assign a nominal value for the size of the manufacturer for the reported drugs. In this category, manufacturers were assigned to one of three groups: ***Large, Medium,*** or ***Small.*** The groups were replaced by numeric values in the following manner: ***Large = 1, Medium = 2*** or ***Small = 3.***

IV. Cluster Analysis

Following the cleaning and preparation of the dataset, the next step requires finding the appropriate method of clustering. The objective of the analysis was to uncover underlying trends by applying unsupervised Machine Learning to the

dataset and group the unique observations. The dataset includes a total of 10 variables, whereas two is of type integer and eight is categorical variable.

Table 1 – Description of variables

Variable	Variable type	Description of variable
Organ	Categorical	Organ variable contains information of which organ is the medication targeted to. There are in total 13 organs categories.
Type	Categorical	Type variable contains information of what kind of type the medication is. The three types are: Solids, liquid, or powder/mixture.
Active substance	Categorical	Active substance contains information about which of the substances in the medication is the primary substance. There are in total 807 categories.
Company size of supplier	Categorical	The company size of supplier is categorized in three different areas; Large, Medium, and Small.
Continent of supplier	Categorical	Continent of supplier contains information about which continent the headquarters for the company is located.
Cause of shortage	Categorical	Cause of shortage contains information on an aggregated level what the underlying cause of shortage for the medication is. The total number of 526 micro causes was divided into 5 macro causes.
Quarter	Categorical	This variable contains information about the calendar quarter of which the medication experienced shortage.
Status	Categorical	Status contains information about whether the medication was available at the end of year the medication went missing.
Dosage	Numeric	Dosage contains information about number of dosages for the medication in each package.
Delta	Numeric	Delta contains information about difference between the date of shortage and the date of shortage end.

As illustrated in Table 1, the dataset contains a mixture of categorical and numeric variables. Furthermore, in unsupervised machine learning and clustering, the K-means algorithm is the most utilized tool to divide data into homogenous groups. Alternatively, a k-mode can also be applied which does not use numerical distances as the K-means, but rather dissimilarities into cluster. The K-modes method is usually applied for dataset containing categorical variables, while the K-means is used for variables of the numerical datatype. For our clustering process, we applied a combination of the two methods mentioned which is named K-prototype. This algorithm was introduced by Zhexue Huang in 1998 as an extension to the K-means and K-modes algorithms. The reasoning for the application of this mixture is grounded in the fact that most datasets regarding drug shortages usually contain a combination of numeric and categorical variables. In the next section, we will elaborate on the logic behind the K-prototype and its notations.

4.1 The K-prototype algorithm

The K-protype algorithm is based on the combination of the mathematical formulation of the K-modes and K-means. The algorithms cost function is simply a minimization of the sum of the distance for the numerical variables, and the dissimilarities in the categorical variables. This can be further shown in the following formulas.

Mathematics formula

Assume that $X = \{X_{1,}X_{2,\dots,}X_n\}$ is a sequence of n objects and $X_i = \{X_{i1}, X_{i2,\dots,}X_{im}\}^m$ with m denoting the variables and i denoting the i-th cluster

The measure of Similarity

The general method of calculating the measure of similarities is defined as follows.

$$d(X_i, Z_l) = \sum_{j=1}^{m} \delta(x_{ij}, z_{lj)}$$

Where $Z_l = \{z_{l1}, z_{l2,\dots,}z_{lm}\}^T$ in this case is a prototype for the given cluster l. Furthermore, the Euclidian distance which is a well-known measure of similarity for numerical variables is denoted as follows.

$$d(X_i, Z_l) = \sqrt{\sum_{j=1}^{m_r} (x_{ij}^r - z_{lj}^r)^2}$$

Where x_{ij}^r represents a value of numerical variables j, z_{lj}^r represents the average of prototypes for numerical variables j in cluster m.

Following the same methodological approach, the next formula denotes the measure of similarity for categorical variables.

$$d(X_i, Z_l) = \gamma_l \sum_{j=1+1}^{m_c} \delta(x_{ij}^c, z_{lj}^c)$$

In which, the basic matching similarity measure is denoted as follows,

$$\delta(x_{ij}^c, z_{lj}^c) = \begin{matrix} 0, & x_{ij}^c = z_{lj}^c \\ 1, & x_{ij}^c \neq z_{lj}^c \end{matrix}$$

where the weight of the unique categorical variables for cluster l, which in this case is the standard deviation of numerical variables in each cluster, is represented by γ_l. Further, categorical variables are denoted by x_{ij}^c, while z_{lj}^c is the mode for each variable in cluster l. The number of categorical variables is denoted by m_c.

Next, a minor modification of the basic matching similarity measure is applied which gives.

$$\delta(x_{ij}^c, z_{lj}^c) = \begin{matrix} 1 - \omega(x_{ij}^c, l), & x_{ij}^c = z_{lj}^c \\ 1 \qquad\qquad\quad , & x_{ij}^c \neq z_{lj}^c \end{matrix}$$

The purpose with the formula and modification above is to increase the object similarity within the clusters with categorical variables. Ultimately, the outcome of the cluster analysis will be improved to the previous solution as $\omega(x_{ij}^c, l)$ denotes the weight for x_{ij}^c were

$$\omega(x_{ij}^c, l) = \frac{f(x_{ij}^c|c_l)}{|c_l| \cdot f(x_{ij}^c|D)}$$

Where frequency of x_{ij}^c in cluster l is denoted by $f(x_{ij}^c|c_l)$, number of objects in cluster l is $|c_l|$ and the frequency of x_{ij}^c in the entire dataset is given by $f(x_{ij}^c|D)$.

In summary, the previous equations show the K-prototype algorithm construct measures of similarity and groups the mixed dataset of numerical and categorical variables into unique clusters through this given mathematical formula,

$$d(X_i, Z_l) = \sqrt{\sum_{j=1}^{m_r}(x_{ij}^r - z_{lj}^r)^2 + \gamma_l \sum_{j=1+1}^{m_c} \delta(x_{ij}^c, z_{lj}^c)}$$

4.2 Huang Cost function

In the normal K-means clustering algorithm, the underlying mechanism of the process is based on minimizing a cost function by changing the means of numerical variables, but not the categorical, which does not have any natural mean. In this relation, the K-prototype algorithm aims to solve this challenge by optimizing the cost function on the whole dataset, for both categorical and numerical variables. In addition, the method guarantees a local optimal clustering and has no limitations in terms of the size of dataset. The cost function defined by Huang for the minimization is as follows,

$$Cost_l = \sum_{l=1}^{K} u_{il} \sum_{j=1}^{m_r} (x_{ij}^r - z_{lj}^r)^2 + \gamma_l \sum_{j=1}^{m_c} u_{il} \sum_{j=1}^{m_c} \delta\left(x_{ij}^c, z_{lj}^c\right)$$

$$Cost_l = Cost_l^r + Cost_l^c$$

In which the total cost of all the numerical variables in cluster l is represented by $Cost_l^r$. The process that is taking place in the optimization is in simple terms that $Cost_l^r$ is being minimized while z_{lk} is calculated with the given equation,

$$z_{lj} = \frac{1}{n_l} \sum_{i=1}^{n} u_{il} \cdot x_{ij} \quad \text{for } j = 1,2, \dots, m$$

in which $n_l = \sum_{i=1}^{n} u_{il} \cdot x_{ij}$ represents the number of objects in cluster l.

Moreover, the categorical variables e.g., C_j is defined as a group of unique value in each categorical variable j and $p(q_{ij}^c \in C_j|l)$ is the given probability for C_j in a cluster l. Hence, the formula of $Cost_l^c$ can be shown as,

$$Cost_l^c = \gamma_l \sum_{j=1}^{m_c} n_l(1 - p(q_{ij}^c \in C_j|l))$$

In which, the objects within cluster l is denoted by n_l.

4.3 Clustering

The clustering consisted of several steps and was conducted primarily in Python. The dataset consisted of 3796 observation of medicine shortages and had no missing values in the sample. In the upcoming paragraphs, we will elaborate on exploration of the data, the clustering process, and the final results.

Data Exploration

Starting off with the categorical variables, there are eight variables present in the dataset. Moreover, in order to conduct a meaningful cluster analysis, it is important to have variables at aggerated level. In other words, the number of unique categories for each of these variables should not be at a detailed level (Rhemtulla, Brosseau-Liard, & Savalei, 2012). This will make it challenging for the model to find patterns for the clusters. The majority of the categorical variables was converted to an aggerated level in the data cleaning process. However, the variable “Active substance” did not seem to have any natural

grouping at a higher level. This led to dropping this variable in the clustering as did not seem to make any difference in the output.

Furthermore, the numerical values in the dataset consisted of number of dosages of each medication and the delta between the first day of shortage and it was first available again. The average number of dosages in the full dataset was 55, while the average delta value was 101 days. The numerical variables seemed to be appropriate and did not require any adjustments.

Clustering process

The general idea of the K-prototype clustering is to differentiate the categorical and numerical variables in the clustering analysis as shown mathematically above. In this relation, the next step after the exploration, was to split the dataset into these two groups. This was done by locating the categorical variables by finding all columns of type "object" in the dataset and assigning these to a new variable. The purpose of this distinction will ease the process of finding the exact number of clusters that are optimal for this dataset.

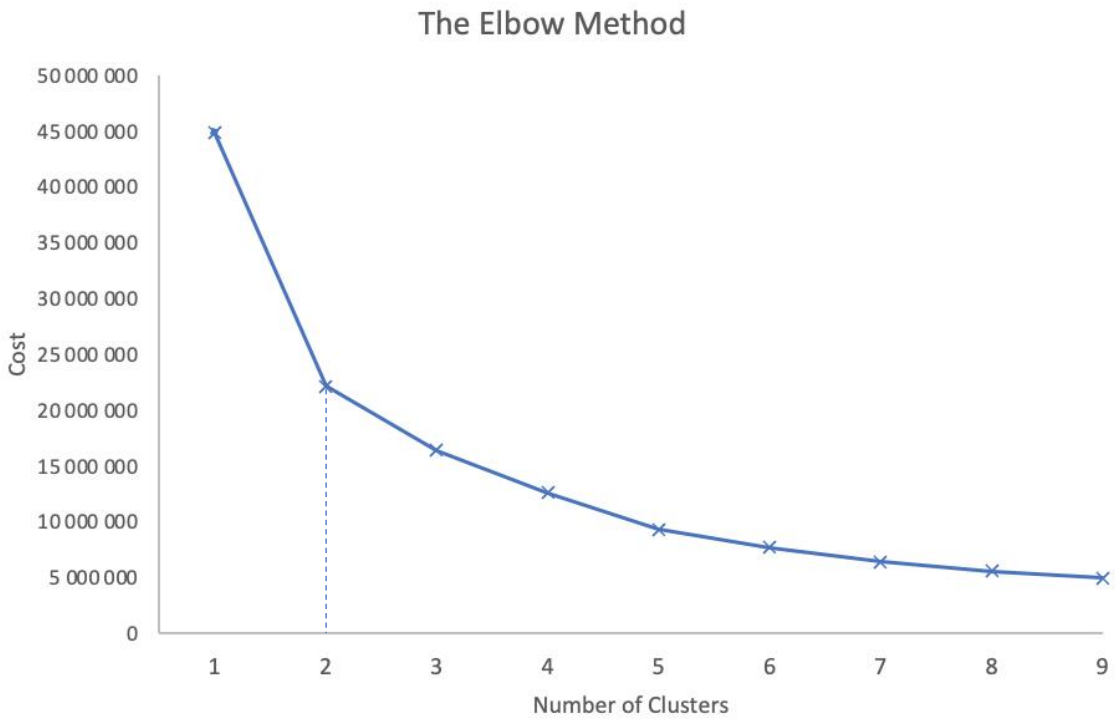
4.4 Finding the optimal K

Prior the clustering itself, it is essential to identify the optimal number of clusters for the dataset as this is an input parameter for the model. There are a variety of methods for determining this number that has been proposed in earlier research, however, we will try one of the most common methods and explore the outcome.

The Elbow method

The elbow method is a technique in which the total sum of squares of each number of clusters is measured and plotted on a graph. Normally, the method is calculated with the Euclidian distance, however, the K-prototype provides a cost function which sums the cost of the categorical and the numerical variables. Moreover, the objective of the plot is to visualize in where in the graph the change of slope from steep to shallow exits (Elbow). This will determine the optimal number of clusters. The technique is not entirely accurate; however, it will give us some indication of how many clusters there could be in the dataset.

Figure 2 – The Elbow Method



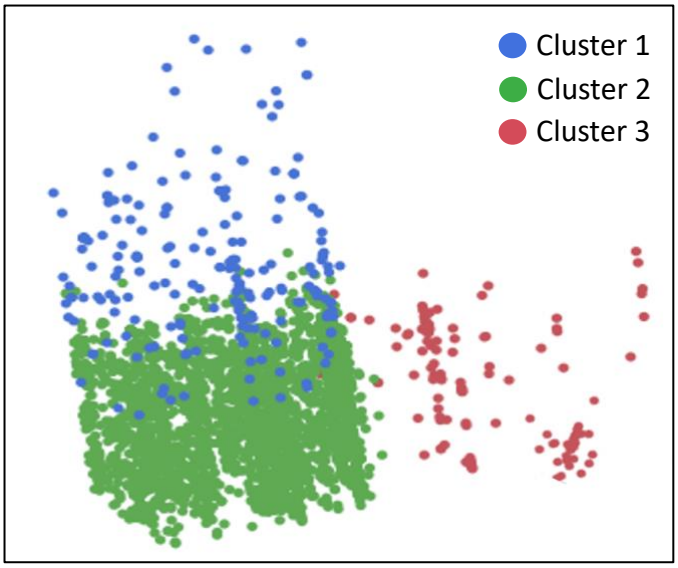
The method, which was iterated for maximum of nine clusters, illustrates that in the second iteration we can spot this elbow area. This essentially means that the optimal number of clusters provided by the elbow method for the K-prototype analysis is 2. However, we will also explore the possibility of 3 and 4 clusters to adjust for the shortcomings of this method.

4.5 Results

After running the K-prototypes cluster analysis for 2, 3 and 4 clusters, the results showed that for both 2 and 4, there were no considerable differences between the groups. For K= 2, It appears that the algorithm fails to identify any significant differences for all variables expect the delta value, of which one cluster possessed a very high average delta and the second cluster was quite low. Moreover, for K=4, the groups seemed to be quite similar where the only distinction was the delta for one of the clusters, which was considerably higher compared to the other three groups.

Nevertheless, when number of clusters was set to 3, the K-prototype clustering appears to have distinguished the dataset into groups with several differences for some of the variables.

Figure 3 – Visualization of clusters



Looking at the plot of the clusters, we can spot that there are three different clusters forming present in the dataset. However, the distance between the clusters appears to be quite small as some of the observations for the different clusters are overlapping. We will in the next section elaborate on key findings that differentiate the clusters.

Cluster descriptions

The means and frequencies for each cluster provides what we could expect from the typical observation. In the total observations for the clusters, we can see that cluster 2 has the largest number of the observations in the dataset, while cluster 1 and 3 are considerably smaller. This may be a result of the similarities between most of the observations, however, the model was able to identify some notable differences in some variables.

Table 2 – Clusters means and frequencies

Cluster	***Total***	***Company size of supplier***	***Quarter***	***Organ***	***Cause of shortage***	***Dosages***	***Delta***
1	375	Medium	Q3	MUSCLES AND SKELETON	production/ price/contract/demand	63	251
2	3155	Large	Q1	THE NERVOUS SYSTEM	production/ price/contract/demand	55	74
3	266	Large	Q2	HEART AND CIRCUIT	admin/deviance/waiting/ delays/missing	42	618

Cluster 1

Cluster 1 is the second largest cluster with a total of 375 observations and mainly consists of medicines that has suppliers of a medium company size. The shortages of majority of the observations usually occurs in the third quarter of the year and the medications in this segment are typically used for muscles and skeleton. Furthermore, the cause of shortages ranges from lack of production, distributor and supplier contract issues and higher demand for the medication. The average number of dosages for the medications are 63 while the average number of days of shortage is 251.

Cluster 2

Cluster 2 is the largest segment in the analysis and contains 83% of the total observations. The company size of the supplier for these medications are large and the majority of the shortages in this segment most likely occurs in the beginning of the year in the first quarter. The typical organ of which the medications in this segment have an effect on is the nervous system. The main causes of shortages are similar to the causes for cluster 1. Lastly, this segment has 55 in average when it comes to number of dosages for the medication, while the 74 days is the average delta. This is the lowest delta for all three segments.

Cluster 3

Cluster 3 has the lowest number of observations with 266, and the majority of the supplier companies is of a large size. Typical for the medications in this segment is that it is likely that the shortages occur in the second quarter of the year. Heart and circuit are the organs most often treated with these medications. Moreover, the main reasons of shortages are mainly due to administrative issues, deviation in delivery schedule, delays due to transportation, and shortages of key raw materials for productions. The average number of dosages for this cluster is the lowest with 42, while the average delta is 618 days, considerably higher than the two other segments.

Summary of cluster analysis

The K-prototype cluster analysis demonstrates there exists different groups in the dataset. While it may be that the three clusters are unbalanced, the analysis

captures certain traits for some of the variables that differentiate. Moreover, Cluster 3 has as mentioned a significantly higher average delta which may be a result of the group's unique causes of shortages, which differentiates from the other two clusters.

The analysis is contributing to gather information which can be used to gain knowledge and understand more about drug shortages. The problem at hand indicates several consequences for various stakeholders in a society when shortages occur. Some of these includes:

- Increased number of consumers with symptoms
- Decrease in revenue for pharmacies
- Increased number of patients at hospitals

These consequences are all important to address, which is why a cluster analysis can assist to anticipate and facilitate for potential shortages. For instance, a possible application of the cluster analysis could be to place all drugs available in clusters. Through such an operation, SLV would be able to compare the results of other drugs in the same cluster to provide estimates for factors such as the expected shortage period (delta), based on the average delta value for that cluster. This confirms that hypothesis 1 is applicable, which suggests that the cluster analysis will contribute to increased accuracy of the expected period of return for drug shortages. In addition, the causes and background for why the shortage occurred will be known so that further actions can be taken to reduce the delta. This confirms that hypothesis 2 is applicable, suggesting that the cluster analysis will assist in providing more practical insights. In the scope of Bohn's (1994) knowledge level framework, the two hypotheses correlate well to the transition from the lower knowledge levels to the higher levels. When the results from a study such as the cluster analysis is confirmed to be able to measure and display some control the variables, this could be seen as evidence of progress in terms of the knowledge level. Consequently, adjustments made with respect to the findings as proposed by hypothesis 2 would provide confirmation of further climbing of the knowledge ladder. Conclusively, by deciphering the structure and patterns of a dataset through a cluster analysis, health policy makers will acquire more insights which can contribute to reduce duration of shortages which will eventually slow

down the effect of the consequences related to drug shortages.

V. Discussion

Past studies on the subject of drug shortages have identified the importance of information gathering for the process of mitigation and sufficient planning (Ventola, 2011; Fox, et al., 2009). In line with these studies, our ambition for this thesis is to evaluate how business analytics methods can help Norwegian health policymakers fight against drug shortages. The main objective of business analytics is to create value and support decision-making through statistical methods and analysis of data. Hence, a business analytical approach to the matter would indicate to assess the process of data collection and evaluate how the collected data could be applied for meaningful analysis. In light of the research question, we can begin to discuss the findings from our study and the literature review.

5.1 The shortage list

Bonsón & Escobar (2006) discuss the importance of standardization of the format for reporting with regards to the facilitation of analysis. Our findings suggest that the format of the reported drug shortages in Norway are currently far from optimal. Our efforts to perform data analysis on the reported drug shortages showed that the data required a considerable amount of cleansing before it was ready for use. Cases of drug shortages in Norway are reported directly to SLV. The Marketing Authorization Holder (MAH) of a drug has a duty to report a shortage that is expected to last longer than 14 days. Shortages should be reported as soon as possible, and no later than two months before the supply cut (Legemiddelverket, 2020). The MAH must also report the shortage to the European Medicines Agency (EMA) in cases where the cause(s) of the shortage includes any of the following reasons:

- The drug is harmful
- The drug has no therapeutic effect
- The benefit/risk ratio is not positive
- The qualitative and/or quantitative composition of the medicinal product is not as it is indicated on the label, or

- The control of the medicinal product and its ingredients, the production, or other conditions of importance for the marketing authorization are not in accordance the legislative requirements.

The reports to SLV are submitted using their own notification form. Since June 1st of 2020 SLV have stopped receiving forms formatted in Microsoft Word and is thereby forcing submitters to complete their submission through an electronic form. The electronic form consists of five or seven pages (dependent on the number of different variants of the specific drug is reported) with various methods of information input, ranging from text fields to drop-down menus. Input-fields such as the 'name of the drug' and 'active substance' are mandatory to progress in the submission process. The two fields are examples of the 'text' input-method in which the submitter must input the information manually. These fields do not offer any sort of correction for potential typos and makes no specification with regards to how many active substances should be included when a drug contains multiple. Hence, these input-methods are allowing for a great variety in terms of quality of information submitted as well as human mistakes. The input-fields of the electronic form directly mirror the different columns from the annual report of 2020, indicating that the report of 2021 will most likely follow the same or similar format as the previous year. The design of the electronic form is almost identical to the reporting template suggested by Pauwel et al. (2014), and as such is subject to the same criticism as the Norwegian form. While the Norwegian Directorate of Health have reported that a more digital process is under development, the current system is compiling the reported drug shortages in excel spreadsheets, such as the ones used in the data analysis portion of this thesis. The spreadsheets include all reported drug shortages from the respective calendar year of the reported shortage, in an unstructured format that makes navigation inconvenient.

Another issue that was uncovered while we were performing the data preparation for analysis was the apparent changing structure of the annual reports. The lack of an established standard over the years has complicated the process of analyzing historical data.

5.2 Findings from data cleansing and analysis

One of the main concerns for applying business analytics methods is the chosen data and its respective quality. The available source of data for health policymakers in Norway is the shortage list maintained by Statens Legemiddelverk (SLV). Through our data cleansing process and subsequent analysis, we identified multiple concerns with the current data quality. Among these concerns is the reoccurring factor that appear to affect most levels of the system, namely the lack of standardization. As discussed by De Weerdt et al. (2015), unlike countries such France and Belgium, Norway does not currently operate with a legal definition for drug shortages. As such, the reported drug shortages in Norway do not coincide to a common definition. As a result, any analysis and subsequent findings based on the current reports do not portray an accurate picture of one interpretation of drug shortages, but rather a mix of many. As the definitions of drug shortages also vary across international borders, international comparisons are to an extent invalidated. With regards to the three levels of standardization discussed by Braa & Sahay (2012), the lack of a uniform definition suggests that the system is yet to achieve the lowest level of standardization: the "syntactic/technical" level. According the Braa & Sahay, a shared understanding of the applied grammar and terminology for the users of the system is the essential foundation for further standardization. In the absence of an internationally agreed upon definition, this disparity in terminology should not come as a surprise. However, a more concerning discovery of our research, as discussed by Jahre et al. (2021) is the apparent divergence of understanding regarding drug shortages within Norway.

The cluster analysis revealed that the underlying information and characteristics of the reported drugs shortages made grouping of the incidents possible. From the perspective of health policymakers, the question remains as to how this information could support their decision-making. The clusters represent certain traits regarding the trend of their respective cluster, such as Cluster 1 mainly consisting of reports regarding shortages reported with production or sales-related causes, with a majority of its incidents submitted in the third quarter of the year. Results of the analysis using the information presently available provides some indications. For instance, wholesalers in charge of ordering the medications of this cluster should closely monitor the availability of these drug from the end of the summer, to stay aware of possible shortages. While this information is useful, the

subsequent questions to answer will relate to why this is the case and how the shortages occur. Unfortunately, this is currently where the cluster analysis falls short. As it stands, what the cluster analysis fails to uncover sufficiently is the underlying trends among the drugs of each cluster, which could be further analyzed once a cluster is discovered (Popovic et al., 2014). Each individual report is recorded as an independent incident, with no relational information provided through the data. Thus, the lack of consistency of the reports makes relational grouping very challenging in its current form. If, for instance, each drug was reported with a producer which had in advance been related to a country or region, further analysis of the clusters could examine the relation between the regions responsible for shortage to possibly uncover unreported reasons for the shortages. Such analysis would require that the data quality and structure support the needed analytics efforts, thus reaffirming our claim regarding the need for standardization.

As the results of our cluster analysis consists of heavily cleansed data based on numerous assumptions, it is reasonable to believe that properly structured input data with standardized information will also result in improved clustering and more accurate results, to further improve the efficiency of the resulting measures to counter them (Provost & Fawcett, 2013). As discussed by Chen et al. (2012), business analytics supports decision-making through combinations of applications and techniques, further suggesting that the results of the cluster analysis should merely be part of the foundation in which decisions are supported through. Studies regarding the advantages of digital reporting such as Bonsón & Escobar (2006) suggests that organization should make use of the potential it provides for analysis. In line with this claim we suggest that standardization will act as the enabler of other analytical methods that could be used to further examine, support, and explain the results of studies such as the cluster analysis.

5.3 Purpose of standardization and business analytics methods

Simons (1987, 1994) reflects on drivers of strategic management and identifies the value of diagnostic control. Diagnostic controls enable monitoring of outcomes and deviation from expectations. For implementation of diagnostic control, Simons defines the need for feedback and measurement systems. Establishing significant feedback and measurement requires sufficient

standardization of the information in question, ensuring valuable comparisons between the expected and actual outcomes. With regards to framework of technological knowledge presented by Bohn (1994) discussed in the literature review of this thesis, we suggested that Norway's knowledge level regarding drug shortages ranked somewhere among the three lowest levels. In the process of attaining further insight, and thereby progressing in their knowledge level on the subject, the next levels of the framework are focused on the measurement, control, and processing capabilities of variables. This coincides well with Simons (1994) emphasis on feedback and measurements.

Standards simplify sharing, thus promoting pooled analysis and cross-product comparisons. As discussed by De Weerdt et al. (2015), the ambition should be a uniform European or even global definition for drug shortages to facilitate international cooperation in terms of analysis and comparisons. While standardization alone does not ensure data quality, high-quality data is usually a result of certain degree of standardization. The current data quality of the reported drug shortages does not sufficiently support analysis. The data is rather chaotic and a follows complex structure, requiring a lot of preparation before meaningful analysis can be performed. As seen in the data cleansing section of this thesis, numerous operations and assumptions are required, making analytical activities time-consuming and inefficient. An essential part of business analytics is the capturing, cleansing, and accessing of data, all of which refers to the category of data management. Stored and structured datasets are referred to as databases, and actions made by a database-owner to control or manipulate the input data is called database management. Manual maintenance and supervision of databases can be tedious and require a lot of effort. However, the efforts towards automating the process have already come a long way. Harrington (2009) discuss implementation of relational databases, and defines software known as database management systems (DBMS) as software that can translate between the user's request or input of data to the physical data storage. For most databases, the person interacting with the data, for instance the individual who is reporting a drug shortage, has most likely no reason to worry about how the data is physically stored. Thus, a digital tool such as a DBMS can help ease the process between input of data and proper storage, facilitating clean and accessible data for extraction. Analytical activities would then not only be supported through ease of access to clean and

useful data, but this would also support the use of automated algorithms that could perform live analysis on the current shortage list, allowing for real-time reports and updates. Thus, a functioning and well-structured database could well be the start of a successful 'warning-system', such as the one suggested by Ventola (2011).

5.4 Digitalization

As discussed in the literature review of this thesis, The Norwegian Directorate of Health has suggested that increased efforts towards digitalization may facilitate analysis and categorization. We believe that a major step towards this digitalization could be to increase the use of digital tools for data collection in the reporting process of drug shortages. With regards to the knowledge framework suggested by Bohn (1994), our findings suggests that comprehensive analysis is required to continue climbing the knowledge ladder. It is our belief that such analysis is unattainable in the absence of a revised platform for digital reporting. In its current form, the format of the reported shortages promotes ambiguity due to the lack of standardization. We suggest that the repercussions of this lack of standardization are reflected in the entire Norwegian system for handling of drug shortages, due to causal relationships. To illustrate how we believe improved digitalization and subsequent standardization in the reporting process would affect the current system, we can apply methods from the field of system dynamics. System dynamics is a method of addressing and learning from complex systems. The concept refers to the process of mapping complex systems through easily interpretable illustrations, that can be used to simulate a real-life process. The representation of the systems allows for display of critical feedback that would otherwise easily be forgotten in mental models (Sterman, 2000). System dynamic models are behavioral theories used to indicate how a real system works. Among the various models used for representation of the feedback structure of systems is the Causal Loop Diagrams (CLDs).

The CLDs are well-suited for display of a hypothesis regarding the causes of dynamics, to capture mental models or to communicate important feedback that is assumed to be responsible for a problem (Sterman, 2000). CLDs are a way of *'describing the system'* by applying various bits of information about the real world into a coherent and unifying theory (Forrester, 1994). Causal diagrams

consist of variables that are connected through arrows, in which the arrows indicate the causal influence between them. A causal relationship between two variables is either assigned a positive or a negative polarity, indicating how a potential change in one variable will affect the other. If for instance the polarity of the causal link between ‘variable x’ and ‘variable y’ is positive, an increase in ‘variable x’ would result in an increase for ‘variable y’. In contrast, if the polarity of the link is negative, an increase in ‘variable x’ would result in a decrease for ‘variable y’. In the effort to illustrate our findings and suggested solutions of improvements, we illustrate the causal relationships between the active variables in a causal loop diagram. The diagram is meant to represent the current system in which drug shortages are reported. This indicates how the lack of important features such as the forementioned required standardization is affecting the total system.

CAUSAL LOOP DIAGRAM

Figure 4 – Current System

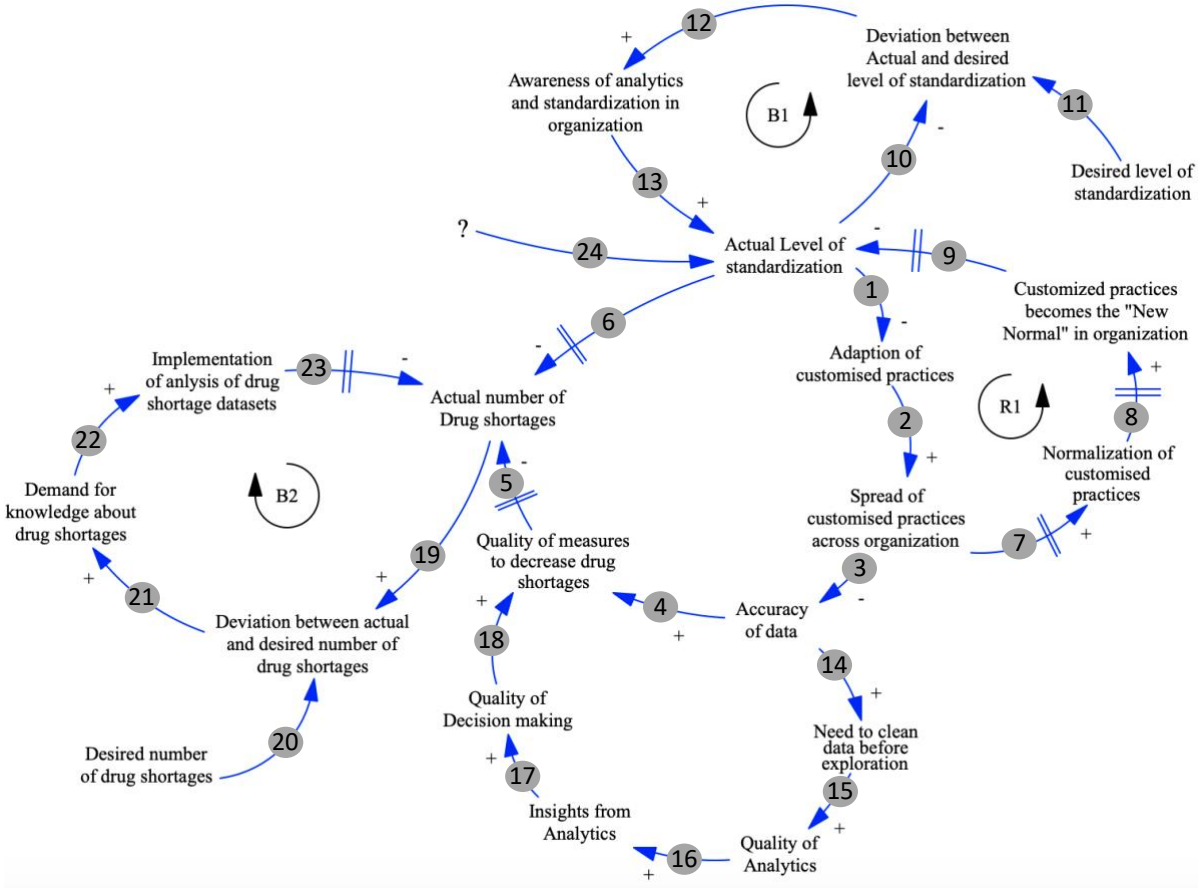


Table 3 – Current reporting system

Link	*Assertions about causal relationships in Figure 4*	*References*
1	Higher level of standardization leads to lower level of customized practices	(Osgood et al., 1996)
2	As the organization adapt to customized process, the spread of customized practices will increase across organization.	(Weick & Quinn, 1999), (Vaughan, 1996)
3	Spread of customized practices has a negative effect on accuracy of data.	(Lintz, 2018), (Bonsón & Escobar, 2006)
4	High level of Accuracy of data will have a positive effect on quality of measures to decrease drug shortages	(Lintz, 2018), (Weber, 1987)
5	High level of quality of measures for drug shortages will contribute to decrease number of Drug shortages over time.	(Redman, 1998), (Federgruen, 2012)
6	Low level of standardization has a negative effect on number of Drug shortages over time.	(Ritter, 2009), (Federgruen, 2012)
7	Spread of customized practices across organization will enter a process of normalization for a period of time.	(Pinto, 2014), (Vaughan, 1996)
8	After the normalization period, customized practices within the organization will become the “New Normal”.	(Vaughan, 1996), (Pinto, 2014), (Prielipp et al., 2010)
9	The “New Normal” of customized practices will further increase the lack of standardization.	(Vaughan, 1996), (Pinto, 2014), (Prielipp et al., 2010)
10	Low level of actual standardization has a negative effect (Higher) on the deviation between actual and desired level of standardization. (Actual – desired)	(Forrester J. W., 1994)
11	Desired level of standardization is a constant in the deviation of standardization.	(Forrester J. W., 1994)

Table 4 – Current reporting system (continued)

Link	*Assertions about causal relationships in Figure 4*	*References*
12	Higher deviation between actual and desired level of standardization will increase awareness of analytics and standardization in organization.	(Forrester J. W., 1994)
13	Higher awareness of analytics and standardization in organization will contribute to increase the actual level of standardization.	(Karim et al., 2016), (Busemeyer & Pleskac, 2009)
14	High accuracy of data has a positive effect on the need of cleaning data before exploration.	(Haug, Zachariassen, & van Liempd, 2011)
15	Sufficient data exploration and cleaning has a positive impact on quality of analytics.	(Wang & Strong, 1996), (Popovic et al., 2014)
16	High quality of Analytics contributes to increased insights from analytics	(Wang & Strong, 1996), (Popovic et al., 2014)
17	More insights from analytics gives better decision-making.	(Braman, 1989), (Provost & Fawcett, 2013)
18	Better decision-making will have a positive impact on measure against Drug shortages.	(Ritter, 2009) (Brynjolfsson, Hitt, & Kim, 2011)
19	Increased number of actual drug shortages will increase the deviation between actual and desired number of drug shortages. (Actual – Desired)	(Forrester J. W., 1994)
20	Desired number of drug shortages is a constant in the deviation of drug shortages.	(Forrester J. W., 1994)
21	Higher deviation between actual and desired number of drug shortages will increase the demand for knowledge about drug shortages.	(Forrester J. W., 1994)

Table 5 – Current reporting system (continued)

Link	*Assertions about causal relationships in Figure 4*	*References*
22	Increased (demand) knowledge about drug shortages will contribute to increase in analysis of drug shortage datasets.	(Willingham, 2006), (Dobrev, Kralovic, & Kralovic, 2013)
23	Increase in analysis of drug shortages will contribute to decrease the actual number of drug shortages.	(Vessey, 1994), (Dvir et al., 2003)
24	The link from "?" to "Lack of standardization will be elaborated in the implications section.	

5.5 About the causal loop diagram

The Causal loop diagram illustrates a potential system of the reporting and standardization for SLV. The process begins by assessing the actual level of standardization, which is considered as low given the unstructured datasets of drug shortages. Nevertheless, The CLD demonstrates that; higher level of standardization will further lead to a decrease in the adaptation of customized practices amongst the loggers in the organization. As discussed by Weick & Quinn (1999), variations in practice leads to continuous change, which may gradually take form of tendencies to normalization. Furthermore, adaption of customized practice will contribute to spread across organization through training of new and existing loggers. In fact, Weick & Quinn (1999) associates such behavior to organizational know-how, suggesting that tacit tasks such as schemas over time will become routine and taken for granted. Low accuracy of data is a consequence of customized practices within an organization, and which is likely to occur due to increased number of duplicates and grammar errors (Krishnan et al., 2016). The data cleansing process proved that this was present in the drug shortages dataset. Nonetheless, the quality of measures to decrease drug shortages will increase as the presence of high accuracy data will give valuable insight to implement suitable measures. An organization will tend to struggle with their strategies if the analytics is not conducted in an appropriate matter (Ritter, 2009). However, when the quality of measures is high, the actual number of drug shortages will decrease.

On the contrary, Low accuracy of data creates challenges for executing data exploration and cleaning effectively (Lintz, 2018). A clean dataset will contribute to increase overall productivity and provide valuable insights in the exploration phase. Furthermore, high quality of analytics will influence the level of insights provided in a positive direction (Provost & Fawcett, 2013). It is a common understanding in the world of analytics that increased quality of the input data will improve the quality of the output. Ultimately, sufficient insights will lead to enhanced decision-making within the organization. Provost & Fawcett (2013) discuss data-driven decision-making (DDD) and evaluates the fundamental concepts of the approach. They suggest that for DDD, data science processes are applied to draw causal conclusions, whereas inaccurate data or the presence of confounding factors could lead to illegitimate findings. Conclusively, a high level of standardization will, over time, contribute to decrease drug shortages.

5.5.1 Reinforcing loop 1 (R1)

The R1 loop is demonstrating how customized practices are being further spread into the organization. The term "Normalization of deviance" was coined by Vaughan (1996) to describe the process in which deviance from correct behavior becomes normalized in corporate culture. Hence, in this loop the spread of customized practices across organizations will, over time, converge to normalization of customized practices. In other words, employees within organization will presume that this is the way of doing this and will perceive customized practices as normal practice. Following, this will in the longer run be implemented as the "New Normal" within the organization (Pinto, 2014). The aftermath of this kind of behavior will have a negative effect on the actual level of standardization and increase number of drug shortages over time.

5.5.2 Balancing loop 1 (B1)

The B1 loop is goal-seeking and aims to balance the reinforcing negative effect on level of standardization (Forrester J. W., 1994). The loop starts out with a low level of standardization which has occurred as a result of the mechanism that plays out in the current system. These mechanisms will trigger a need and desired level of standardization as the current system is contributing to a larger gap between the actual level and the desired level of standardization. The needed and

desired level will eventually come as a result of an increase in the actual number of drug shortages which has occurred in the past.

As the deviation between the actual and the desired level of standardization increases, more awareness will be drawn to this as the threshold for tolerating the trend will be reached (Pfuhl & Henry, 1993). Furthermore, the awareness of the problem will harness solutions which will assumingly be through analytics and standardization (Willingham, 2006). This will, in return, have a positive effect and increase the actual level of standardization.

5.5.3 Balancing loop 2 (B2)

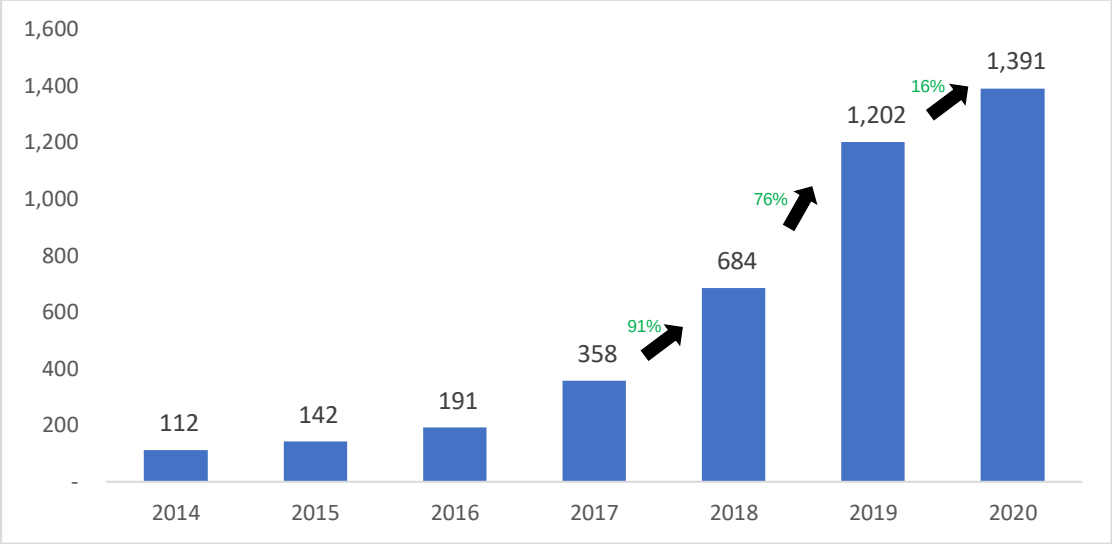
The B2 loop is goal-seeking loop in the diagram aims to reduce the actual number of drug shortages. The loop builds on a similar idea as balancing loop 1, which entails that SLV will develop a desired number of drug shortages due to the increasing trend, as shown in figure 5. As the ambition of any mitigation strategy regarding drug shortages would be to minimize the number of reports, the desired value is assumed to be as low as possible, presumably zero. The increasing actual number of drug shortages, which currently high, will continue to increase the deviation between the actual number and the desired number of drug shortages (Forrester J. W., 1994). There are many examples of urgent societal concerns regarding issues such as ecological devastation or human health, which creates an immediate demand for further knowledge (Fuchs, Blachenfellner, & Bichler, 2007). Should the number of reported drug shortages continue to rise to where it is unmanageable, SLV may adopt a similar perception and develop an urgent need for more knowledge. Considering that SLV already possess datasets of registered drug shortages, a possible further effect of the demand of knowledge will lead to initiation and implementation of analysis on these shortage datasets. In line with our reasoning regarding successful implementation of data-driven decision-making, if the analytical effort on the data is conducted in a proper matter, the actual number of drug shortages will decrease over time.

5.5.4 Summary of Causal loop diagram

The causal loop diagram displays several dynamics in play which contributes to both decrease and increase the number of drug shortages. The aftermath of low-

level standardization is assumingly reflected by the reinforcing loop (R1). This loop will continue to be the main driver of the increase of drug shortages due to the normalization of customized processes. This is also reflected in figure 5 where the increase in number of registered drug shortages by SLV in the past six years has been close to exponential. This seems to indicate that the reinforcing loop could be driving the behavior. Nevertheless, if SLV manages to focus and adopt the behaviors which are included in the balancing loops, the growth of drug shortages could potentially be slowed down in the future. This would be a possible outcome as the goal-seeking mechanism will tend to dominate within the organization. Nevertheless, as there may be other reinforcing factors causing the exponential growth, which are not included in this causal loop diagram, the adaption of behaviors in the balancing loops is not guaranteed to decrease the growth.

Figure 5 – Number of registered Drug shortages by SLV in the past six years



VI. Implications

6.1 Practical Implications

6.1.1 Analytics maturity

The concept of Analytics maturity models refers to different frameworks for measurement of an organization's analytics capabilities with regards to their overall understanding and adaptation (Harris & Davenport, 2018; Grossman, 2018; Gartner, 2018). The various models applied to determine the analytics maturity score the organizations based on multiple factors. These factors consist

of everything from the organizational structure in place, including the human resources, their approach to analytics and the available tools and experience, to aspect such as the quality and accessibility of the necessary data used in the analytics process. While the maturity score of the different frameworks provides an estimate of an organization's current analytical capabilities, it also contributes towards the understanding of how the organization could improve. As the various required factors are ranked, this presents an opportunity to prioritize which areas are in most need of improvement. Through application of the concept of analytics maturity in the scope of the technical knowledge framework by Bohn (1994), we believe the results justify our recommendations.

The framework of choice for our assessment of the analytics maturity regarding drug shortages in Norway is the DELTA Plus or DELTA TA Model (Davenport, 2018). The name of the model is an acronym, consisting of the elements regarded as most the critical factors for a successful analytics program: data, enterprise, leadership, targets, analysts, technology, and analytics techniques. The most pressing issue in this regard is the emphasis of this thesis, namely the data gathered from reported shortages. We argue that standardization of the input data to increase the data quality will consequently act as the enabler for other success factors. One such ensuing factor will be the deployment and use of analyst. Improvements of the input data should accelerate analytic efforts and employment of analysts with the required expertise to make use of the improved data. Correspondingly, employment and delegation of specialized analysts will induce the application of new analytics techniques and new technology. The priority of these elements is in our opinion essential to establish the required foundation to decide the suggested focus areas and actions for the remaining success factors of the Delta Plus model: The enterprise, leadership, and targets. These three factors perfectly correspond the health policymakers in Norway, their targets, and the actions they take in the fight against drug shortages. With regards to maturity of each factor, an organization ranks as analytically mature when the leaders support analytical efforts and analytics is integral to the organization's strategy. Correspondingly to our proposal that implementation and application of analytics is required to progress in terms of the technical knowledge framework of Bohn (1994), suggesting that the analytics maturity and the technical knowledge level will increase in parallel. Alternatively, further development of the analytics

maturity can also be seen as required step towards an increased knowledge level. Regardless of the perception for how the concepts relates, the conditional factor for growth relies on an improved data structure.

The causal loop diagram in Figure 4 illustrates the causal relationships in the current reporting system. Link 24 from this diagram depicts a causal relationship between an unnamed variable and the actual level of standardization in the reporting system. This unspecified variable is meant to represent all the various factors that impacts an organization's efforts toward standardization for analytical application. Although it can be argued that the level of standardization is affected by other factors, some of the most important factors will undoubtably be the same success factors as the ones considered for the analytics maturity. Hence, if subscribed to the concept of analytics maturity, this suggests that this unnamed variable could in fact be replaced by the analytics maturity of the organization. This implies that when the maturity increases, so will the actual level of standardization. Consequently, while the maturity remains unchanged, as to will the level of standardization. The ripple-effects of increased level of maturity will be exposed in the sense that increased level of standardization will facilitate further development for all of the other elements of the system, as depicted in the causal loop diagram.

6.1.2 Data structure

In order for an organization to fully utilize and harness the power of analytics to their day-to-day, it is essential to implement a reliable data structure which is appropriate for data analytics (Schuh & Blum, 2016). This will further contribute to assist for a better use of the company's resources and assist the employees to conduct more standardized approaches. In the following section, we will introduce a possible data structure for SLV to keep track of drug shortages.

6.1.2.1 Database

As mentioned in previous sections, the current reporting system for drug shortages has proved to be ineffective and lacks a systematic approach. The challenge with this approach is that there are many repetitive inputs in the different columns, and these also contain grammar errors which creates several unnecessary data points. Database normalization is the process of structuring a

database in accordance with a series of so-called normal forms to reduce data redundancy and improve data integrity (Beeri, Bernstein, & Goodman, 1989). The technique is a process of structuring a database's columns (attributes) and tables (relations) such that database integrity constraints can enforce their dependencies appropriately. This is done by following a set of formal rules, either via synthesis (forming a new database design) or decomposition (breaking down an existing database design). SLV assumingly do not have a clear database design which led us to construct a new database design based on the cleansed data from section 3.3.

6.1.2.2 Database Design and modeling

Figure 6 – Database Design

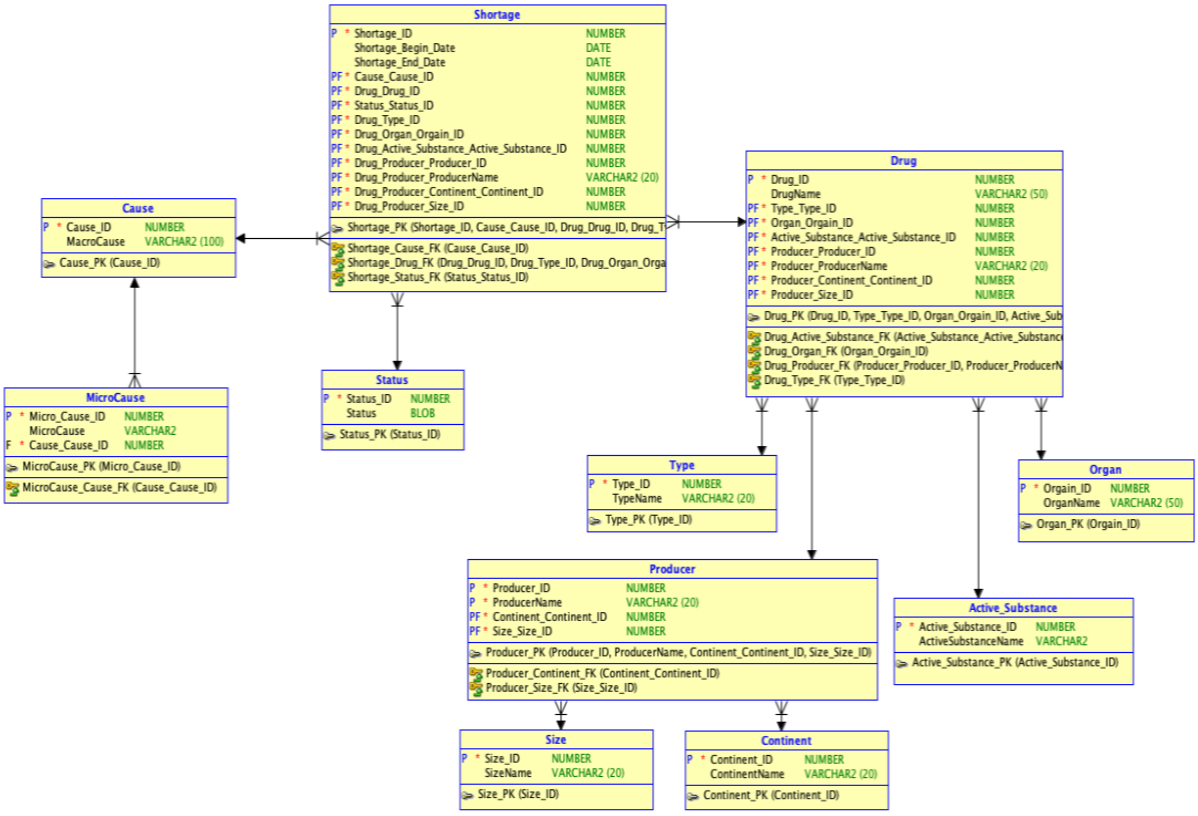


Figure 6 demonstrates a possible database design for SLV. The modeling follows numerous sets of rules which are general for Normalization. The first step of the modeling was to convert the data into the first Normal form (1NF). In order for this form to be valid, two rules have to be fulfilled:

1. The cell of each table should contain maximum one value.
2. Each record needs to be unique.

The original dataset given by SLV did not have any of these rules present which naturally means that the dataset was unstructured and difficult to conduct analytics on. In our cleansing process from chapter *3.3 Data Cleansing and Preparation,* we manage to transform the dataset into the first normal form, the basis dataset for the cluster analysis, which ultimately formed the minimum requirement for database design. One of the main advantages of 1NF is that it avoids unneeded recurring values, which eliminates any issues with adding, removing, and changing entries in the database. Furthermore, the second form of normalization (2NF) is only applicable if the requirements for 1NF is already fulfilled. This is given in the rules for 2NF which are:

1. The dataset must already be in 1NF.
2. No Partial dependency.

Partial dependency is present when a non-prime attribute is functionally dependent on part of a candidate key. An example of this in our 1NF design is that there were many columns in the same table that was not necessary to have, and they all were depended on the observation ID. For instance, if SLV wanted to get the cause of a given specific drug at a given time, this would have not been feasible because the drug may be linked to different causes and the same the drug may diverse variants in terms of size and dosages. In simpler terms, a query would struggle to return the desired output because the drug is not fully dependent on one column, but many different columns. Nevertheless, there are many ways to convert 1NF to 2NF, and our chosen method was to create several tables where partial dependencies were avoided. One example of this is the Drug table which is linked to other sub-tables by Many-to-one relationships. For instance, this suggests that a specific drug can only belong to one producer, while a producer can be linked to multiple different drugs. Accordingly, a specific drug can only be of one type, but the type can cover several drugs. By implementing such logical relations in the database design, we avoid partial dependencies. Consequently, each column on every table becomes depended on only one primary key, which provides full dependency.

Lastly, a third form of normalization (3NF) was applied in order to give the database design a practical and effective structure. The rules for the third form

normalization are given as:

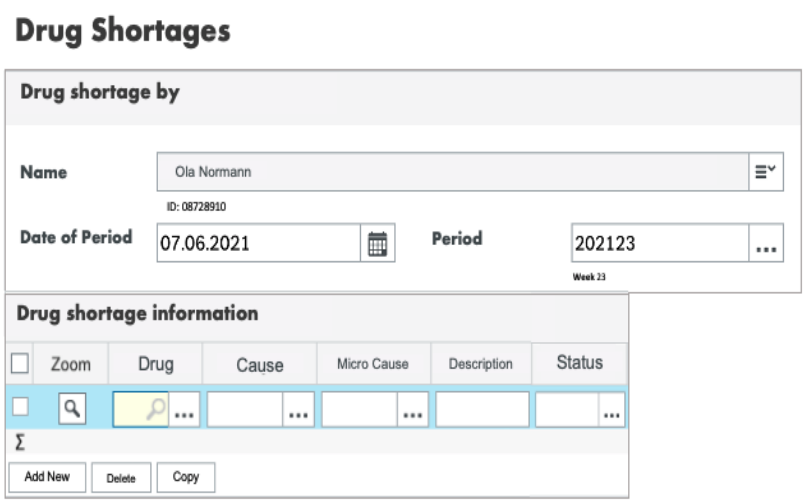
1. It should be in the Second Normal form (2NF)
2. No transitive dependency.

Transitive Dependency is present when a non-prime attribute is depending on other non-prime attributes. For example, in figure 6, the table “Producer” is linked two other tables below which are “Size” and Continent”. Originally, in our 2NF, these tables were not made and were given as column inside the “Producer” table. In this case, we had transitive dependency because both continent and size columns (non-prime attributes) where depended on producer name (another non-prime attributes). This was solved to a 3NF by constructing the two tables and placing their primary key as foreign key in the producer table. The benefits of breaking the database design to a 3NF is that the amount of data duplication becomes significantly reduced and date integrity is achieved. This is an advantage and relevant for effective analytics.

6.1.2.3 Database Graphical User-interface

Sensible and friendly graphical user-interface (GUI) is essential for any businesses in order to increase efficiency amongst employees (Jansen, 1998). The implementation of the database design would contribute to improve the logging of drug shortages as a result of the normalization. To illustrate how this could be a practical solution for SLV, we have designed a Graphical User Interface that is coherent with the database design.

Figure 7 – Graphical User Interface for Drug shortages



The current reporting system has 14 columns that needs to be filled out for a single drug shortage. The normalization of the database design has contributed to shorten this into 5 columns[3] in comparison. The proposed GUI has several properties which reflects the structure of the database. An example is the column "Drug", which provides information about all the other columns that are connected to it in the structure. This means that the logger does not have to manually fill out the active substances, producer or any other column which is linked to the "Drug" table. This will automatically be embedded in the chosen drug. Essentially, the only information required by the logger is the cause and micro-cause of shortage, a description of the shortage and the status. The presence of accountability is key for functional and effective flow of work. This is the background of including the employee's name and ID for each log of shortage. It important that the logging is being executed in the appropriate method. Deviance will usually lead to insufficient data points which goes against the purpose of the database design and the GUI. Hence, by having assigned each shortage to the logger, it makes simpler to inform the logger if any mistakes where to occur.

The GUI is a demonstration for improving the effectiveness of logging and at the same time, building a database of information which can be used for analytics. Other relevant columns can be easily added to the database design and GUI for more informative observations.

6.2 Theoretical implications

[3] Name, Date of Period and Period will be filled in automatically.

While the extent of literature on the field of drug shortages has remained scarce for a long time, the topic has recently seen a gradual rise in attention. However, tendencies among the published literature suggests that most research thus far remain as case-studies and mainly related to the causes of shortages (e.g., Pauwels et al., 2014; De Weerdt et al., 2015; Aguiar & Ernest, 2020). Among the studies that explore the various causes, a great deal appears to base its conclusions on calculations regarding the number of cases and their respective basis of reporting, such as the sum of reports with the same stated cause. Little or no existing literature goes beyond these simple calculations to further explore trends, correlations, or predictions based on thorough analysis. Multiple studies refer to potential mitigation strategies based on careful planning (e.g., Ventola, 2011; Fox, et al., 2009). While the arguments made for this method appear valid, the existing literature fails to clarify or demonstrate how the planning should be accomplished. Although this thesis does not explicitly explore this planning procedure, the study contributes to the existing literature as another step towards to realization of this mitigation strategy.

Our research suggests that planning in terms of decision-making by health policymakers will be facilitated through application of comprehensive analysis. The findings from our cluster analysis indicate that available data contains valuable information for strategical decisions and mitigation planning. However, the quality of the data requires extensive preparation for such analysis to take place. For continuous and further analysis of increased accuracy, our discoveries imply that improved data management is essential. The field of system dynamics is introduced in the thesis through the application of a causal loop diagram (CLD) to illustrate the current reporting process, and the causal relationship between the various factors. The findings from the CLD reveal some very important discoveries, that emphasizes the need for change in the data management. The current process implies a system that encourages customized approaches, which may continue to increase the problem of drug shortages rather than solving it. However, more standardized procedures and increased data quality will lead to improved insights from analytics efforts and subsequent decision-making.

The thesis makes creative use of Bohn's (1994) framework for measurement of technological knowledge, as a method of demonstrating the opportunities and

benefits that follows knowledge growth. The framework categorizes knowledge based on its display, from where the lower levels display little or no application of knowledge, to the higher levels in which knowledge is applied in coding, equations, and formulas. As such, although the framework was originally designed for the production process, the method works well with respect to data-driven decision-making. From the perspectives of business analytics, the approach suggests that decision-making for health policymakers should be data-driven (Provost & Fawcett, 2013). Throughout the entirity of our research process, the reoccuring topic and fundemental requirement for such decision-making is the necessary standardization. Our findings implies that the current knowledge level regarding drug shortages in Norway is a results of the lack of standardization, and that progression towards a higher level will require an increased efforts in this regard (Bohn, 1994).

Our research displays how digital tools applied in the reporting process for shortages and the subsequent data management may counteract the lack of standards, thereby promoting further climbing of the ladder of knowledge. Existing literature implies that the effectiveness of prevention and mitigation strategies will rely on the collected infomration (Fox, et al., 2009). Enhanced data management and automation will also enable further mitigation strategies, such as a potential warning-system, as suggested by Ventola (2011).

Our study is confined to explore the impact of our research question for the Norwegian health system. The decision to limit our research specifically to Norway is rooted in several factors. First, the lack of a uniform international definition for drug shortages has limited the possibility of valid international comparisons, resulting in most of the existing literature relying on country or area-specific case studies (e.g., Gatesman & Smith, 2011; Pauwels et al., 2014). Second, Norway represents a resourceful country with a strong economy and relatively small population of approximately 5,4 million (SSB, 2020). Furthermore, Norway has consistently ranked among the top countries in the world with regards to healthcare (The Lancet, 2017; OECD, 2019). Based on this rationale, we find it reasonable to assume that Norway makes for a good example in terms of reporting of drug shortages, and that the findings of our research may also be applicable to other countries of similar prerequisites.

6.2.1 Future research and limitations

The thesis relies on a few assumptions that can be considered as limitations of the study. SLV suggests that the pandemic of Covid-19 has shown little impact on the of drug shortage situation (SLV, 2020b). Future research could explore this presumption as more data becomes available. While this study is focused on shortages related to medicinal products for human consumption, it is assumed that the findings also apply to veterinary products, as these are reported to SLV through the same system. Hence, future research could apply the same or similar methods as for this thesis on the category of veterinary drugs to explore and compare the results. Furthermore, as the data quality for reported shortages of both categories improves, future studies could apply comparative analysis on the categories to identify potential correlations or possible predictions based on either category. The thesis' replicability can be questioned with regards to number of assumptions required to perform the necessary data cleansing prior to the performed cluster analysis. As such, efforts to replicate the results may achieve different results to the once presented in this paper. Despite the possible discrepancies between the results of our analysis and attempted replications, the overall recommendations and findings from the thesis remain.

The assumption of invalidated international comparisons is applied in the reasoning as to why the thesis is concentrated on a single country. However, efforts have been made by researchers to explore the results of countries in which the definitions are similar enough to support a basis for comparison (Bogaert et al., 2015). As such, at the time when Norway follows a uniform definition for reported drug shortages, future research could explore the contrasts of the reports to those of other countries in which the definitions are harmonized enough to facilitate comparisons.

Finally, the findings of this thesis are assumed to be applicable for countries with similar prerequisites in terms of aspects such as population size and level of healthcare. Future research could apply similar studies to such countries to evaluate this assumption or explore whether the findings of the research will apply to countries of contrasting prerequisites to Norway.

VII. Conclusion

The aim of the thesis was to review the information currently available to Norwegian health policymakers concerning drug shortages, and evaluate to what degree practises and perspectives from the field of business analytics can be applied to assist in their decision-making. This allowed us to formulate the following research question:

"How can business analytics methods be applied to help health policymakers in Norway in the fight against drug shortages?"

Our study highlights the relationship between knowledge levels and the accuracy of strategies supported by data-driven decision-making. For health policymakers, increased knowledge levels will allow for application of knowledge in the coding, equations, and formulas used to control variables and increase the accuracy of the results. By categorizing the reported shortages using a cluster analysis, we illustrate that the currently available data does supports decision-making to an extent, while increased accuracy of such analysis will require increased data quality. Our study implies that the lack of standardization offered by the current reporting system prevents knowledge growth. Our use of the causal loop diagram displays how this will impact the situation regarding drug shortages, and why adjustments are essential. Consequently, alternative changes as our suggested implementation of digital reporting tools, data management systems and relational database design will enable enhanced data quality to facilitate analysis and subsequent climb of knowledge levels.

In conclusion, our findings suggests that standardization is the first step towards a higher knowledge level, of which the enabled analysis will continue the climb. Consequently, a higher knowledge level will result in more accurate results, which enables more efficient and precise measures for health policymakers to tackle the problem of drug shortages.

VIII. Bibliography

Abbott, K. W., & Snidal, D. (2001). International 'standards' and international governance. *Journal of European Public Policy 8:3 Special Issue*, 345-370.

Accenture. (2017). *Technology for Patients: The New Age of Human Empowerment*. Retrieved from Accenture.com: https://www.accenture.com/t20180320T075055Z__w__/us-en/_acnmedia/Accenture/next-gen-5/insight-tech-life-sciences-2017/Life-Sciences-Tech-Vision-2017_ONLINE.pdfla=en#zoom=50

Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons, Volume 51, Issue 5*, 565-570.

Acosta, A., Vanegas, E. P., Rovira, J., Godman, B., & Bochenek, T. (2019). Medicine Shortages: Gaps Between Countries and Global Perspectives. *Frontiers in Pharmacology (10)*, 763.

Aguiar, E., & Ernest, K. (2020). Letter to the editor - Factual inaccuracies of the article Causes of drug shortages in the legal pharmaceutical framework. *Regul Toxicol Pharmacol. vol 111*.

Beeri, C., Bernstein, P. A., & Goodman, N. (1989). A sophisticate's introduction to database normalization theory. *Readings in Artificial Intelligence and Databases*, 468-479.

Besancon, L., & Chaar, B. (2013). *2013 Summit on Medicines Shortages.* Toronto: FIP (International Federation of Pharmacists).

Bochenek, T., Abilova, V., Alkan, A., Asanin, B., Beriain, I. D., Besovic, Z., . . . Gulbun. (2018). Systemic Measures and Legislative and Organizational Frameworks Aimed at Preventing or Mitigating Drug Shortages in 28 European and Western Asian Countries. *Frontiers in Pharmacology*.

Bocquet, F., Degrassat-Theas, A., Peigne, J., & Paubel, P. (2017). The new regulatory tools of the 2016 Health Law to fight drug shortages in France. *Health Policy, volume 121, Issue 5*, 471-476.

Bogaert, P., Bochenek, T., Prokop, A., & Pilc, A. (2015). A Qualitative Approach to a Better Understanding of the Problems Underlying Drug Shortages, as Viewed from Belgian, French and the European Union's Perspectives. *PloS ONE 10(5)*.

Bohn, R. E. (1994). Measuring and Managing Technological Knowledge. *Sloan Management Review, 36;1*, 61-73.

Bonsón, E., & Escobar, T. (2006). Digital reporting in Eastern Europe: An empirical study. *International Journal of Accounting Information Systems, Volume 7(4)*, 299-318.

Braa, J., & Sahay, S. (2012). *Integrated Health Information Architecture: Power to users.* New Dehli: Matrix Publishers.

Braman, S. (1989). Defining information: An approach for policymakers. *Telecommunications Policy, Volume 13, Issue 3*, 233-242.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011, April 24). *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?* Retrieved from SSRN: https://ssrn.com/abstract=1819486

Busemeyer, J. R., & Pleskac, T. J. (2009). Theoretical tools for understanding and aiding dynamic decision making. *Journal of Mathematical Psychology Volume 53, Issue 3,* 126-138.

Büyüközkan, G., & Göçer, F. (2018). Digital Supply Chain: Literature review and a proposed framework for future research. *Computers in Industry, Volume 97,* 157-177.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, Vol. 35, No. 4,* 1165-1188.

Davenport, T. (2018). DELTA Plus Model & Five Stages of Analytics Maturity: A Primer. *International Institute for Analytics*.

De Weerdt, E., Simoens, S., Casteels, M., & Huys, I. (2015). Toward a European definition for a drug shortage: a qualitative study. *Frontiers in Pharmacology, vol 6,* 253.

De Weerdt, E., Simoens, S., Casteels, M., & Huys, I. (2017). Clinical, Economic and Policy Implications of Drug Shortages in the European Union. *Applied Health Economics and Health Policy*, 441-445.

De Weerdt, E., Simoens, S., Hombroeck, L., Casteels, M., & Huys, I. (2015). Causes of drug shortages in the legal pharmaceutical framework. *Regulatory Toxicology and Pharmacology, v.71(2),* 251-258.

Diedrichsen, E., Waning, B., Jambert, E., Bärnighausen, T., Li, Y., Pouw, M., & Moo, S. (2010). The global pediatric antiretroviral market: analyses of product availability and utilization reveal challenges for development of pediatric formulations and HIV/AIDS treatment in children. *BMC Pediatrics volume 10, Article number: 74* .

Dobrev, S., Kralovic, R., & Kralovic, R. (2013). Computing with advice: when knowledge helps. *Bulletin of EATCS, 2(110)*.

Dunning, D. (2011). Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance. *Advances in Experimental Social Psychology, Volume 44,* 247-296.

Dvir, D., Lipovetsky , S., Shenhar, A. J., & Tishler, A. (2003). What is really important for project success? A refined, multivariate, comprehensive analysis. *International Journal of Management and Decision Making, 4.4,* 382-404.

EMA. (2021, June 15). Who we are. Retrieved from European Medicines Agency: https://www.ema.europa.eu/en/about-us/who-we-are

EMA-HMA. (2019). *Guidance on detection and notification of shortages of medicinal products for Marketing Authorisation Holders (MAHs) in the Union (EEA).* Amsterdam: European Medicines Agency & Heads of Medicines Agencies.

EU Pharma Ltd. (n.d.). *Parallel export and import*. Retrieved April 2020, from eupharmaltd.com: http://eupharmaltd.com/en/parallel-export-and-import/

FDA. (2011). *A review of FDA's approach to medical product shortages.* United States Food and Drug Administration.

FDA. (2019). *Report on Drug Shortages for Calendar Year 2019.* U.S. Food and Drug Administration, Department of Health and Human Services .

FDA. (2020). *Federal Food, Drug, and Cosmetic Act (FD&C Act).* Food and Drugs Administration.

FDA. (2021, March 16). *FDA.gov*. Retrieved from Generic Drugs: Questions and Answers: https://www.fda.gov/drugs/questions-answers/generic-drugs-questions-answers

Federgruen, A. (2012, March 1). *The Drug Shortage Debacle—And How to Fix It.* Retrieved from Wall Street Journal: https://www.wsj.com/articles/SB10001424052970203986604577253242913478400

Forrester, I., & Dawes, A. (2008). Parallel trade in prescription medicines in the European Union: the age of Reason? *Yearbook of Antitrust and Regulatory Studies (YARS), Issue 1*, 9-31.

Forrester, J. W. (1994). System dynamics, systems thinking, and soft OR. *System Dynamics Review (10)*, 245-256.

Fox, E. R., & Tyler, L. S. (2017). Potential Association between Drug Shortages and High-Cost Medications. *The Journal of Human Pharmacology and Drug Therapy, Volume37, Issue1*, 36-42.

Fox, E. R., Birt, A., James, K. B., Kokko, H., Salverson, S., & Soflin, D. L. (2009). ASHP Guidelines on Managing Drug Product Shortages in Hospitals and Health Systems. *American Journal of Health-System Pharmacy, Volume 66, Issue 15*, 1399–1406.

Fuchs, C., Blachenfellner, S., & Bichler, R. (2007). The urgent need for change: Rethinking knowledge and management. *In Knowledge Management: Innovation, Technology and Cultures*, 293-307.

Gartner. (2018). *Gartner Survey Shows Organizations Are Slow to Advance in Data and Analytics.* Press Release: Gartner.

Gatesman, M. L., & Smith, T. J. (2011). The Shortage of Essential Chemotherapy Drugs in the United States. *The New England Journal of Medicine: 365*, 1653-1655.

Gray, A., & Manasse, H. R. (2012). Shortages of medicines: a complex global challenge. *Bulletin of the World Health Organization, Volume 90, Number 3*, 157-244. Retrieved from Bulletin of the World Health Organization: https://www.who.int/bulletin/volumes/90/3/11-101303/en/

Grossman, R. L. (2018). A framework for evaluating the analytic maturity of an organization. *International Journal of Information Management Volume 38, Issue 1*, 45-51.

Harrington, J. L. (2009). *Relational Database Design and Implementation : Clearly Explained.* Elsevier Science & Technology.

Harris, & Davenport. (2018). *DELTA Plus Model & Five Stages of Analytics Maturity: A Primer.* International Institute for Analytics.

Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management, Volume: 4, Issue 2*, 168-193.

Helsedirektoratet. (2019, June 21). *Nasjonal legemiddelberedskap: Vurderinger og anbefalinger.* Retrieved from Helsedirektoratet: https://www.helsedirektoratet.no/rapporter/nasjonal-legemiddelberedskap

ICAEW. (2004). *Digital reporting: a progress report.* Institute of Chartered Accountants in England & Wales.

Jahre, M., van Oorschot, K., Soldal, O. B., Viana, J., & Årdal, C. (2021). *Explorative Interviews Norwegian Stakeholders.* Oslo: MIA - Measures for Improved Availability of Medicines and Vaccines.

Jansen, B. J. (1998). The graphical user interface. *ACM SIGCHI Bulletin, Volume 30, Issue 2*, 22-26.

Jenks, S. (2011). Efforts Underway To Curb Drug Shortages. *Journal of the National Cancer Institute, Volume 103, Issue 12*, 914-915.

Johnson, T. (2011). Drug shortages: an increasing problem for patients and clinicians. *the Journal of the South Dakota State Medical Association, 64(1)*, 14-15.

Karim, R., Westerberg, J., Galar, D., & Kumar, U. (2016). Maintenance Analytics – The New Know in Maintenance. *IFAC-PapersOnLine Volume 49, Issue 28*, 214-219.

Khan, F. A. (2019). Anesthetic drugs shortage in lower and middle income countries: a safety and quality issue. *Anesthesiology, Pain & Intensive Care, Vol. 23 (4)*, 337-339.

Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. *Communications of the ACM, Volume 45, Issue 8, Evolving data mining into solutions for insights*, 5-106.

Krishnan, S., Wang, J., Franklin, M. J., Goldberg, K., Kraska, T., Milo, T., & Wu, E. (2015). SampleClean: Fast and Reliable Analytics on Dirty Data. *IEEE Data Eng. Bull., 38(3)*, 59-75.

Krishnan, S., Wang, J., Wu, E., Franklin, M. J., & Goldberg, K. (2016). ActiveClean: interactive data cleaning for statistical modeling. *Proceedings of the VLDB Endowment, Volume 9, Issue 1*, 948–959.

Lasi, H., Fettke, P., Kemper, H.-G., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering, volume 6*, 239-242.

Legemiddelverket. (2020, June 16). *Legemiddelverket.no*. Retrieved May 2021, from Godkjenning av legemidler: https://legemiddelverket.no/godkjenning/godkjenning-av-legemidler/avregistrering-melding-om-legemiddelmangel-og-reaktivering-av-vare#melding-om-legemiddelmangel

Lintz, J. (2018). A Qualitative Study of Master Patient Index (MPI) Record Challenges from Health Information Management Professionals' Perspectives. *Perspectives in Health Information Management* , 1-13.

Madakam, S., Ramaswamy, R., & Tripathi, S. (2015). Internet of Things (IoT): A Literature Review. *Journal of Computer and Communications, Vol.03 No.05*.

Muljani, N., & Ellitan, L. (2019). Developing Competitiveness in Industrial Revolution 4.0 . *International Journal of Trend in Research and Development., 6 (1),* 1-3.

OECD. (2019). *Health at a Glance 2019: OECD Indicators.* Paris: OECD Publishing.

Osgood, W. D., Wilson, J. K., O'Malley, P. M., Bachman, J. G., & Johnston, L. D. (1996). Routine Activities and Individual Deviant Behavior. *American Sociological Review, Vol. 61, No. 4,* 635-655.

Pauwels, K., Huys, I., Casteels, M., & Simoens, S. (2014). Drug shortages in European countries: a trade-off between market attractiveness and cost containment? *BMC Health Services Research, 438*.

Pfuhl, E. H., & Henry, S. (1993). *The deviance process.* New York: Transaction Publishers.

Pinto, J. K. (2014). Project management, governance, and the normalization of deviance. *International Journal of Project Management, Volume 32, Issue 3*, 376-387.

Popovic, A., Hackney, R., Coelho, P. S., & Jaklic, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. *The Journal of Strategic Information Systems, Volume 23, Issue 4*, 270-283.

Prielipp, R. C., Magro, M., Morell, R. C., & Brull, S. J. (2010). The Normalization of Deviance: Do We (Un)Knowingly Accept Doing the Wrong Thing? *Anesthesia & Analgesia: Volume 110, Issue 5*, 1499-1502.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data, Vol. 1, No. 1*, 51-59.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems volume 2, Article number: 3*.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM, Volume 41, Issue 2*, 79-82.

Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. . *Psychological methods, 17 3*, 354-73.

Ritter, A. (2009). How do drug policy makers access research evidence? *International Journal of Drug Policy, 20,* 70–75.

Rohrmann, B. (1986). *Evaluating the usefulness of decision aids: A methodological perspective.* Amsterdam: B. Brehmen, H. Jungermann, P Lourens, & G. Sevon (Eds.).

Roos, G., & Roos, J. (1997). Measuring your Company's Intellectual Performance. *Long Range Planning, Special Issue on Intellectual Capital, Vol. 30, No. 3*, 413-426.

Schuh, G., & Blum, M. (2016). Design of a data structure for the order processing as a basis for data analytics methods. *Portland International Conference on Management of Engineering and Technology (PICMET)*, 2164-2169.

Simons, R. (1987). Accounting control systems and business strategy: An empirical analysis. *Accounting, Organizations and Society, 15(1–2),* 127-143.

Simons, R. (1994). *Levers of Control: How managers use innovative control systems to drive strategic renewal.* Harvard Business School Press.

SLV. (2014). *Årsrapport 2014.* Statens legemiddelverk.

SLV. (2020a, June 21). *Legemiddelmangel og avregistreringer tidligere år*. Retrieved from Statens legemiddelverk: https://legemiddelverket.no/legemiddelmangel/legemiddelmangel-og-avregistreringer-tidligere-ar

SLV. (2020b). *Årsrapport 2020.* Statens legemiddelverk.

SNL. (2020, August 11). *Store Norske Leksikon*. Retrieved from Store Medisinske Leksikon : https://sml.snl.no/ATC-systemet

SSB. (2020). *Befolkning.* Oslo: Statistics Norway.

Sterman, J. (2000). *Business Dynamics: Systems thinking and modelig for a complex world.* Homewood: Irwin/McGraw-Hill.

Strömgren, E. (2021). *SAS Workshop - Analytics Use Case Cancas II.* Oslo: SAS.

Sweet, B. V., Kaakeh, R., Reilly, C., Bush, C., Deloach, S., Higgins, B., & Stevenson, J. (2011). Drug Shortages: A Complex Health Care Crisis. *American Journal of Health-System Pharmacy, Volume 68, Issue 19,* 1811–1819.

The Lancet. (2011). Shortages of cancer drugs in the USA. *The Lancet Oncology, Volume 12, Issue 4,* 33.

The Lancet. (2017). Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015. *ISSN: 0140-6736, Vol: 390, Issue: 10091,* 231-266.

Todnem By, R. (2005). Organisational change management: A critical review. *Journal of Change Management, 5(4),* 369-380.

Torreya. (2020). *Top Global Pharmaceutical Company Report: The Pharma 1000.* Torreya Capital LLC.

Tucker, E. L., Cao, Y., Fox, E. R., & Sweet, B. V. (2020). The Drug Shortage Era: A Scoping Review of the Literature 2001–2019. *Clinical pharmacology and therapeutics, Vol.108 (6),* 1150-1155.

Valentinetti, D., & Muñoz, F. F. (2021). Internet of things: Emerging impacts on digital reporting. *Journal of Business Research*.

Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture and deviance at NASA.* Chicago: University of Chicago.

Ventola, C. (2011). The Drug Shortage Crisis in the United States. *Pharmacy and Therapeutics, 36(11),* 740-757.

Vessey, I. (1994). The effect of information presentation on decision making: A cost-benefit analysis. *Information & Management Volume 27, Issue 2,* 103-119.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems, Volume 12, Issue 4,* 5-33.

Weber, M. (1987). Decision making with incomplete information. *European Journal of Operational Research, Volume 28, Issue 1*, 44-57.

Weick, K. E., & Quinn, R. E. (1999). Organizational change and development. *Annual Review of Psychology, Vol. 50*, 361-386.

WHO. (2015). *Technical Consultation on Preventing and Managing Global Stock Outs of Medicines.* Geneva: World Health Organization.

WHO. (2016). *Addressing the global shortage of medicines and vaccines – WHA69.25.* Geneva: World Health Organization.

Willingham, D. T. (2006). How knowledge helps: It speeds and strengthens reading comprehension, learning-and thinking. *American Educator, 30, 1* , 30.

8.1 Table of figures

8.2 Table of tables