Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Thomas Akselberg Hatlebrekke

Use of Association Rule Mining to Identify Trigger Foods in Irritable Bowel Syndrome

Master's thesis in Applied Computer Science
Supervisor: Dr. Sule Yildirim Yayilgan
June 2019

NTNU
Norwegian University of Science and Technology

Thomas Akselberg Hatlebrekke

Use of Association Rule Mining to Identify Trigger Foods in Irritable Bowel Syndrome

Master's thesis in Applied Computer Science
Supervisor: Dr. Sule Yildirim Yayilgan
June 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of
Science and Technology

Use of Association Rule Mining to Identify Trigger Foods in Irritable Bowel Syndrome

Thomas Akselberg Hatlebrekke

01-06-2019

Master's Thesis
Master of Science in Applied Computer Science
30 ECTS
Department of Computer Science
Norwegian University of Science and Technology,

Supervisor: Dr. Sule Yildirim Yayilgan

Acknowledgment

First and foremost, I would like to thank my supervisor, Dr. Sule Yildirim Yayilgan, for continuous and invaluable guidance throughout the work with this thesis.

The results of the thesis would not have been the same without the help from Birgitte Berentsen, manager of "National competency service for functional gastrointestinal diseases", who provided important professional expertise and made the patient recruitment possible. I also want to thank Anita Lyssand in Helse Bergen for giving me the opportunity to combine work with taking a master's degree. Lastly, my gratitude goes out to the patients who participated in the study.

T.A.H.

Abstract

This work applies association rule mining (ARM) to dietary- and symptom registrations collected from six irritable bowel syndrome (IBS) patients over periods of 12 to 16 days. The Apriori algorithm is used to generate frequent itemsets from the data and discover association rules identifying relationships between foods and IBS symptoms. The results were interpreted by six clinicians from "National competency service for functional gastrointestinal diseases" at Haukeland University Hospital.

The results and interpretations from the clinicians show that relationships between foods and IBS symptoms can be identified using ARM, and that the approach can be useful in a decision support system for clinical nutritionists in consulting IBS patients.

Contents

Acknowledgment . . . . **i**
Abstract . . . . **ii**
Contents . . . . **iii**
List of Figures . . . . **v**
List of Tables . . . . **vi**
1 Introduction **1**
1.1 Topic covered by the project 1
1.2 Keywords 1
1.3 Problem description 1
1.4 Motivation and benefits 2
1.5 Research questions 2
1.6 Contributions 2
1.7 Clarification of terms and acronyms 2
1.8 Ethical and legal considerations 2
2 Background **4**
2.1 Association rule mining 4
2.1.1 Objective interestingness measure of association rules 4
2.1.2 Apriori algorithm 5
2.2 Related work 6
2.2.1 IBS and the role of diet 6
2.2.2 Identifying relationships between IBS symptoms and diet 7
2.2.3 Use of ARM in medicine 8
2.2.4 Food logging 9
2.2.5 Summary and challenges 9
3 Methodology **11**
3.1 Data collection 11
3.1.1 Recruiting IBS patients 11
3.1.2 Design and implementation of logging system 14
3.1.3 Logging via mobile application 16
3.2 Data processing 18
3.2.1 Transforming collected data to input datasets 18
3.2.2 Discovering association rules 21
3.2.3 Extracting interesting association rules 22
3.3 Interpretation of association rules with domain experts 23

3.4 Tools and frameworks . . . 24
4 Experiments and results . . . 25
4.1 Filtering discovered association rules . . . 25
4.2 Deciding on input parameters . . . 26
4.3 Deciding time interval in registration periods . . . 26
4.4 Experiment 1 . . . 27
4.5 Experiment 2 . . . 28
4.6 Experiment 3 . . . 29
4.7 Feedback from domain experts . . . 31
4.7.1 Feedback on design of the study . . . 32
4.7.2 Feedback on value of association rules . . . 32
5 Discussion . . . 34
5.1 Collecting data . . . 34
5.1.1 Amount of collected data . . . 34
5.1.2 Recruitment . . . 35
5.1.3 Quality of collected data . . . 35
5.1.4 Usability of the logging system . . . 36
5.1.5 Additional predictors . . . 37
5.2 Discovery of association rules . . . 38
5.2.1 Deciding length of registration periods . . . 38
5.2.2 Deciding on input parameters . . . 38
5.2.3 Presentation of discovered association rules . . . 38
5.2.4 Involvement of domain experts . . . 39
6 Conclusion . . . 40
Bibliography . . . 41
A Information e-mail . . . 47
B Instruction e-mail . . . 49
C Final e-mail . . . 52
D REK approval . . . 53
E Association rules of format ($X \Rightarrow y$) . . . 57

List of Figures

1 Architecture of the logging system, from a patient logging dietary intake and IBS symptoms, to domain experts interpreting the discovered association rules. 12
2 Flowchart of the patient recruitment process. 13
3 Screenshots of the mobile application. **Upper left:** menu. **Upper middle:** search and selection of food items. **Upper right:** registration of selected food item and time of consumption. **Lower left:** registration of symptom strength and time of occurrence. **Lower middle:** history of previous registrations. 15
4 Simplified overview of entity relationships in the database. 16
5 Illustration of how a timeline of registrations from an IBS patient was converted to registration periods, using different time intervals. 20
6 Compressed sample of a data frame representing registered food groups and symptoms 22
7 Compressed sample of a data frame representing registered food names and symptoms 22
8 Association rules indicating relationships between food groups and IBS symptoms for all patients. 28
9 Association rules indicating relationships between food names and IBS symptoms for all patients. 29
10 Association rule indicating a relationship between "Sweets mix, without chocolate" and IBS symptoms for patient 1. 31
11 Association rules indicating relationships between food names and IBS symptoms for patient 2. 31
12 Association rules indicating relationships between food names and IBS symptoms for patient 4. 31
13 Top 20 association rules of the format $(X \Rightarrow y)$ from experiment 2. 57
14 Top 20 association rules of the format $(X \Rightarrow y)$ from experiment 3. 58
15 All association rules of the format $(X \Rightarrow y)$ from experiment 4. Patient 1. 58
16 Top 20 association rules of the format $(X \Rightarrow y)$ from experiment 4. Patient 2 59
17 Top 20 association rules of the format $(X \Rightarrow y)$ from experiment 4. Patient 4. 60

List of Tables

1 Overview of IBS patients completing a satisfactory registration of their dietary habits and symptoms. 14
2 Complete list of available food groups obtained from Matvaretabellen. 17
3 Overview of datasets used in the experiments. 25
4 No. of frequent itemsets generated from the Apriori algorithm with different min. support thresholds. 26
5 Number of association rules discovered from dataset 1-3. 27
6 Number of association rules discovered from dataset 2. 27
7 Number of association rules discovered from dataset 4. 28
8 Number of association rules discovered from dataset 5-10. 30

1 Introduction

1.1 Topic covered by the project

The topic of this project is the use of machine learning techniques to discover relationships between irritable bowel syndrome (IBS) symptoms and diet. Machine learning is a subfield of computer science and artificial intelligence. Association rule mining (ARM) algorithms and feedback from clinical experts in IBS will be used to analyze collected patient data in order to identify food items that trigger IBS symptoms.

1.2 Keywords

IBS symptoms & diet associations; Association rule mining; Food logging

1.3 Problem description

A lot of people struggle with health problems related to diet. IBS is a generic term for issues related to the digestive system, and is estimated to affect 10-20% of the Norwegian population [1]. Symptoms of IBS varies and can cause problems like stomach pain, bloating, constipation and diarrhea. IBS is not dangerous, but the problems can be a major obstacle in the patients' daily life, which can lead to decreased quality of life [2].

The direct cause of the disease is uncertain, but a change of diet has in many cases helped overcoming or easing the symptoms [3]. However, dietary advice differs between websites, forums, social networks and even doctors. In addition, patients react individually on different diets. What eases the symptoms for one patient, might worsen the symptoms for another one. A lot of testing and failing is therefore often required before an ideal personal diet eventually is found, causing a lot of frustration for the patient. This emphasizes the need for a solution that can give accurate and individualized dietary advice.

In recent years there has been a growing focus on individualized treatment of various diseases. At the same time we can see an increasing use of self logging. Collected data from food logging is used as a possible solution by health professionals to get an overview of what their patients eat and drink over a certain period of time. This can be used to give feedback to the patients about what needs to be changed in order to overcome the disease. However, potentially huge amounts of data can come in, causing a lot of manual work for the health professionals.

At Haukeland University Hospital today, clinical nutritionists ask IBS patients to record their diet on paper for two or three days before a consultation. The recordings are manually analyzed by the clinical nutritionists in order to get an overview of the patient's diet. Based on the recordings, potential IBS trigger foods are identified and dietary advice is given. This method is ineffective, time demanding and vulnerable to errors as trigger foods can be overlooked. Another problem with food logging is the effort that is needed from the patient. Many find it demanding to record their food intake over a longer period of time [4]. This affects the accuracy of the data and the amount that is collected.

1.4 Motivation and benefits

It is predicted that there will be a lack of health professionals in the future [5]. At the same time, the need for individualized treatment is increasing.

If this study delivers the results expected, association rules can be used in a decision support system for clinical nutritionists to identify individual IBS trigger foods. This can hopefully lead to more accurate dietary advice for the IBS patients. In addition, the clinical nutritionists can use less time manually looking for relationships between diet and symptoms.

The primary target group in the present study are IBS patients. The intention is to make their every day life easier by offering individualized dietary suggestions to soothe IBS symptoms and increase their quality of life.

1.5 Research questions

Research goal: Discover association rules that can assist clinical nutritionists in understanding the relationships between food intake and IBS symptoms.

- **RQ1:** Can relationships between consumed food groups and IBS symptoms be identified using association rule mining?
- **RQ2:** Can relationships between consumed food items and IBS symptoms be identified using association rule mining?
- **RQ3:** Can relationships between consumed food items and IBS symptoms be identified for individual IBS patients, using association rule mining?
- **RQ4:** Can association rules be useful in a decision support system for clinical nutritionists in their work on consulting IBS patients?

1.6 Contributions

This thesis will hopefully give useful answers on which food items and food groups the IBS patients should avoid in order to overcome their symptoms. Association rules indicating relationships between foods and IBS symptoms will be presented.

1.7 Clarification of terms and acronyms

- IBS - Irritable Bowel Syndrome
- ARM - Association Rule Mining

1.8 Ethical and legal considerations

In order to collect and use data from IBS patients in the present study, an approval from "REK - Regional committees for medical and health research ethics" [6] was obtained. The approval is included in Appendix D. All patients gave their written consent to participate in the study.

2 Background

The intent of this chapter is to give an introduction to ARM and how it has been applied to different fields of medicine in the recent years. An overview of the current knowledge of the relationships between diet and IBS is also given, in addition to an inspection of state of the art solutions for revealing these relationships.

To the best of my knowledge, ARM has not previously been used to discover relationships between foods and IBS symptoms. The data to be collected in the present study can not be labelled. For that reason, unsupervised learning will be used.

Unsupervised learning is a subcategory of machine learning and is used to discover hidden patterns in datasets consisting of items that are not labelled. There are two main unsupervised learning approaches: Clustering and ARM. Clustering is the approach of separating a dataset into subsets where the items in each subset are similar to each other [7]. On the other hand, ARM is used to discover relationships between the items in a dataset.

2.1 Association rule mining

In ARM, the discovered relationships between items are presented as rules that are easy to understand and interpret by the reader. ARM has been widely and successfully used in different fields of medicine.

ARM was introduced by Agrawal et al. [8] in 1993. Using the history of purchases at a supermarket as an example, the authors' algorithm could reveal the probability of a certain product being a part of a purchase, based on the other products existing in the same purchase. This is called market basket analysis and has become the usual illustration of use case when different association rule algorithms are presented. Agrawal et al. defined ARM the following way:

Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of m binary attributes called items. Let T be a database of transactions. Each transaction t is represented as a binary vector, with $t[k] = 1$ if t bought the item I_k, and $t[k] = 0$ otherwise.

An association rule is an expression of the form X $\Rightarrow$ Y, where X and Y are itemsets in I. The rule states that if X occurs, then Y will occur. The left side of the rule (X) is called antecedent, while the right side of the rule (Y) is called consequent.

2.1.1 Objective interestingness measure of association rules

When large datasets are used for ARM, a great number of rules might be the result. In many cases, this can make it impossible for a human being to extract the interesting rules regarding the purpose of the rule discovery. In order to determine the interestingness of an association rule, several objective measurements exist. These are statistics based on frequency of the rules in the dataset. [9] state that a good interestingness measure of association rules should include both generality and reliability.

Support is a measure fulfilling the requirement regarding generality. It shows how frequent a given rule is occurring within the full dataset and says something about the statistical significance of a rule [8]. It is a good measurement, since rules having a very low support may be present because of a coincidence.
Support is defined as:

$$support(X \Rightarrow Y) = \frac{\text{Number of transactions containing both X and Y}}{\text{Total number of transactions}} \quad (2.1)$$

Confidence is another objective measurement, representing the reliability of the assumption made by a rule. It measures the probability of Y being in a transaction when it also contains X. The higher the confidence for a rule, the more likely it is that Y is present in transactions containing X. Confidence measures the strength of the rule [8].
Confidence is defined as:

$$confidence(X \Rightarrow Y) = \frac{\text{Number of transactions containing both X and Y}}{\text{Number of transactions containing X}} \quad (2.2)$$

Setting a minimum threshold of support and confidence (a so-called support-confidence framework) will limit the number of association rules. However, using these two measurements alone is in many cases not enough to end up with a manageable number of rules that can be considered as interesting. In order to evaluate the interestingness of the rules, an interest factor called lift can be added. Lift measures the correlation between X and Y. This is a measure indicating if the rule is occurring more often than what is expected by chance. Definition:

$$lift(X \Rightarrow Y) = \frac{confidence(X \Rightarrow Y)}{support(Y)} \quad (2.3)$$

2.1.2 Apriori algorithm

Apriori is the most well-known ARM algorithm. It was introduced by Agrawal and Srikant [10] in 1994 and is used to mine frequent itemsets and association rules from a transactional database.

Apriori is assuming that all subsets of a frequent itemset also must be frequent. Additionally, the subsets of an infrequent itemset must be infrequent. Based on a user-specified minimum support threshold, frequent itemsets in the database are identified by using a bottom-up approach. So-called candidate itemsets are tested against the data and drawn out one item at a time if they fulfill the support threshold.

Algorithm 1 Apriori algorithm

begin
  $L_1 \leftarrow Frequent1 - itemset$
  $k \leftarrow 2$
  while $L_{k-1} \neq \phi$ **do**
    $Temp \leftarrow candidateItemset(L_{k-1})$
    $C_k \leftarrow frequencyOfItemset(Temp)$
    $L_k \leftarrow compareItemsetWithMinimumSupport(C_k, minSupport)$
    $k \leftarrow k + 1$
  end while
  return L
end

From the frequent itemsets identified, association rules satisfying a user-specified confidence threshold are discovered.

2.2 Related work

2.2.1 IBS and the role of diet

Many research studies have looked at the impact foods have on IBS symptoms and how a change in diet can improve the quality of life for IBS patients [11]. There is a broad agreement among health professionals that diet plays an important role when it comes to understanding the cause of IBS symptoms. In a review of evidences that relate diet to IBS symptoms, Cuomo et al. [12] point at dietary strategies as one of the key tools in therapeutic management of patients with IBS.

We can see that the studies point at clear relationships between diet and IBS symptoms. Different food groups and nutrients are mentioned as common triggers [13, 14, 15, 16, 17]. However, we know that the effect of different foods varies from patient to patient. Zia et al.'s article [18] supports this statement. The authors investigated the food- and symptom journals from 17 patients registered over a period of 15 days. Using regression analysis, associations between food nutrients and symptoms were found in 13 of the journals. An overview of the results showed that the associations differed for every individual. Dapoigny et al. [19] discussed how difficult it is to give IBS patients general diet advice, as people react differently on various foods. They also stated that food alone is not responsible for all symptoms, but that individual dietary adjustments can lead to an increased quality of life for many of the patients. Also Cuomo et al. [12] pointed at the importance of customizing the diet in order to manage IBS symptoms. Their research suggested that reducing intake of short-chain carbohydrates and sugar alcohols (FODMAPs) have a positive impact on controlling symptoms. A diet restricting foods containing these types of carbohydrates are known as a low FODMAP diet. Staudacher et al.'s study [20] showed that 76% of patients following this diet for nine months experienced an improvement in symptoms control, compared to 54% of patients who followed standard dietary advice. Even if the study indicates that a low FODMAP diet is more effective than standard dietary advice, we can see that the approach is not able to help all of the patients. Following a strict low FODMAP diet may also cause the patient to miss out on necessary nutrients because of drastically restricting the selection of available food items. This underlines the need for individualized diet plans.

Several research studies show a significant effect in improving IBS symptoms on patients that are offered diets tailored to their needs. Ali et al. [3] looked in their research at the efficacy of individualized diet. Blood tests from a group of 58 IBS patients were analyzed to see how their white blood cells (leukocytes) responded when exposed to 200 different food extracts. The foods that caused a moderate or heavy reaction to the leukocytes were categorized as "bad". These foods were excluded from the individual diet plans for half of the group (intervention group), while the other half received a diet plan consisting only of "bad" foods. After following the assigned diet over a period of four weeks, the intervention group saw significantly improvements in quality of life, compared to the comparison group.

Even if an individualized diet is proven to ease the symptoms for a patient, knowing exactly *which* foods to avoid is the hard part when finding the right diet. A common approach to reveal the trigger foods, is to go through an elimination diet, where the patient eliminates one food item at the time to see how it affects the symptoms. This is a time consuming activity and may lead to malnutrition if not done correctly [21, 22]. Zia et al. [23] refer to the fact that there is no standardized method for identifying trigger foods. This may lead to large variations in the conclusions when health professionals interpret their patients' food journals. Their study included eight health professionals that interpreted 17 food journals from IBS patients. They were asked to rate how likely the different food groups were to cause IBS symptoms for the respective patients. The results showed that the agreement between the health professionals in general were poor, meaning that a patient probably will get different advice depending on which health professional is visited.

2.2.2 Identifying relationships between IBS symptoms and diet

Numerous studies are conducted with a goal of revealing the relationships between IBS symptoms and food intake [22, 19, 24]. Some of them make use of self-tracking via smartphone applications in order to discover individual food triggers.

Zia et al. [25] developed a smartphone application for diet- and symptom registration. The application was intended for IBS patients. The authors wanted to evaluate the usefulness of the application, in addition to the effect it had on symptoms. 11 participants were asked to log their gastrointestinal symptoms four times a day along with all meals over two weeks. Regression analysis was performed to find nutrients associated with reported symptoms. At least one strong association was found in 73% of the participants, but only 30% of them showed a statistically significant change on the IBS Severity Scoring System (IBS-SSS) [26]. Many of the participants expressed that they missed clear answers and instructions from the application on which foods to avoid. Regarding usability, inconsistency and complexity was mentioned as reasons for why the application was perceived as time-consuming. Experiences from the study can be useful when developing a food registration system for IBS patients in the future, both when it comes to usability and what IBS patients want in order to overcome their challenges.

Also Schroeder et al. [27] addressed the problems regarding manual identification of IBS triggers. They aimed to ease the understanding of food- and symptom journals for both patients and health professionals. Regression analysis were done on 10 journals to find correlations between food and symptoms. Interactive visualizations of the results were presented to the patients and health professionals. Semi-structured interviews revealed that the visualizations helped to better understand the data through a collaborative review.

2.2.3 Use of ARM in medicine

Even if there are different approaches on identifying individual IBS food triggers, use of ARM to solve the issue is clearly limited to a minimum. However, when it comes to other fields in medicine, ARM has been used frequently to identify relationships and improve knowledge about various conditions and situations. It is a convenient approach in the field of bioinformatics, as the data often are categorical or binary as a default.

In a recently published study from Breuner et al. [28], ARM was applied to find unidentified genotype-phenotype relationships in people with bipolar disorder. The Apriori algorithm was used on three independent datasets consisting of 2,835 phenotypically characterized patients. The authors identified strong relationships between genotype patterns and the phenotypes eating disorder and anxiety.

Heritage et al. [29] used ARM to reveal potentially harmful combinations of drugs for patients suffering from several diseases. Looking at 3,000,000 entries consisting of different medications taken and registered side effects from over 360,000 patients, both commonly known and surprising relationships between drugs and their side effects were revealed.

Diabetes mellitus has been widely targeted in the field of machine learning (ML) and data mining (DM). In a systematic review of the use of ML and DM in the field of diabetes, Kavakiotis et al. [30] went through research studies aiming to gain knowledge about diagnosis, complications, genetic background, environment, and management of the disease. They found that about 15% of the ML approaches used were of the unsupervised type, which is the category of ARM. ARM was mainly used to uncover different predictors occurring together in patients suffering from diabetes.

Finding combinations of risk patterns for type 2 diabetes was the goal in Ramezankhani et al.'s [31] study. The authors gathered variables like demographic characteristics, smoking habits, medical history and drug history from over 6,000 people during a time period of 12 years. The data were analyzed with ARM. This was found to be a useful approach in revealing which frequent combinations of factors leading to an increased risk of developing diabetes.

Also when it comes to diagnosing cancer, ARM has shown promising results. In [32], frequent patterns found in 92 mammogram images were used to distinguish cancerous and non-cancerous tumors. The Apriori algorithm followed by an ESAR (Extraction of strong associations) algorithm resulted in strong association rules that could be further used for diagnosis of digital mammogram images.

Prediction of epidemic outbreaks is another use case where ARM looks to be an effective tool [33]. Input variables like epidemiological data, meteorological data, socioeconomic data and mosquito data were selected based on reviews of available malaria literature. Buczak et al. managed to develop a model that successfully predicted malaria cases 7-8 weeks before they actually happened.

2.2.4 Food logging

Food logging is widely used by people to keep track of daily intake of nutrients. The motivations vary from losing weight, overcoming a disease or simply curiosity about how much and what is actually being consumed. The recordings are shared with friends via social media platforms or shared with health professionals as a part of a supervision. Common to all use cases is that the goal is improved health or well-being.

However, keeping up with the logging can be difficult to maintain. Several research studies mention problems with food logging over longer periods of time. The participants often find it exhausting and are struggling to keep up, mostly because they find the logging too complex, hard to understand, time consuming, or simply because they forget it [34].

Cordeiro et al. [4] studied people's experiences regarding traditional food logging, and it emerged that the main reasons meal registrations were skipped, was due to forgetting (54%), difficult registration process (40%) and not knowing the contents of the meal (40%). It was also highlighted that the respondents stopped the logging because of reduced value (37%), too much work required (25%) and time consumption (16%). Several of these findings illuminate the importance of a tool that is easy to use and demands as little as possible from the user, in order to maintain accurate and complete registrations over time.

The authors found that using photos as method for logging made the registration easier for the participants. It is also worth noting that 65% of the participants reported that they in previous experiences had skipped registrations because they did not know the ingredients in the meal. This issue was completely eliminated with use of photos.

Another way to increase logging rate, is the use of notifications as reminders. Bentley and Tollmar [35] demonstrated how powerful reminders can be to avoid users forgetting to register their food intake. For one month, they examined the behavior of 10 users testing a mobile application for food- and exercise logging. The results showed that only 12% of the users recorded every day. In a follow-up study, the application was updated with simple, non-interrupting status bar notifications that reminded the users to record today's intake and activity. The study revealed that 63% of the users now logged each day.

2.2.5 Summary and challenges

ARM has clearly turned out to be useful in several areas in medicine the recent years. A common goal for all the cases is revealing useful patterns in large datasets where it is difficult for human beings to discover relationships between variables manually. With this in mind, it is surprising that the use of association rules to identify relationships between foods and IBS symptoms is yet to be explored. However, looking at the current solutions linked to food logging and ARM there are challenges important to address.

It is difficult to keep up with food logging over longer periods of time, which can lead to inaccurate and insufficient amounts of data collected. This is a recurring issue in research studies where food logging is a central part. That is why it is so important to offer a solution that is easy to use and requires as little effort as possible.

Association rules should be presented in a clear and understandable way. This can form a common understanding for both health professionals and patients. The health professionals will be supported to make decisions and advice, while patients can trust the advice given. To know that the advice given is based on science has shown to be important for patients in order to gain trust [36].

Based on previous work with association rules in the field of medicine, the strengths of ARM, and the need for individualized dietary advice for IBS patients, ARM seems like a good approach to discover relationships between foods and IBS symptoms from dietary logs. To see if this is true, the research questions in section 1.5 will be challenged in the present study.

3 Methodology

This chapter describes the overall system architecture as illustrated in figure 1. Methods used to recruit patients, collect data, process collected data and interpret discovered association rules are presented. A repository with source code used to obtain the results in the study, can be found online [37].

3.1 Data collection

To collect relevant data to analyze in this study, historical dietary- and symptom data from people experiencing problems related to their stomach and digestive system were needed. IBS patients logged what they were eating and drinking in addition to the strength of encountered IBS symptoms. The logging period lasted for 12-16 days and were completed via a smartphone application made available for the patients attending.

3.1.1 Recruiting IBS patients

For identification of foods triggering IBS symptoms, diet logs from people diagnosed with IBS were needed. Logs from people sticking to their "normal" diet was preferable in order to be able to confirm the well-known trigger foods, in addition to reveal new ones.

"National competency service for functional gastrointestinal diseases" at Haukeland University Hospital [38] had an ongoing research study targeting IBS patients in an online course about the diagnose. In that connection, a change in the existing research application were submitted to include collecting dietary- and symptom data from patients via the smartphone app. Waiting for this approval took longer time than expected and delayed the process. However, the changes were finally approved (Appendix D) and the recruitment of patients could start.

A presentation about this study was held in an ongoing course [39] where 20 IBS patients attended. From this course, three patients registered an interest in participating in the study. Further, 70 patients waiting for attending the same course were called by phone and informed about the study. 23 of these showed an interest in participating.

E-mail addresses from a total of 26 patients were collected. Two weeks before the start of the study, an e-mail (Appendix A) was sent to each of the 26 patients. This contained background information about the study, a brief explanation of their role, and contact info in case they wanted further information. The respondents were asked to confirm if they were interested in participating. An informed consent form was then sent by traditional mail to each of the patients who confirmed their interest.

When the signed consent form were returned, a new e-mail (Appendix B) was sent to each patient. This contained a link to the mobile application in App Store and Google Play, detailed information about how the application worked and what to register. In the same e-mail, a unique

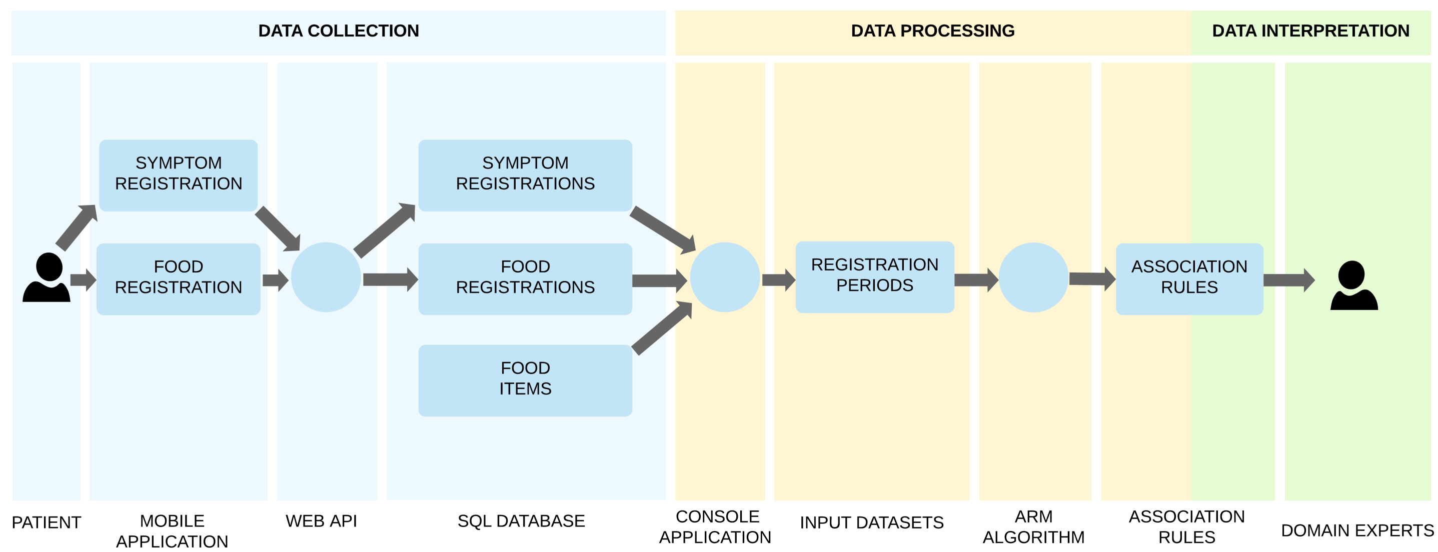

Figure 1: Architecture of the logging system, from a patient logging dietary intake and IBS symptoms, to domain experts interpreting the discovered association rules.

username and password were given for accessing the mobile application. With this information in hand, the patient was ready to start the registration.

14 out of the 26 potential participants signed and returned the consent form. Six of these never started the registration, while two started the registration but quit shortly after due to personal reasons. Figure 2 shows the different stages of the recruitment process.

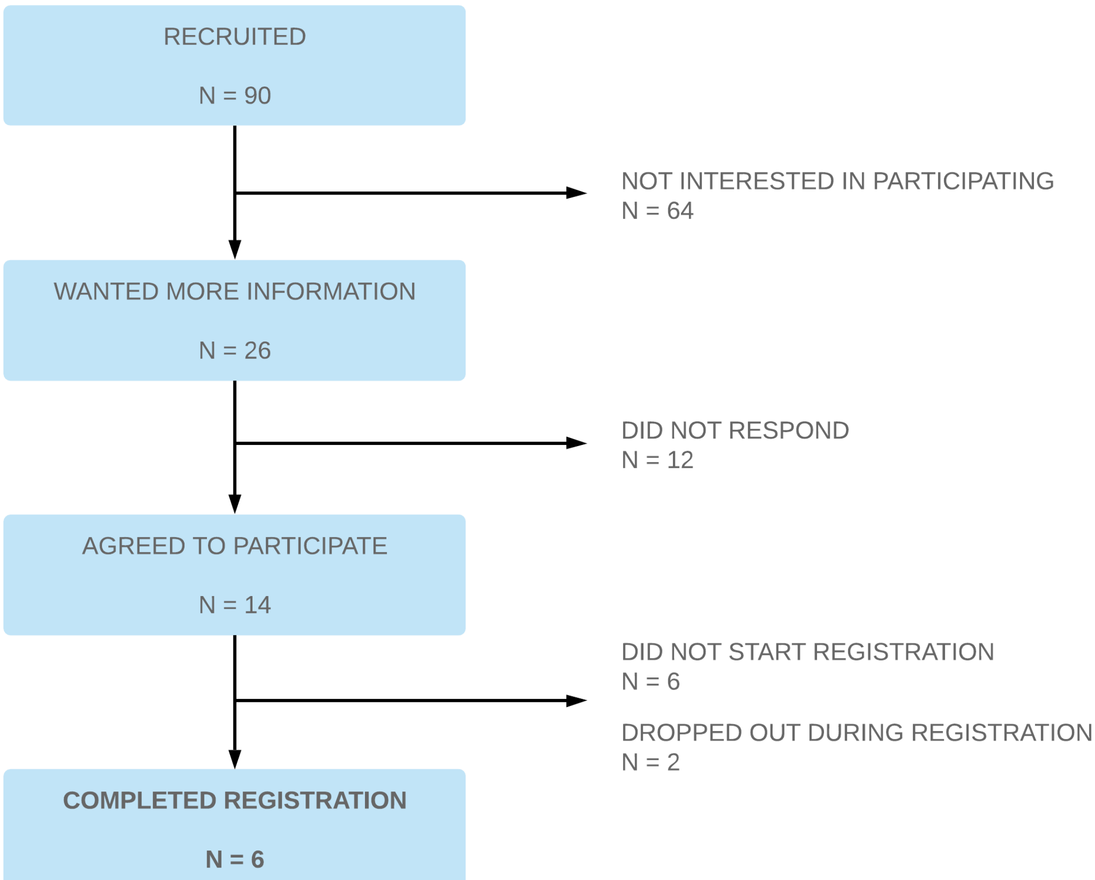

Figure 2: Flowchart of the patient recruitment process.

Six IBS patients completed a satisfactory registration of their dietary habits and IBS symptoms over periods ranging from 12 to 16 days. The group consisted of three females and three males aged from 28 to 53. An overview of the patients completing the registration is given in table 1.

A final e-mail (Appendix C) was sent to each of the six patients who completed the registration. I expressed my gratitude for their participation and asked three questions about how they experienced the logging period and the usefulness of the mobile application.

Table 1: Overview of IBS patients completing a satisfactory registration of their dietary habits and symptoms.

Patient	Gender	Age	Days of registration	No. of food registrations	No. of symptom registrations
1	F	32	14	207	21
2	F	51	15	280	33
3	M	53	16	187	17
4	F	51	14	195	25
5	M	40	15	172	23
6	M	28	12	119	2

3.1.2 Design and implementation of logging system

A mobile application and a web API were developed for the purpose of collecting dietary and symptom registrations from the patients. A mobile application is convenient for food logging, as most people have it available at all times. This is in contrast to a pen and paper approach.

Mobile application

The mobile application was developed with Ionic [40], a UI framework built on top of Angular [41] and Cordova [42] for deployment as a native mobile application. A native application was chosen over a web application due to easy access via established marketplaces that the patients already were familiar with.

To collect as much accurate data as possible, it was decided to keep the registration period to 14 days. Smaller amounts of accurate data were preferable to bigger but less accurate datasets. Regarding concerns on time consumption, the focus was to create a mobile application that was easy to understand and demanding as little effort as possible for each registration.

Three menu choices were visible in the mobile application, after logging in with a username and password:

- "**Register food consumption**" to search for food items, and register time of when they had been consumed
- "**Register symptom**" to register strength and time of when a symptom had occurred
- "**History**" to see previously registered food- and symptom registrations, with an option to delete each registration

Figure 3 is showing screenshots from the mobile application.

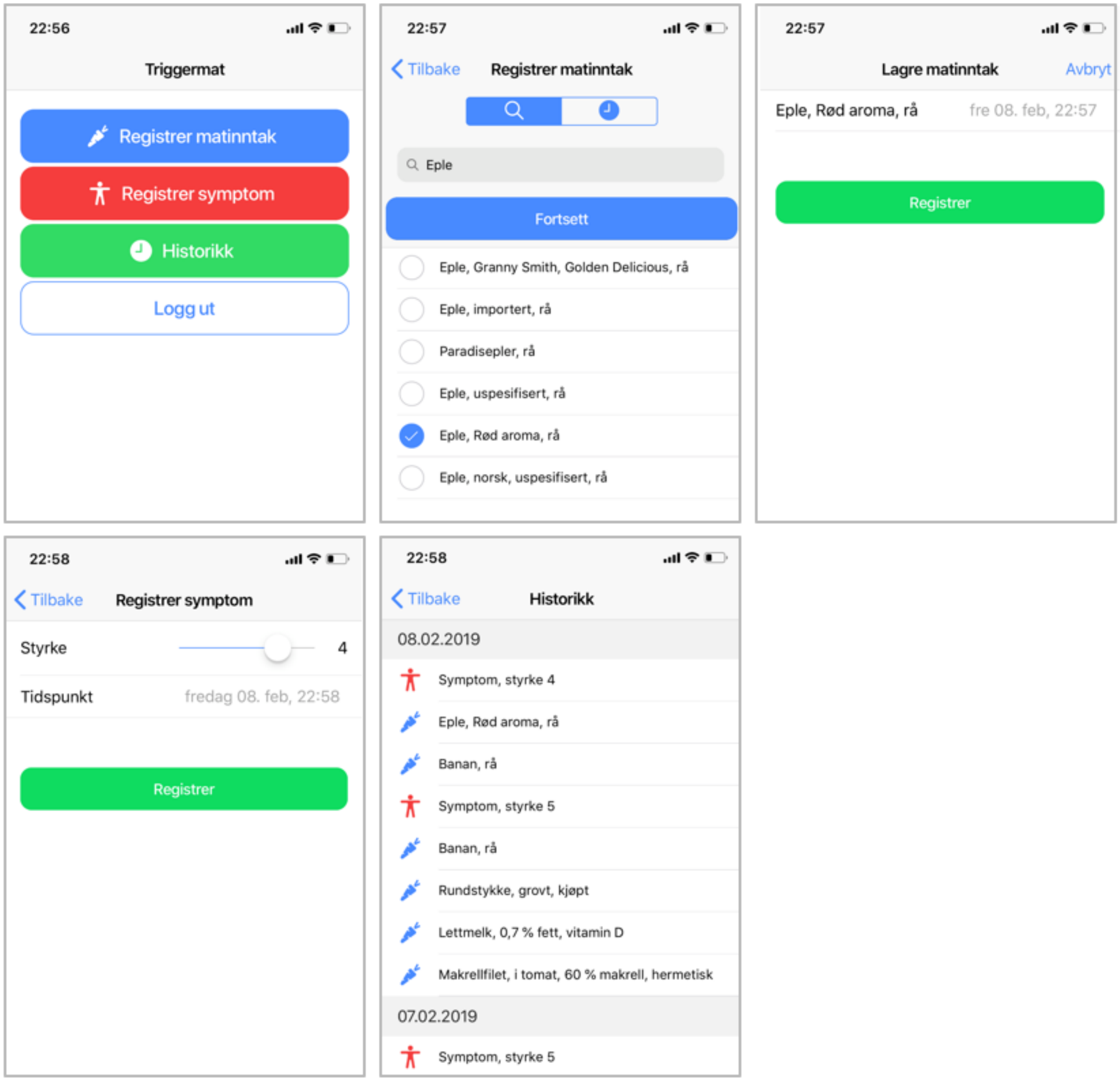

Figure 3: Screenshots of the mobile application. **Upper left:** menu. **Upper middle:** search and selection of food items. **Upper right:** registration of selected food item and time of consumption. **Lower left:** registration of symptom strength and time of occurrence. **Lower middle:** history of previous registrations.

Web API

An ASP.NET web API [43] was developed to offer the mobile application an endpoint to communicate with. The API authenticated requests from the mobile application, listed available food items, received food- and symptom registrations and listed a history of food- and symptom registrations for the authenticated patient.

To be able to look at individual registrations, the patients logged in to the mobile application with a given username and password. This ensured a connection between all of their registrations and a User object in the database. User objects were added to the database prior to the study. Figure 4 is showing the entity relationships in the database.

Before the patients were given access to the mobile application, all food items from the Excel format of Matvaretabellen [44] were extracted and stored in the database. This made all food items available via the web API.

When the patient searched for a food item in the mobile application, a request was sent to the API, who looked for food items containing the search term in either name or group. The found food items were returned so the patient could select the consumed food item, set time of when it was consumed and then post it back. The API then stored the food registration in the database.

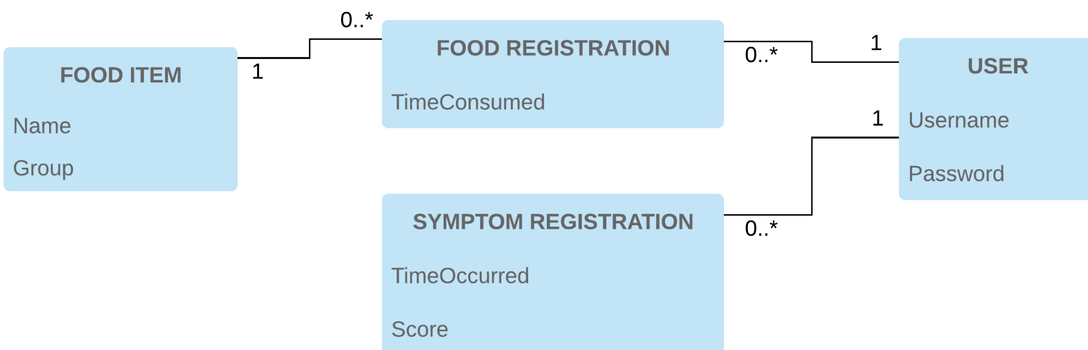

Figure 4: Simplified overview of entity relationships in the database.

3.1.3 Logging via mobile application

All patients were instructed to log their dietary intake and IBS symptoms for 14 days. The time of diet intake and symptom occurrence were also logged.

Diet

Diet implies every single food- and drink item digested. E.g. a breakfast might consist of low-fat milk, whole wheat bread, egg, Swiss cheese, butter and an apple.

Matvaretabellen was selected as a useful source of food- and drink items, as it contains many of the most common food items and meals consumed in Norway, the country where all of the patients were located. A total of 1,718 food items are listed, all with their respective nutrients, and categorized according to which food group it contains. See table 2 for the complete list of food groups. Having these values available was necessary to extract the actual food groups from the patients' diet habits, in order to reveal potential relationships to their symptoms.

Knowing that items from Matvaretabellen do not cover all dietary intake for a person, the pa-

Table 2: Complete list of available food groups obtained from Matvaretabellen.

Beef, veal, prepared	Beef, veal, raw
Beer, wine, spirit etc	Bread, rolls etc, industry made
Bread, rolls, etc, home-made	Breakfast cereals, muesli
Cheese, extra fat	Cheese, full fat
Cheese, reduced fat	Chocolate and other sweets
Cod liver oil	Cookies, sweet biscuits, rusks
Cream, sour cream and replacer	Crisp bread, crackers etc
Dessert, ice cream etc	Dishes with poultry or meat
Egg, prepared	Egg, raw
Fatty fish, prepared	Fatty fish, raw
Fish products, prepared	Fish products, sandwich fish
Flour	Fruit and berries, raw/fresh
Fruit and berry products	Grain, rice, pasta, prepared
Grain, rice, pasta, raw	Herbs and spices
Infant food	Juice, fruit drink, soda etc
Lamb, mutton, prepared	Lamb, mutton, raw
Lean fish, prepared	Lean fish, raw
Legumes	Margarine and butter
Mayonnaise, dressing etc	Meat products, sandwich meats
Milk and milk based beverages	Miscellaneous ingredients
Nuts, almonds and seeds	Oil, frying fat etc
Other cakes etc	Other meats, minced, offal, prepared
Other meats, minced, offal, raw	Pizza, pie, taco etc
Pork, prepared	Pork, raw
Porridge	Potatoes
Poultry, prepared	Poultry, raw
Powder base, dry	Sausages
Shellfish, fish offal	Snacks
Soup, sauce/gravy, stew base	Sugar, honey and sweet spreads
Vegetable products	Vegetables, prepared
Vegetables, raw and frozen	Vegetarian products and dishes
Water, coffee, tea	Yeast cake, griddle cake, waffle
Yoghurt	

tients were instructed that in cases where a food item was not found, the most related alternative was preferable, as looking at relationships between food groups and symptoms was one of the research questions. The option of letting the patients register own food items was assessed, but abandoned to keep the logging experience as easy as possible.

IBS symptoms

In addition to logging diet, the patients were instructed to register when IBS symptoms emerged. The symptom registration was chosen as a numerical value between one and five, where one was some discomfort and five was the feeling of great discomfort.

The type of symptoms in IBS patients varies, and comes in form of diarrhea, bloating, constipation and pain. Logging the different types was seen as an option, but omitted as this information is irrelevant regarding the research questions. The important point was if the patient felt discomfort, not the type of discomfort. Leaving out registration of symptom type would also be easier for the patient and decrease the chances of quitting the registration.

3.2 Data processing

The data collected in the present study was unlabeled. When looking for patterns and relationships in large, unlabeled datasets, unsupervised machine learning can be a powerful approach. That is why ARM is applied in this study. ARM algorithms discover rules presenting patterns and relationships in a human understandable way. This is a clear advantage since extracting specific results from large datasets is difficult or impossible to do manually.

As described in section 2.1, ARM is discovering rules based on the frequency of present items in a dataset's transactions. The data processing stage describes how collected data from the patients was transformed into different datasets of a required format. These datasets were used as input in separate executions of the Apriori algorithm. Based on frequent itemsets generated from the executions, association rules were discovered. The rules indicated which food items were causing IBS symptoms for the patients. Objectively interesting rules were extracted and prepared for interpretation by domain experts.

The data processing stage consisted of three steps and each step is described in this section:

1. Transforming collected data to input datasets
2. Discovering association rules
3. Extracting interesting association rules

3.2.1 Transforming collected data to input datasets

The wanted output from the ARM algorithm was association rules identifying relationships between consumed foods and occurred IBS symptoms. To get an impression of the common trigger foods in the patient group, registration data from all six patients together were included, as well as registration data from each patient individually. Foods were specified as food groups or food names. This formed a basis to transform the collected data into different input datasets that had to be executed separately.

Creating a registration timeline

The first step towards a dataset ready to be used as input, was to create a timeline for each patient. The timeline consisted of every food- and symptom registration the patient had made during their logging period.

Food- and symptom registrations from each patient were extracted from the SQL database using a .NET console application and Entity Framework Core [45]. The registrations were added to the timeline one by one, ordered by the date of when the food was consumed or when the symptom occurred. Each food registration was connected to a food item extracted from Matvaretabellen, so each food item's group and name were accessible.

Splitting timeline into registration periods

When an ARM algorithm identifies patterns, it looks for similarities between items present in the transactions of a dataset. To get a dataset of such form, the timeline with registrations had to be divided into different registration periods, each representing a transaction. Each registration period had the length of a predefined time interval and consisted of food- and symptom registrations present in the period.

The length of the registration periods were initially set to four hours, based on [18]'s findings regarding IBS patient reports on the time it takes from food is consumed until symptoms occur. However, this is highly individual from person to person. Symptoms may occur long before four hours and also long after. With this in mind, registration periods were generated based on the following time intervals:

- Two hours each
- Four hours each
- Six hours each

To calculate the total number of registration periods for each patient's timeline, the number of hours between the first- and the last registration was divided by the chosen time interval. All registrations (food items and symptoms) were then added to its corresponding period based on the time of when the food item was consumed or when the symptom had occurred.

Figure 5 illustrates how a timeline of registrations from an IBS patient was converted to registration periods, using different time intervals.

Creating CSV files of registration periods

The registration periods now consisted of food items and symptom registrations. This information had to be transformed into a format that was readable by the Apriori algorithm. To get the registration periods from the .NET console application to the Python [46] environment where the ARM algorithm was going to be executed, the registration periods for each use case listed below were stored as a CSV file. The CSV files were read and included in the Python script at a later stage of the process.

The first line in a CSV file is a comma-separated list of attributes. Due to several of the attributes

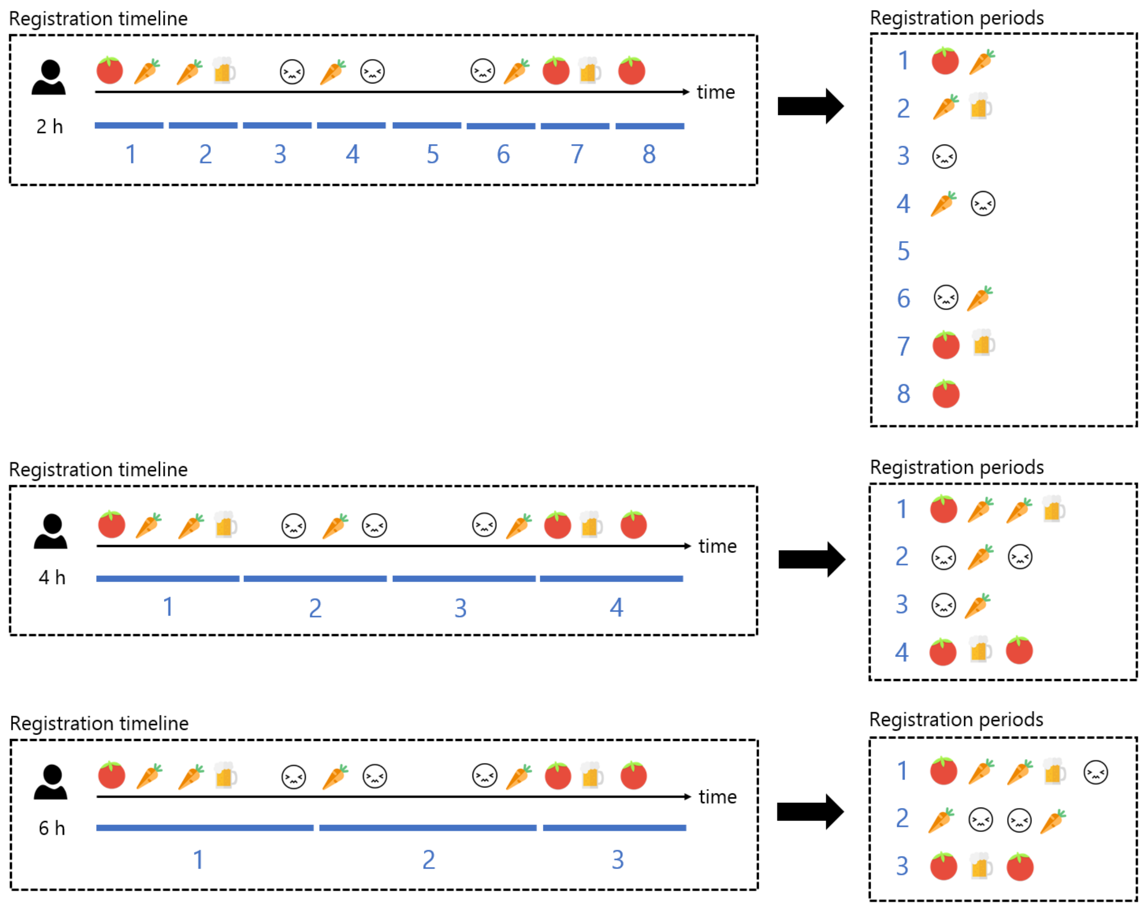

Figure 5: Illustration of how a timeline of registrations from an IBS patient was converted to registration periods, using different time intervals.

of interest containing commas, semicolon was used as a separator instead of comma. Both food groups and individual food items (specified by food name) were of interest in the present study, so CSV files had to be generated for both cases. To declare the attributes of interest, "Had symptoms" was added to the first line followed by either:

- All food groups existing among the food items in the database
- All food names existing among the food items in the database

The rest of the lines in a CSV file consist of values associated to the corresponding attribute on the first line. These lines each represented a registration period and indicated if a symptom had occurred in the period, as well as which food groups or food names that had been consumed in the same period.

To decide if a symptom occurred in the registration period or not, at least one symptom registration with a score equal to- or above one had to exist in the period. The target was to identify which foods were causing symptoms in general, not which foods were causing the strongest symptoms. Therefore, the threshold value was set to "1".

To decide if a food was consumed in the registration period or not, every food registration in the period were iterated through. A check was made against the relevant attribute (food group or food name) to see if it was present. If it was present, the value of the corresponding attribute was set to "1". If it was absent, the value was set to "0".

A total of 42 CSV files were generated using the .NET console application. The files contained the following:

- From all six patients together:
  - Absent/present symptoms and food groups in registration periods of two hours each
  - Absent/present symptoms and food groups in registration periods of four hours each
  - Absent/present symptoms and food groups registration in periods of six hours each
  - Absent/present symptoms and food names in registration periods of two hours each
  - Absent/present symptoms and food names in registration periods of four hours each
  - Absent/present symptoms and food names in registration in periods of six hours each
- From each of the six patients individually:
  - Absent/present symptoms and food groups in registration periods of two hours each
  - Absent/present symptoms and food groups in registration periods of four hours each
  - Absent/present symptoms and food groups in registration periods of six hours each
  - Absent/present symptoms and food names in registration periods of two hours each
  - Absent/present symptoms and food names in registration periods of four hours each
  - Absent/present symptoms and food names in registration in periods of six hours each

3.2.2 Discovering association rules

To discover association rules based on the different registration periods, three experiments were carried out executing a Python [46] script in Jupyter Notebook [47]. The script read selected CSV files from the list above and transformed each to a pandas [48] data frame. The data frame and a minimum support threshold were used as input to the Apriori algorithm from the Python library MLxtend [49].

Figure 6 and 7 display compressed samples of data frames used as input to the Apriori algorithm. Each row is representing a registration period of a predefined time interval, while each value in a row indicate if its corresponding attribute existed in the specific registration period (1) or not (0).

Based on a minimum support threshold, the Apriori algorithm generated frequent itemsets. An itemset in this context is a food registration, a symptom registration or a combination of food- and symptom registrations being present in a registration period. Generating frequent itemsets is the first step when discovering association rules.

Based on the frequent itemsets given as output from the Apriori algorithm, association rules were

id	Had symptoms	(Beef, veal, prepared)	(Beef, veal, raw)	(Beer, wine, spirit etc)	(Bread, rolls etc, industry made)	(Bread, rolls, etc, home-made)	(Breakfast cereals, muesli)	(Cheese, extra fat)	(Cheese, full fat)	(Cheese, reduced fat)	...	(Snacks)	(Soup, sauce/gravy, stew base)
1	0	0	0	0	1	0	0	0	0	0	...	0	0
2	1	0	0	0	0	0	0	0	0	0	...	0	1
3	0	0	0	0	0	1	0	0	1	0	...	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0

Figure 6: Compressed sample of a data frame representing registered food groups and symptoms

id	Had symptoms	(Adzuki beans, uncooked)	(Alaska pollock, raw)	(Alfalfa seeds, sprouted, raw)	(Almond beverage)	(Almond beverage, with calcium and vitamins)	(Almond flour)	(Almonds)	(Almonds, without peel)	(Amaranth, uncooked)	...	(Yeast ring, sweet, with dried fruit)
1	1	0	0	0	0	0	0	0	0	0	...	0
2	1	0	0	0	0	0	0	0	0	0	...	0
3	0	0	0	0	0	0	0	0	0	0	...	0
4	0	0	0	0	0	0	0	0	0	0	...	0
5	0	0	0	0	0	0	0	0	0	0	...	0

Figure 7: Compressed sample of a data frame representing registered food names and symptoms

discovered using "association rules" from MLxtend [49]. Input to the algorithm was the frequent itemsets from the Apriori algorithm and a minimum confidence threshold.

3.2.3 Extracting interesting association rules

In order to say something about the interestingness of the discovered association rules, there are a great number of objective measures available. These are calculated based on the frequency of the present itemsets in the dataset. Support, confidence and lift are already mentioned in section 2.1.1 and the minimum thresholds of support and confidence were manually set when the rules were discovered. As a starting point when looking for interesting rules, all of the rules were ordered by the lift measure.

Tweaking the support and confidence threshold gradually, eventually led to a great number of association rules discovered. However, even when satisfying these thresholds, many of the rules can give similar results with slight variations. As an example, a selection of discovered rules could look like this:

1. (Snacks, Sausages, Potatoes, Yoghurt $\Rightarrow$ Had symptoms)
2. (Snacks, Potatoes, Yoghurt $\Rightarrow$ Had symptoms)
3. (Sausages, Potatoes, Yoghurt $\Rightarrow$ Had symptoms)
4. (Sausages $\Rightarrow$ Had symptoms)
5. (Sausages, Potatoes, Snacks $\Rightarrow$ Had symptoms)
6. (Potatoes, snacks $\Rightarrow$ Had symptoms)
7. (Potatoes $\Rightarrow$ Had symptoms)

A large amount of rules is in many cases difficult to deal with for those who are going to interpret and make use of them. Many rules offering the same implications can make it hard to extract specific ones that are interesting and are offering real value.

When shortening the number of rules discovered, the focus was to end up with the rules that possibly could help get an answer to the research questions. Rules that are easy to read and interpret by a person with expertise in the field of nutrition and IBS. Straightforward and clean rules indicating foods that were likely to cause symptoms for the patients were favored. With this in mind, the discovered association rules were filtered to display rules with only one food (specified by group or name) as the antecedent (left side) and only the attribute "Had symptoms" as the consequent (right side) of the rule.

3.3 Interpretation of association rules with domain experts

To answer the research questions regarding the usefulness of ARM, feedback were needed from domain experts. The study was presented to six clinicians with expertise in the field of IBS, all working at "National competency service for functional gastrointestinal diseases" at Haukeland University Hospital. Three clinical nutritionists, one doctor, one nurse and one neuroscientist were present. With their approval, the sound of the presentation and the discussion session was recorded using a mobile phone. This was to review the discussions and the input that emerged during the session.

The research goal, patient recruitment and how the data was collected were explained and presented using Power Point slides on a presentation screen. A brief introduction to ARM was also given, describing the basics of how rules are discovered and what the different objective measures mean. This formed a foundation for the domain experts (clinicians) to interpret the discovered rules. Based on the interpretations, the domain experts could make up their minds about ARM as a future decision support system in consultations with IBS patients regarding their dietary habits.

After the introduction, a selection of discovered and filtered association rules were handed out to each of the clinicians and displayed on the presentation screen. The clinicians were allowed to speak freely and share their thoughts about the rules. To touch upon topics that were important in order to answer the research questions, the following questions were raised during the discussion:

- Are the results realistic in terms of current knowledge about IBS?

- Are the results surprising in terms of current knowledge about IBS?
- What are your opinions regarding the time intervals on the registration periods?
- Are association rules useful to identify IBS trigger foods?

3.4 Tools and frameworks

This section gives an overview of tools and frameworks used for data collection and data processing.

- **Visual Studio Code** [50] is a code editor. It was used when developing the Web API and mobile application, preprocessing the collected data and for continuous deployment of the Web API to Microsoft Azure [51].
- **Ionic** [40] is an open source toolkit for native mobile application development using the basic web technologies HTML, JavaScript and CSS. It is built on top of Angular [41] to interact with UI components and Corodva [42] for accessing mobile API features. Ionic was used to deploy a native mobile application both to Android and iOS using a single language and a single code base.
- **Xcode** [52] is an integrated development environment (IDE). It was used to sign and deploy the iOS mobile application to App Store [53].
- **Jupyter Notebook** [47] is a web application for coding and executing Python [46] scripts. It was used to discover association rules.
- **Pandas** [48] is a Python library. It was used to transform CSV files into data frames being used as input to the Apriori algorithm. It was also used to visualize the discovered association rules.
- **MLxtend** [49] is a Python library used to generate frequent itemsets and association rules from the input datasets.

4 Experiments and results

This chapter gives an overview of experiments that were carried out in order to discover association rules and provide answers to the research questions in section 1.5. 10 datasets based on the CSV files listed in 3.2.1 were used in different experiments as input to the Apriori algorithm. An overview of the datasets and their parameters are given in table 3.

Table 3: Overview of datasets used in the experiments.

Dataset ID	Food specification	Patient ID	No. of attributes	No. of registration periods	Time interval in each registration period
1	Food group	All	66	2,136	2 h
2	Food group	All	66	1,068	4 h
3	Food group	All	66	714	6 h
4	Food name	All	1,719	1,068	4 h
5	Food name	1	1,719	86	4 h
6	Food name	2	1,719	91	4 h
7	Food name	3	1,719	97	4 h
8	Food name	4	1,719	88	4 h
9	Food name	5	1,719	94	4 h
10	Food name	6	1,719	82	4 h

4.1 Filtering discovered association rules

When discovering association rules indicating relationships between foods and IBS symptoms, only rules showing foods as the antecedent (left side) and the attribute "Had symptoms" as the consequent (right side) were of interest. Therefore, all discovered rules from the experiments were filtered into two formats:

1. **(X $\Rightarrow$ y)**
   - X is the antecedent of the rule, being one or more food items specified by either food group or food name
   - y is the consequent of the rule, being the attribute "Had symptoms"
2. **(x $\Rightarrow$ y)**

- x is the antecedent of the rule, being one food item specified by either food group or food name
- y is the consequent of the rule, being the attribute "Had symptoms"

Due to readability, only discovered rules of the format ($x \Rightarrow y$) are listed in the experiments. Top 20 rules of the format ($X \Rightarrow y$) can be found in Appendix E.

4.2 Deciding on input parameters

To decide on which input parameters to use in the experiments, the Apriori algorithm were executed using different minimum support thresholds on three different datasets. These datasets contained registration periods with foods specified by their group, from all patients. The time interval of the registration periods is what separated them (two-, four- and six hours). The support threshold was first set to 0.10, and gradually decreased to 0.01. The number of frequent itemsets generated in the different executions can be seen in table 4.

Table 4: No. of frequent itemsets generated from the Apriori algorithm with different min. support thresholds.

Min. support	**No. of frequent itemsets, dataset 1 (2h)**	**No. of frequent itemsets, dataset 2 (4h)**	**No. of frequent itemsets, dataset 3 (6h)**
0.10	10	88	384
0.08	19	146	702
0.06	37	304	1,490
0.04	92	719	7,727
0.02	337	4,950	51,948
0.01	1,719	53,021	311,852

When deciding the threshold, a certain amount of frequent itemsets were desirable, without making the most infrequent ones available for rule discovery. After an evaluation of the numbers, 0.02 were selected as minimum support. This means that only combinations of foods and symptoms occurring in 2% or more of the registration periods were made available for discovery of association rules.

Rules with high confidence were desired in order to get rules indicating with a high certainty which foods that were present in the periods where symptoms occurred. Therefore, the confidence threshold was set to 0.80.

Support threshold of 0.02 and confidence threshold of 0.80 were used in all of the following experiments.

4.3 Deciding time interval in registration periods

As mentioned in 3.2.1, datasets with registration periods of different time intervals were generated. To assess how the results varied with different time intervals, the Apriori algorithm was executed three times using datasets with registration periods of two-, four- and six hours. These registration periods contained foods specified by their group. The number of association rules discovered from the three executions are displayed in table 5. This was used a basis for the upcoming experiments.

Table 5: Number of association rules discovered from dataset 1-3.

Dataset ID (see table 3)	**No. of association rules (X ⇒ y)**	**No. of association rules (x ⇒ y)**
1 (2h)	12	1
2 (4h)	1,305	10
3 (6h)	23,563	23

The number of rules increase along with increasing the time intervals, and so does the complexity of the rules containing multiple food groups as the antecedent. This makes it more difficult to extract interesting rules for the domain experts who are the intended interpreters of the rules. In addition, when increasing the time interval, the probability of a symptom registration being present also increases. This will lead to a higher probability of food groups being associated to a symptom, despite the fact that the food may have been consumed a long time before the symptom occurred.

Selecting a time interval in the middle seemed reasonable. By using four hour intervals, the time between meals was likely to be covered and the rules could be filtered into the format (x ⇒ y) without losing too much information on combinations of foods. With this in mind, datasets with registration periods of four hours were used as input in the following experiments.

4.4 Experiment 1

To identify relationships between food groups and IBS symptoms in respect to research question 1, association rules were discovered based on frequent itemsets from the Apriori algorithm. Dataset 2 was used as input. This dataset contained registration periods with foods specified by their group, from all patients. Table 6 shows the number of rules discovered.

Table 6: Number of association rules discovered from dataset 2.

Dataset ID (see table 3)	**No. of association rules (X ⇒ y)**	**No. of association rules (x ⇒ y)**
2 (4h)	1,305	10

Figure 8 shows discovered association rules of the format **(x ⇒ y)**. The rules are ordered by their lift value. They indicate different food groups being associated with symptoms. ("Poultry, raw" ⇒ "Had symptoms") is the strongest one according to the objective measures. The support measure of the rule is 0.028, meaning that both "Poultry, raw" and a symptom is registered in 2.8% of the 1,068 registration periods in dataset 2. Further, a confidence of 1.000 means that in 100% of the registration periods where "Poultry, raw" was registered, a symptom was also registered. This indicates high reliability for the rule. A lift value of 2.438 is suggesting that the occurrence of this rule is not just a coincidence.

The lift and confidence measures of the remaining rules are gradually decreasing. However, they still show fairly strong measures in terms of reliability and correlation as the confidence is above 0.8 and lift is well above 1.

The results of experiment 1 show that relationships between consumed food groups and IBS symptoms can be discovered using ARM. Accordingly, and also based on feedback from domain experts, the answer to research question 1 is "yes". The feedback is provided later in this chapter.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Poultry, raw))	(Had symptoms)	0.028090	0.410112	0.028090	1.000000	2.438356
2	((Cheese, extra fat))	(Had symptoms)	0.044944	0.410112	0.039326	0.875000	2.133562
3	((Other meats, minced, offal, raw))	(Had symptoms)	0.044944	0.410112	0.039326	0.875000	2.133562
4	((Dishes with poultry or meat))	(Had symptoms)	0.039326	0.410112	0.033708	0.857143	2.090020
5	((Nuts, almonds and seeds))	(Had symptoms)	0.073034	0.410112	0.061798	0.846154	2.063224
6	((Vegetable products))	(Had symptoms)	0.089888	0.410112	0.073034	0.812500	1.981164
7	((Grain, rice, pasta, raw))	(Had symptoms)	0.117978	0.410112	0.095506	0.809524	1.973907
8	((Beer, wine, spirit etc))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685
9	((Snacks))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685
10	((Soup, sauce/gravy, stew base))	(Had symptoms)	0.056180	0.410112	0.044944	0.800000	1.950685

Figure 8: Association rules indicating relationships between food groups and IBS symptoms for all patients.

4.5 Experiment 2

Experiment 2 was carried out to identify relationships between individual food items ad IBS symptoms. This was in order to provide an answer to research question 2. Dataset 4 was used as input to the Apriori algorithm. This dataset contained registration periods with foods specified by their name, from all patients. Table 7 shows the number of rules discovered.

Table 7: Number of association rules discovered from dataset 4.

Dataset ID (see table 3)	**No. of association rules (X $\Rightarrow$ y)**	**No. of association rules (x $\Rightarrow$ y)**
4 (4h)	431	19

Figure 9 shows discovered association rules of the format **(x $\Rightarrow$ y)**. The rules are ordered by their lift value. They indicate different food names being associated with symptoms. Naturally, several of the food names present in the rules in this experiment belong to the food groups listed in experiment 1. However, when the food names are listed instead of the food groups, it is easier to give accurate dietary advice to a patient.

The top six rules all indicate associations between food items specified by name, and symptoms. Each of the rules has a confidence value of 1.000 and a lift value of 2.438. This means that a symptom was registered in the same registration period every single time gluten-free bread, boiled egg, tomato ketchup, jasmin rice, boiled potatoes or warm smoked mackerel were registered.

Again, we can see that all 19 rules on the list have a lift value indicating that the associations between food item and symptom did not happen by chance.

The results of experiment 2 show that relationships between consumed food items (specified by name) and IBS symptoms can be discovered using ARM. Accordingly, and also based on feedback from domain experts, the answer to research question 2 is "yes". The feedback is provided later in this chapter.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Bread, gluten-free, white, with water, home-made))	(Had symptoms)	0.028090	0.410112	0.028090	1.000000	2.438356
2	((Egg, organic, boiled))	(Had symptoms)	0.028090	0.410112	0.028090	1.000000	2.438356
3	((Tomato ketchup))	(Had symptoms)	0.022472	0.410112	0.022472	1.000000	2.438356
4	((Rice, Jasmin, uncooked))	(Had symptoms)	0.022472	0.410112	0.022472	1.000000	2.438356
5	((Potatoes, storage, boiled with skin))	(Had symptoms)	0.022472	0.410112	0.022472	1.000000	2.438356
6	((Mackerel, warm smoked))	(Had symptoms)	0.022472	0.410112	0.022472	1.000000	2.438356
7	((Linseeds, flax seeds, crushed))	(Had symptoms)	0.056180	0.410112	0.050562	0.900000	2.194521
8	((Sesame seeds, without shell))	(Had symptoms)	0.056180	0.410112	0.050562	0.900000	2.194521
9	((Sunflower seeds))	(Had symptoms)	0.056180	0.410112	0.050562	0.900000	2.194521
10	((Margarine, vegetable fat, Soft Flora))	(Had symptoms)	0.084270	0.410112	0.073034	0.866667	2.113242
11	((Yoghurt, lactose free, vanilla))	(Had symptoms)	0.039326	0.410112	0.033708	0.857143	2.090020
12	((Cheese, hard, Norvegia))	(Had symptoms)	0.073034	0.410112	0.061798	0.846154	2.063224
13	((Butter))	(Had symptoms)	0.106742	0.410112	0.089888	0.842105	2.053353
14	((Cheese, ripened, Brie))	(Had symptoms)	0.033708	0.410112	0.028090	0.833333	2.031963
15	((Egg, boiled))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685
16	((Jam, without sugar, 60 % berries))	(Had symptoms)	0.056180	0.410112	0.044944	0.800000	1.950685
17	((Sweet pepper, red, cooked))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685
18	((Syrup))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685
19	((Casserole, with chicken, tomato, onion and mushroom))	(Had symptoms)	0.028090	0.410112	0.022472	0.800000	1.950685

Figure 9: Association rules indicating relationships between food names and IBS symptoms for all patients.

4.6 Experiment 3

The focus in experiment 3 was to find an answer to research question 3. The discovered association rules were supposed to identify trigger foods for the patients individually. The Apriori algorithm was executed using dataset 5 to 10 respectively as input. These datasets contained registration periods with foods specified by their name, each from one patient. Table 8 displays the number of rules discovered from the frequent itemsets in each of the executions.

Table 8: Number of association rules discovered from dataset 5-10.

Dataset ID (see table 3)	**No. of association rules (X ⇒ y)**	**No. of association rules (x ⇒ y)**
5 (4h)	3	1
6 (4h)	63	7
7 (4h)	0	0
8 (4h)	227	13
9 (4h)	0	0
10 (4h)	0	0

Association rules satisfying the thresholds were only found for three of the six patients. Figure 10 to 12 show the discovered rules from patient 1, 2 and 4. The rules are ordered by their lift value. They indicate different food names being associated with symptoms.

The rules are affected by the low amounts of collected data. When looking at datasets from patients individually, the number of registration periods is naturally much lower than what the case is when looking at data from all patients together. Between 82 and 97 transactions is not sufficient.

As an example, looking at figure 10, where "Sweet mix, without chocolate" is the only rule of the format (x ⇒ y) satisfying the confidence threshold. The confidence value of the rule is 1.000, which is strong. However, looking at the parameter called antecedent support, showing 0.023, meaning that this food item is registered in 2.3% of the 86 registration periods. In other words, two times. So, to achieve a confidence of 1.000, a symptom registration had to be present in both of these two registration periods. The lift value of 4.526 is also strong, and takes this into account, but the fact is that 86 registration periods, or transactions are insufficient. This is the case for all of the discovered rules where datasets for individual patients have been used as input. The rules can give useful indications about relationships between food and symptoms, but more data is needed.

The results of experiment 3 show that relationships between consumed food items (specified by name) and IBS symptoms can be discovered for individual IBS patients using ARM. Accordingly, and also based on feedback from domain experts, the answer to research question 3 is "yes". The feedback is provided later in this chapter. However, it must be mentioned that the size of the datasets used in experiment 3 are smaller than the datasets used in experiment 1 and 2. The rules in experiment 3 are therefore less conclusive.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Sweets mix, without chocolate))	(Had symptoms)	0.023256	0.22093	0.023256	1.0	4.526316

Figure 10: Association rule indicating a relationship between "Sweets mix, without chocolate" and IBS symptoms for patient 1.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Beef, minced meat, pan-fried, without fat))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
2	((Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.054945	0.351648	0.054945	1.0	2.84375
3	((Corn starch))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
4	((Cucumber, Norwegian, raw))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
5	((Egg, boiled))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
6	((Egg, fried in fat))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
7	((Tomato ketchup))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375

Figure 11: Association rules indicating relationships between food names and IBS symptoms for patient 2.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Avocado, raw))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
2	((Cheese, ripened, Brie))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
3	((Chicken, minced meat, raw))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
4	((Cream, sour, 35 % fat))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
5	((Egg, organic, boiled))	(Had symptoms)	0.056818	0.25	0.056818	1.0	4.0
6	((Greek salad, with feta cheese and olives))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
7	((Mackerel, warm smoked))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
8	((Soft drinks, cola drinks, artificially sweetened))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
9	((Sweet pepper, green, raw))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
10	((Water, tap))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
11	((Linseeds, flax seeds, crushed))	(Had symptoms)	0.113636	0.25	0.090909	0.8	3.2
12	((Sesame seeds, without shell))	(Had symptoms)	0.113636	0.25	0.090909	0.8	3.2
13	((Sunflower seeds))	(Had symptoms)	0.113636	0.25	0.090909	0.8	3.2

Figure 12: Association rules indicating relationships between food names and IBS symptoms for patient 4.

4.7 Feedback from domain experts

This section includes feedback from domain experts regarding design of the present study and their subjective interpretations of the discovered association rules.

4.7.1 Feedback on design of the study

Length of registration periods

One of the challenges in the data processing stage was to decide a time interval for the registration periods used as input to the Apriori algorithm. The results showed differences between two-, four- and six hour intervals.

When addressing the challenge with the clinicians, they expressed that the used intervals made sense. Four hours per registration period was according to the clinicians' experience likely to cover the time in between meals and thus catch symptoms and potential trigger foods in the same periods. However, they stated that individual variations from patient to patient regarding time from food intake to symptom occurrence make it difficult to select one interval. Some get symptoms right after the food is consumed, while others get symptoms the day after.

Combinations of trigger foods

The association rules presented for the clinicians all contained one food group or food name as the antecedent and "Had symptoms" as the consequent. However, from a clinician's perspective, it was pointed at the importance of being able to look at combinations of foods that might have caused symptoms. This in addition to single food groups or food names. Different types of foods can influence each other when consumed together and have an effect on the symptoms.

Registration of food quantity

Collecting the food quantity being consumed was requested by the clinicians, as this has an impact on the symptoms for an IBS patient. However, they agreed that it made sense to leave it out in this study to lower the burden for the patients when registering.

Classification of symptoms

To see if some foods were causing stronger symptoms than others, the clinicians wanted the possibility to filter the results by the symptom score threshold. From a clinical point of view, foods causing the strongest symptoms are the most interesting ones. Also the type of symptoms occurring were mentioned as an option being nice to have. The possibility to see associations between pain, bloating, diarrhea or nausea, and certain foods was suggested.

4.7.2 Feedback on value of association rules

When looking at the discovered association rules, different observations were made by the clinicians. Some of the rules stood out and confirmed their existing knowledge and expectations regarding foods triggering IBS symptoms.

"Vegetable products" was a food group that was presented among the rules. This food group was especially highlighted by the clinicians as typical trigger food and mentioned as a result they expected to see. Vegetables often contain so-called FODMAPs [12] and are in many cases recommended to be avoided for IBS patients.

Another food item showing up frequently among the rules, was egg. This matched with the clinical nutritionists' experiences. They explained that egg is frequently mentioned by their patients as a food item triggering IBS symptoms.

Several rules confirming the clinicians current knowledge about IBS trigger foods indicated that the discovered association rules were realistic. However, not all of the foods listed among the rules did make sense. Some of them were probably present due to coincidences rather than triggering symptoms. Food items that are consumed frequently in general, could be a possible false positive. When occurring in many of the registration periods, they can be considered as a trigger food without necessarily being one. For a person drinking water all day, water can show up as a potential trigger. Water triggering IBS symptoms was highlighted by the clinicians as unlikely.

When asked if some of the rules were surprising in terms of their current knowledge about IBS, a clinical nutritionist stated that rarely any presented trigger foods are surprising. Her experience was that patients in consultations regularly announce new foods that they allegedly react to. All association rules in the lists that were presented were not discussed one by one. However, the clinicians themselves pointed out rules that they perceived as interesting.

The general opinion among the clinicians was that association rules can be become highly beneficial in their work on consulting IBS patients and giving individual advise regarding their diet. The rules were seen as useful in order to assist them in identifying relationships between the diet and symptoms. They told that use of association rules can increase knowledge about the patient group and make it easier to give them dietary advice.

The presented system was also seen as useful in order to collect more data about the patients' dietary habits. This can be used to assist the clinical nutritionists in their work on counseling their patients. The clinicians mentioned that their patients could be given access to the mobile application prior to the consultation, so that the consultant could be ready with advice right away. This in contrast to the current solution, where patients log their diet with pen and paper over three days before clinical nutritionists manually review the logs to look for associations between foods and symptoms.

The results in the previous experiments and feedback from the domain experts, show that association rules can be used to identify relationships between consumed foods and IBS symptoms. Association rules can also be useful in a decision support system for clinical nutritionists in their work on consulting IBS patients. Based on these findings, research question 1, 2, 3 and 4 can all be answered as "yes".

5 Discussion

The goal of the present study was to discover association rules that can assist clinical nutritionists in understanding relationships between diet and IBS symptoms. Four research questions were asked initially:

1. Can relationships between consumed food groups and IBS symptoms be identified using association rule mining?
2. Can relationships between consumed food items and IBS symptoms be identified using association rule mining?
3. Can relationships between consumed food items and IBS symptoms be identified for individual IBS patients, using association rule mining?
4. Can association rules be useful in a decision support system for clinical nutritionists in their work on consulting IBS patients?

Through experiments and discussion with clinicians, the answer to all four research questions were "yes". Association rules can be used to identify relationships between foods and IBS symptoms. The rules can also be useful in a decision support system for clinical nutritionists in their work on consulting IBS patients.

The findings may prove to be beneficial when it comes to obtaining knowledge about what are triggering IBS symptoms. Use of ARM can help clinicians extracting useful relationships from large collections of dietary logs from patients, which is impossible to extract manually. ARM can support giving individualized dietary advice for patients. This is much needed for a patient group where the symptom triggers are highly individual and general advice often is not enough to overcome the symptoms.

However, there are various challenges related to collecting dietary logs from IBS patients and the process of discovering interesting association rules. These challenges and limitations are discussed in this chapter, together with ideas regarding future work on how to make use of ARM in a decision support system for clinicians consulting IBS patients.

5.1 Collecting data

Challenges, limitations and considerations regarding data collection from IBS patients, are presented in this section.

5.1.1 Amount of collected data

The four research questions were all confirmed, but the amount of collected data must be seen as the main limitation of the present study. The more data being available, the more accurate the discovered association rules become. However, the amount of data that can be collected from each patient was limited in the present study, especially over a short period of time. This makes it difficult to determine trigger foods for individual patients with a high certainty. On the other hand, the datasets from all patients in total are bigger and might give a better picture of the overall status when it comes to IBS trigger foods.

The amount of collected data can be increased by extending the logging period beyond 14 days. However, the challenges tied to logging are more than only the length of the logging period.

5.1.2 Recruitment

The number of patients completing the logging period was an obvious challenge in the study, which in turn limited the amount of collected data. This harmed the generalizability of the results. IBS patients are a group of wide variations when it comes to the degree of their illness. Most are managing their lives well despite their symptoms, while the quality of life for others are greatly impaired. This was also reflected during the recruitment process, where some of the participants dropped out due to lack of energy. Results from a selection of six patients are not enough to generalize from. Two of the intended participants had to withdraw from the study after a short period of time. They mentioned exhaustion from the disease as a reason for the withdrawal. They did not have enough energy to register their intake.

The reasons given for the dropouts, indicate that the patients completing the registration might not be the ones struggling the most. The fact that patients suffering most from the disease might not have been present, caused valuable data from this selection of the patient group to be missing. Collecting data from the most ill patients is important in order to gain more knowledge about this group. At the same time, this is a big challenge due to the fact that they already have enough to struggle with. Participating in a study is probably not their top priority.

Asking too much from a patient who is already struggling with their health can lead to lack of motivation and end up in no collected data at all. The ethical aspects should also be considered when including ill patients. Collecting data is important for the sake of research, but it should not be at the expense of the patients' health and well-being.

In order to collect data from more patients in future research, an alternative could be to implement a two week registration period for all IBS patients before attending a consultation with a clinical nutritionist. However, the challenges mentioned above have to be managed carefully.

5.1.3 Quality of collected data

Not only the amounts of collected data made an impact to the results in the present study. The quality of the data must also be taken into consideration when evaluating the results. There is no guarantee that the collected data actually represent the real dietary habits of the patients. Patients might leave out foods that they know are unhealthy or foods that they think are irrelevant for the research. Registrations might also slip under the radar simply because they are forgotten.

All incomplete registrations were not only due to the patients, though. A limitation in the study was the number of food items available via the mobile application (1,718). The entire assortment was limited to food items retrieved from Matvaretabellen. Foods consumed beyond this were therefore not registered, which gave an inaccurate representation of the dietary habits of the patients. Biased or lacking registrations can damage the accuracy of the results, and lead to actual trigger foods not showing up in the discovered rules.

To facilitate for patients logging their dietary habits, foods from more than one source should be made available for selection. In addition to making existing food items available for selection, an option might be to let the patients themselves add food items to the database. This can potentially boost the accuracy of data collected, as the chance of missing registrations will decrease.

Not only incomplete registrations can be a threat to the quality of the collected data. As a patient group suffering from a decreased quality of life, the desperation to overcome the disease often leads to testing out new diets regularly. Even if the patients during the logging period were encouraged to follow their normal diet without making modifications, a potential source of error is that the patients intentionally avoided foods that they suspected would trigger symptoms.

Another aspect that had to be taken into consideration when collecting data, was the length of the study. The logging period was set to two weeks. In the feedback from the patients, two weeks were seen as an appropriate length by some, while others thought that it was too long. The amount of collected data is highly decisive regarding the end result, so on one hand it is a need to stretch the logging period as much as possible to collect sufficient amounts of data. On the other hand, stretching it too much may cause the collected data to be of lower quality. Finding a solution in the middle should lead to the best results.

5.1.4 Usability of the logging system

When dealing with fragile patient groups, the last obstacle to prevent a completed data collection, should be the data collection system itself. Keeping the registration process as simple and effortless as possible is important to avoid irritations that may cause the patient to lose motivation and potentially quit. Feedback from some of the patients in the present study indicated that the mobile application for diet- and symptom logging in general was simple to use, but that a few minor bugs at times made the logging more complicated than necessary. These bugs could and should have been avoided by testing the mobile application more carefully before the data collection. This illustrates how small details are important when it comes to food logging. It is important to offer solutions and tools that require as little effort as possible from the patient.

To provide effortless solutions, artificial intelligence is likely to play an important role in food logging the coming years. Typing every food item and the exact quantity of what has been consumed is not manageable to most people over longer periods of time. Especially not for patients already struggling with their condition. Image recognition technologies are becoming increasingly powerful when it comes to identification of foods. Eventually, the exact quantity of every food consumed can be identified by using the smartphone to take a picture of a complete meal.

In order to be able to collect huge amounts of dietary data, the effort needed from the patients' point of view should be much less than what is the case today. Ideally, dietary habits should be reported automatically without any further involvement from the patients. However, this is a topic beyond this thesis. Still, this is important to look into when it comes to collecting accurate and enough data in research on dietary habits. Huge amounts of data can be extremely valuable.

5.1.5 Additional predictors

It is important to remember the fact that association rules discovered in the present study do not include other predictors than food. Discovered rules might look interesting due to objective measures or because they make sense based on current knowledge. However, the real reason why they show up can be due to factors that are not taken into account. Rules might have been rejected if the whole picture was known. Quantity of the food consumed, or the patient's state of mind at the time of consumption are examples of predictors that potentially could have been of interest. Symptoms that are registered by the patients can come from other sources than the consumed foods. This must be taken into consideration when the rules are interpreted, and should also be looked into in future research. The scope should be expanded to include more predictors.

The collected data and findings from this thesis can be used to take a deeper look at relationships between diet and IBS symptoms. Instead of using food groups and food names, the impact of specific nutrients can be investigated. Detailed nutritional contents for all food items are available in Matvaretabellen. However, for this to offer value, the quantity of each consumed food has to be registered.

A patient in the study raised a question about why the food quantity was not logged, as the patient knew by experience that this had an impact on her symptoms. This might have decreased her motivation for participating, as she had trouble seeing useful results being produced without logging the quantity of consumed food. However, as already mentioned, quantity registration was deliberately discarded in order to lower the registration burden for the patients. There is also a broad agreement among health professionals that quantity of food intake does have an impact on whether it will trigger symptoms or not. However, making the patients measure grams of all food items in every meal would increase the work required, and undoubtedly shorten the period they would be capable of keeping up the logging of quality data. With this in mind, a choice were made to leave out the quantity logging and focus on the food items.

Another aspect that can be investigated based on the current data collected, is which foods are causing the strongest symptoms. Can trigger foods be "rated" in terms of how bad they are? Are there specific foods standing out to cause symptoms of higher strength than others? For the sake of the present study, all symptom registrations with a score of 1 or above were included in the datasets. However, all registrations are collected with a score, so in future research these scores can be used to associate foods with symptoms of a certain strength.

When talking about symptoms, a potential future work could be to collect data identifying what type of symptoms that occurred (diarrhea, bloating, pain etc.). This might give answers on how different foods cause different types of symptoms, which in turn implies why the symptoms are occurring. Also other factors regarding diet can be collected in addition to the actual quantity consumed. Time between meals, how long each meal lasted, or time of the day when the food was consumed are all seen as possible factors that may have an impact on the symptoms.

Changing the focus away from diet, ARM can in future works be used to identify relationships between other lifestyle factors and IBS symptoms. Food is suspected not to be the only trigger, so to collect data representing activity, stress, sleep, anxiety, emotions, social happenings or other factors in the patients' lives could provide more answers on what is triggering the symptoms. Demographic parameters are also input that might help mapping different factors to symptoms.

There is no doubt that including different predictors potentially can give useful answers to what is triggering IBS symptoms. However, finding the right line between collecting *enough* data and collecting *quality* data is a great challenge. The difficulties of getting patients to keep logging over longer periods of time is setting limitations when it comes to prioritizing which type of data to collect. The fact that IBS is a complex disease with a lot of potential sources causing symptoms, makes it tempting to include several predictors in the data collecting stage. When choosing the level of data to collect, both the effort required from the patients and what is actually needed to discover useful results must be considered.

5.2 Discovery of association rules

In this section, challenges and limitations regarding the process of transforming collected data to useful association rules, are discussed.

5.2.1 Deciding length of registration periods

The choice of time interval when dividing registration timelines into registration periods is a challenge worth mentioning from the data processing stage. Depending on the selected time interval, relationships between foods and symptoms might have been left out of the results due to casually ending up in different registration periods. There is not one correct answer here, but finding a way to optimize the length is left as future work.

5.2.2 Deciding on input parameters

A drawback with the support-confidence framework, is that interesting rules of lower support than the set threshold might go unnoticed. When a great number of rules are discovered, it is difficult to extract those of interest. Different thresholds were being tested on the way to obtain a manageable amount of rules. The measures of the discovered rules should ideally be as strong as possible, without compromising rules with lower values that still could be of interest. The fact that there is not one minimum support- or minimum confidence threshold that is correct, makes it difficult to decide which thresholds to use. Small adjustments can potentially lead to very different results.

5.2.3 Presentation of discovered association rules

A common challenge when discovering association rules, is displaying the rules in a way that is understandable for the reader without pruning valuable information. Redundant rules showing more or less the same information should be removed. In the present study, this was done by filtering the rules to contain only one food as the antecedent. This method has its advantages and disadvantages. A benefit with filtering the rules so that they only contain one food group or food name as the antecedent, is that the discovered rules are reduced to a manageable amount being easy to read and interpret. As seen in experiment 1, the number of rules were reduced from 1,305 to 10 with this method. On the other hand, interesting rules containing combinations of foods could get lost when removing all rules containing more than one food as the antecedent.

Discovered association rules were presented "as is" for the clinicians. When applied to a decision support system, the rules should be presented in a more interactive way. It should be possible to filter rules manually and explore them on different detail levels. This was emphasized during the discussion with the clinicians, as they wanted to look at potential combinations of foods that trigger symptoms.

It is also worth mentioning that there are a number of alternative measures of association rules that can be used to rank their objective interestingness [54]. Lift was applied in this thesis, but use of different measures is left as future work.

5.2.4 Involvement of domain experts

The strength of ARM is its ability to extract useful relationships between items in a dataset where it is difficult for a human to extract the relationships manually. However, the discovered association rules alone can not give meaningful results. They are meant to support clinical nutritionists in consulting IBS patients about diet, and must be critically evaluated and put into a context where they make sense.

Objective measures like support, confidence and lift can say a lot about the strength of a rule, but to be really useful and provide information that can be converted into value in the real world, they must be evaluated and interpreted by a domain expert. This must be a person with in-depth knowledge of the field being investigated. In the present study, this was people with in-depth knowledge of IBS and nutrition. The domain expert should, in addition to know their domain, also be familiar with statistical terms, what the objective measures actually stand for and how the data was collected. With these criteria being met, this person is well equipped to tell which rules are interesting. Both expected ones and unexpected ones.

When rules are interpreted by domain experts, both expected and unexpected results may occur. It is easy to dismiss an unexpected result as a coincidence. However, when unexpected rules show up time after time in independent datasets, their implications can possibly point the clinicians in the right direction. Unexpected rules should be looked further into as they could reveal specific relationships between foods and IBS symptoms that are currently unknown.

Six clinicians in the field of IBS were involved in the evaluation of the discovered rules. Obtaining domain experts' opinion about the results is a strength of the study. However, an even greater strength would have been to involve the domain experts during the whole process. Both when setting input parameters and when making decisions regarding transformation of collected data. Multidisciplinary collaboration is necessary and can ensure a continuous improvement of the process towards the goal of discovering accurate and useful association rules.

6 Conclusion

The present study has applied ARM to collected dietary- and symptom registrations from six IBS patients. The goal was to discover association rules that can assist clinical nutritionists in understanding the relationships between food intake and IBS symptoms.

Association rules identifying relationships between food groups and symptoms, as well as specific food items and symptoms, have been discovered. The rules were interpreted and evaluated by domain experts. Their feedback indicate that ARM can be useful in a decision support system for clinical nutritionists in consulting IBS patients. The discovered rules gave a realistic reflection of the domain expert's current knowledge about IBS trigger foods.

The findings in this thesis can be used in future research to obtain more knowledge about IBS trigger foods and to give accurate, individualized dietary advice to IBS patients. Eventually, other IBS triggers beyond diet might be identified by applying ARM to patient data representing different lifestyle factors like stress, activity and sleep. IBS is affecting a lot of people and leads to a severely reduced quality of life. If it turns out that association rules can point at specific factors which can be targeted for intervention, they can be highly valuable for the patients affected.

Decision support systems making use of association rules should be easy to use for clinicians, offer clear and straightforward advice in addition to give an opportunity to dig deeper into the rules and see more details.

On the other hand, challenges need to be addressed in order to gain more knowledge and to provide accurate advice to individuals. ARM is best exploited when large amounts of data are processed. This is the main limitation of the present study, and makes it difficult to draw definite conclusions. More data is needed, especially when it comes to identifying trigger foods on patients individually.

Further studies should be conducted to collect more data, but it is important to simplify the logging process as much as possible for the patients, in order to avoid inaccurate data. Additionally, domain experts should always be involved during the data processing stage and when interpreting the association rules.

Bibliography

[1] Irritabel tarm | LMF | landsforeningen mot fordøyelsessykdommer. https://www.lmfnorge.no/Magesmerter/Diagnoser/IBS. Accessed: 14-05-2018.

[2] Simren, M., Axelsson, J., Gillberg, R., Abrahamsson, H., Svedlund, J., & Bjornsson, E. S. February 2002. Quality of life in inflammatory bowel disease in remission: the impact of IBS-like symptoms and associated psychological factors. *The American Journal of Gastroenterology*, 97(2), 389–396. URL: https://doi.org/10.1111/j.1572-0241.2002.05475.x, doi:10.1111/j.1572-0241.2002.05475.x.

[3] Ali, A., Weiss, T. R., McKee, D., Scherban, A., Khan, S., Fields, M. R., Apollo, D., & Mehal, W. Z. sep 2017. Efficacy of individualised diets in patients with irritable bowel syndrome: a randomised controlled trial. *BMJ Open Gastroenterology*, 4(1), e000164. URL: https://doi.org/10.1136/bmjgast-2017-000164, doi:10.1136/bmjgast-2017-000164.

[4] Cordeiro, F., Bales, E., Cherry, E., & Fogarty, J. 2015. Rethinking the mobile food journal. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI 15*. ACM Press. URL: https://doi.org/10.1145/2702123.2702154, doi:10.1145/2702123.2702154.

[5] Eggen, F. W., Røtnes, R., & Steen, J. 2018. Helse-norge 2040 - hvordan vil framtiden bli? Helsedirektoratet. URL: https://static1.squarespace.com/static/576280dd6b8f5b9b197512ef/t/5a6ed2bd0d92971475cc69aa/1517212361297/Helse-Norge+2040.pdf.

[6] REK - Regional committees for medical and health research ethics. Accessed: 01-05-2019. URL: https://helseforskning.etikkom.no/?_ikbLanguageCode=us.

[7] James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. *An Introduction to Statistical Learning*. Springer New York. URL: https://doi.org/10.1007/978-1-4614-7138-7, doi:10.1007/978-1-4614-7138-7.

[8] Agrawal, R., Imieliński, T., & Swami, A. jun 1993. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207–216. URL: https://doi.org/10.1145/170036.170072, doi:10.1145/170036.170072.

[9] Geng, L. & Hamilton, H. J. sep 2006. Interestingness measures for data mining. *ACM Computing Surveys*, 38(3), 9–es. URL: https://doi.org/10.1145/1132960.1132963, doi:10.1145/1132960.1132963.

[10] Agrawal, R. & Srikant, R. 1994. Fast algorithms for mining association rules. In *Proc. of 20th Intl. Conf. on VLDB*, 487–499.

[11] Kortlever, T. L., Huinink, S. T. B., Offereins, M., Hebblethwaite, C., Brien, L. O., Leeper, J., Mulder, C. J. J., Barrett, J. S., & Gearry, R. B. jan 2019. Low-FODMAP diet is associated with improved quality of life in IBS patients-a prospective observational study. *Nutrition in Clinical Practice*. URL: https://doi.org/10.1002/ncp.10233, doi:10.1002/ncp.10233.

[12] Cuomo, R., Andreozzi, P., Zito, F. P., Passananti, V., De Carlo, G., & Sarnelli, G. July 2014. Irritable bowel syndrome and food interaction. *World J. Gastroenterol.*, 20(27), 8837–8845.

[13] Volta, U., Pinto-Sanchez, M. I., Boschetti, E., Caio, G., Giorgio, R. D., & Verdu, E. F. oct 2016. Dietary triggers in irritable bowel syndrome: Is there a role for gluten? *Journal of Neurogastroenterology and Motility*, 22(4), 547–557. URL: https://doi.org/10.5056/jnm16069, doi:10.5056/jnm16069.

[14] Pereira, M. T., Malik, M., Nostro, J. A., Mahler, G. J., & Musselman, L. P. nov 2018. Effect of dietary additives on intestinal permeability in both drosophila and a human cell co-culture. *Disease Models & Mechanisms*, 11(12), dmm034520. URL: https://doi.org/10.1242/dmm.034520, doi:10.1242/dmm.034520.

[15] Puertolas, M. & Fifi, A. nov 2018. The role of disaccharidase deficiencies in functional abdominal pain disorders—a narrative review. *Nutrients*, 10(12), 1835. URL: https://doi.org/10.3390/nu10121835, doi:10.3390/nu10121835.

[16] Solar, I., Santos, L. A. O., Yamashita, L. M., Barret, J. S., Nagasako, C. K., Montes, C. G., Louzada, M. L., Calsa, M. S., & Vasques, A. C. J. oct 2018. Irritable bowel syndrome: associations between FODMAPS intake, problematic foods, adiposity, and gastrointestinal symptoms. *European Journal of Clinical Nutrition*. URL: https://doi.org/10.1038/s41430-018-0331-7, doi:10.1038/s41430-018-0331-7.

[17] Saha, L. 2014. Irritable bowel syndrome: Pathogenesis, diagnosis, treatment, and evidence-based medicine. *World Journal of Gastroenterology*, 20(22), 6759. URL: https://doi.org/10.3748/wjg.v20.i22.6759, doi:10.3748/wjg.v20.i22.6759.

[18] Zia, J. K., Chung, C.-F., Schroeder, J., Munson, S. A., Kientz, J. A., Fogarty, J., Bales, E., Schenk, J. M., & Heitkemper, M. M. sep 2016. The feasibility, usability, and clinical utility of traditional paper food and symptom journals for patients with irritable bowel syndrome. *Neurogastroenterology & Motility*, 29(2), e12935. URL: https://doi.org/10.1111/nmo.12935, doi:10.1111/nmo.12935.

[19] Dapoigny, M., Stockbrügger, R., Azpiroz, F., Collins, S., Coremans, G., Müller-Lissner, S., Oberndorff, A., Pace, F., Smout, A., Vatn, M., & Whorwell, P. 2003. Role of alimentation

in irritable bowel syndrome. *Digestion*, 67(4), 225–233. URL: https://doi.org/10.1159/000072061, doi:10.1159/000072061.

[20] Staudacher, H. M., Whelan, K., Irving, P. M., & Lomer, M. C. E. May 2011. Comparison of symptom response following advice for a diet low in fermentable carbohydrates (FODMAPs) versus standard dietary advice in patients with irritable bowel syndrome. *Journal of Human Nutrition and Dietetics*, 24(5), 487–495. URL: https://doi.org/10.1111/j.1365-277x.2011.01162.x, doi:10.1111/j.1365-277x.2011.01162.x.

[21] Ligaarden, S. C. & Farup, P. G. may 2011. Low intake of vitamin b6 is associated with irritable bowel syndrome symptoms. *Nutrition Research*, 31(5), 356–361. URL: https://doi.org/10.1016/j.nutres.2011.04.001, doi:10.1016/j.nutres.2011.04.001.

[22] Monsbakken, K. W., Vandvik, P. O., & Farup, P. G. dec 2005. Perceived food intolerance in subjects with irritable bowel syndrome – etiology, prevalence and consequences. *European Journal of Clinical Nutrition*, 60(5), 667–672. URL: https://doi.org/10.1038/sj.ejcn.1602367, doi:10.1038/sj.ejcn.1602367.

[23] Zia, J., Chung, C.-F., Xu, K., Dong, Y., Schenk, J., Cain, K., Munson, S., & Heitkemper, M. nov 2017. Inter-rater reliability of provider interpretations of irritable bowel syndrome food and symptom journals. *Journal of Clinical Medicine*, 6(11), 105. URL: https://doi.org/10.3390/jcm6110105, doi:10.3390/jcm6110105.

[24] Karkar, R., Zia, J., Schroeder, J., Epstein, D. A., Pina, L. R., Scofield, J., Fogarty, J., Kientz, J. A., Munson, S. A., & Vilardaga, R. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI 17*. ACM Press. URL: https://doi.org/10.1145/3025453.3025480, doi:10.1145/3025453.3025480.

[25] Zia, J., Schroeder, J., Munson, S., Fogarty, J., Nguyen, L., Barney, P., Heitkemper, M., & Ladabaum, U. March 2016. Feasibility and usability pilot study of a novel irritable bowel syndrome food and gastrointestinal symptom journal smartphone app. *Clinical and Translational Gastroenterology*, 7(3), e147. URL: https://doi.org/10.1038/ctg.2016.9, doi:10.1038/ctg.2016.9.

[26] Betz, C., Mannsdörfer, K., & Bischoff, S. October 2013. Validierung des IBS-SSS. *Zeitschrift für Gastroenterologie*, 51(10), 1171–1176. URL: https://doi.org/10.1055/s-0033-1335260, doi:10.1055/s-0033-1335260.

[27] Schroeder, J., Hoffswell, J., Chung, C.-F., Fogarty, J., Munson, S., & Zia, J. 2017. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW 17*. ACM Press. URL: https://doi.org/10.1145/2998181.2998276, doi:10.1145/2998181.2998276.

[28] Breuer, R., Mattheisen, M., Frank, J., Krumm, B., Treutlein, J., Kassem, L., Strohmaier, J., Herms, S., Mühleisen, T. W., Degenhardt, F., Cichon, S., Nöthen, M. M., Karypis, G., Kelsoe, J., Greenwood, T., Nievergelt, C., Shilling, P., Shekhtman, T., Edenberg, H., Craig, D., Szelinger, S., Nurnberger, J., Gershon, E., Alliey-Rodriguez, N., Zandi, P., Goes, F., Schork, N., Smith, E., Koller, D., Zhang, P., Badner, J., Berrettini, W., Bloss, C., Byerley, W., Coryell, W., Foroud, T., Guo, Y., Hipolito, M., Keating, B., Lawson, W., Liu, C., Mahon, P., McInnis, M., Murray, S., Nwulia, E., Potash, J., Rice, J., Scheftner, W., Zöllner, S., McMahon, F. J., Rietschel, M., & Schulze, T. G. nov 2018. Detecting significant genotype–phenotype association rules in bipolar disorder: market research meets complex genetics. *International Journal of Bipolar Disorders*, 6(1). URL: https://doi.org/10.1186/s40345-018-0132-x, doi:10.1186/s40345-018-0132-x.

[29] Heritage, J., McDonald, S., & McGarry, K. sep 2017. Integrating association rules mined from health-care data with ontological information for automated knowledge generation. In *Advances in Intelligent Systems and Computing*, 3–16. Springer International Publishing. URL: https://doi.org/10.1007/978-3-319-66939-7_1, doi:10.1007/978-3-319-66939-7_1.

[30] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. 2017. Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104–116. URL: https://doi.org/10.1016/j.csbj.2016.12.005, doi:10.1016/j.csbj.2016.12.005.

[31] Ramezankhani, A., Pournik, O., Shahrabi, J., Azizi, F., & Hadaegh, F. apr 2015. An application of association rule mining to extract risk pattern for type 2 diabetes using tehran lipid and glucose study database. *International Journal of Endocrinology and Metabolism*, 13(2). URL: https://doi.org/10.5812/ijem.25389, doi:10.5812/ijem.25389.

[32] Deshmukh, J. & Bhosle, U. 2016. Image mining using association rule for medical image dataset. *Procedia Computer Science*, 85, 117–124. URL: https://doi.org/10.1016/j.procs.2016.05.196, doi:10.1016/j.procs.2016.05.196.

[33] Buczak, A. L., Baugher, B., Guven, E., Ramac-Thomas, L. C., Elbert, Y., Babin, S. M., & Lewis, S. H. jun 2015. Fuzzy association rule mining and classification for the prediction of malaria in south korea. *BMC Medical Informatics and Decision Making*, 15(1). URL: https://doi.org/10.1186/s12911-015-0170-6, doi:10.1186/s12911-015-0170-6.

[34] Cordeiro, F., Epstein, D. A., Thomaz, E., Bales, E., Jagannathan, A. K., Abowd, G. D., & Fogarty, J. 2015. Barriers and negative nudges. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI 15*. ACM Press. URL: https://doi.org/10.1145/2702123.2702155, doi:10.1145/2702123.2702155.

[35] Bentley, F. & Tollmar, K. 2013. The power of mobile notifications to increase wellbeing logging behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems - CHI 13. ACM Press. URL: https://doi.org/10.1145/2470654.2466140, doi:10.1145/2470654.2466140.

[36] Schroeder, J., Karkar, R., Fogarty, J., Kientz, J. A., Munson, S. A., & Kay, M. September 2018. A patient-centered proposal for bayesian analysis of self-experiments for health. *Journal of Healthcare Informatics Research*, 3(1), 124–155. URL: https://doi.org/10.1007/s41666-018-0033-x, doi:10.1007/s41666-018-0033-x.

[37] Bitbucket - ThomasHatlebrekke / mastersthesis. Accessed: 26-05-2019. URL: https://bitbucket.org/ThomasHatlebrekke/mastersthesis/.

[38] Nasjonal kompetansetjeneste for funksjonelle mage-tarmsykdommer. Accessed: 22-05-2019. URL: https://helse-bergen.no/seksjon-avdeling/Sider/Nasjonal-kompetansetjeneste-for-funksjonell-mage-tarmsykdommer.aspx.

[39] IBS-Skolen, Helse Bergen. Accessed: 22-05-2019. URL: https://helse-bergen.no/avdelinger/medisinsk-avdeling/gastroenterologisk-seksjon/nasjonal-kompetansetjeneste-for-funksjonelle-mage-tarmsykdommer/ibs-skolen.

[40] Ionic Framework. Accessed: 20-05-2019. URL: https://ionicframework.com/.

[41] Angular Docs. Accessed: 20-05-2019. URL: https://angular.io/.

[42] Apache Cordova. Accessed: 20-05-2019. URL: https://cordova.apache.org/.

[43] ASP.NET Web API. Accessed: 09-05-2019. URL: https://dotnet.microsoft.com/apps/aspnet/apis.

[44] Matvaretabellen. Accessed: 22-05-2019. URL: http://www.matportalen.no/verktoy/matvaretabellen/#tabs-1-1-anchor.

[45] Entity Framework Core. Accessed: 08-05-2019. URL: https://docs.microsoft.com/en-us/ef/core/.

[46] Python.org. Accessed: 01-05-2019. URL: https://www.python.org/about/.

[47] Project Jupyter - Home. Accessed: 01-05-2019. URL: https://jupyter.org/.

[48] pandas: Python Data Analysis Library. Accessed: 01-05-2019. URL: https://pandas.pydata.org/.

[49] Home - mlxtend. Accessed: 01-05-2019. URL: http://rasbt.github.io/mlxtend/.

[50] Visual Studio Code. Accessed: 20-05-2019. URL: https://code.visualstudio.com/.

[51] Microsoft Azure. Accessed: 20-05-2019. URL: https://azure.microsoft.com/nb-no/.

[52] Xcode - Apple Developer. Accessed: 20-05-2019. URL: https://developer.apple.com/xcode/.

[53] App Store - Apple. Accessed: 20-05-2019. URL: https://www.apple.com/no/ios/app-store/.

[54] Tan, P.-N., Kumar, V., & Srivastava, J. jun 2004. Selecting the right objective measure for association analysis. *Information Systems*, 29(4), 293–313. URL: https://doi.org/10.1016/s0306-4379(03)00072-3, doi:10.1016/s0306-4379(03)00072-3.

A Information e-mail

The following e-mail were sent after a phone call where the patient confirmed an interest in the research:

Hei,

Som avtalt følger informasjon om forskningsprosjektet som du forhåpentligvis ønsker bidra i.

Bakgrunn
Prosjektet er en del av "Mage-tarmskolen på internett og mobilapplikasjon" som driftes av Nasjonal kompetansetjeneste for funksjonelle mage-tarmsykdommer ved Haukeland Universitetssjukehus.

Målet med prosjektet er å finne koblinger mellom matvarer og IBS-symptomer. Dette skal gjøres ved å analysere kostholdsdata som du registrerer i en mobilapplikasjon.

Hva skal du gjøre?

1. Du vil i løpet av de neste dagene få tilsendt et samtykkeskjema i posten. Dette må signeres av deg og returneres til oss så snart som mulig. Ferdig frankert konvolutt med utfylt adresse følger med skjemaet.
2. Når vi har mottatt signert samtykkeskjema fra deg, vil du få en e-post om hvordan du får tilgang til mobilapplikasjonen. I denne skal du gjennom to uker registrere:
    - hva du inntar av mat og drikke
    - styrke på IBS-symptomer som oppstår

Mobilapplikasjonen er enkel å bruke, og det kreves ingen måling eller veiing av mat som er inntatt. Det er hva som inntas som er viktig. Du skal spise og drikke som du pleier.

Når?
Du kan starte med registreringen fortløpende når du har fått tilgang til mobilapplikasjonen. Registrering bør starte så snart som mulig og senest innen utgangen av februar 2019.

Er du med?
Ønsker du å bidra i prosjektet, vennligst besvar denne e-posten med "Ja". Dersom du lurer på noe, ikke nøl med å stille spørsmål via e-post eller telefon 48 60 67 11.

NB: Samtykkeskjemaet du får tilsendt omhandler "Mage-tarmskolen på internett og mobilapplikasjon". Dette prosjektet fokuserer kun på mobilapplikasjon-delen, så avsnittene som beskriver behandlingsprogram via internett og utfylling av diverse spørreskjema er ikke relevant for deg i dette tilfellet.

Mvh
Thomas Akselberg Hatlebrekke
Masterstudent, NTNU
Webutvikler, Helse Vest IKT

B Instruction e-mail

The following e-mail was sent after a signed informed consent from the patient was received:

Hei,

Takk for at du ønsker å delta i forskningsprosjekt om koblinger mellom matvarer og IBS-symptomer. Under følger instruksjoner for bruk av mobilapp, samt praktisk informasjon om prosjektet.

Instruksjoner for bruk av mobilapp

Installasjon
Dersom du har en Android-telefon, last ned appen her:
https://play.google.com/store/apps/details?id=hatlebrekke.thomas.triggerfoods

Dersom du har en iPhone, last ned appen her:
https://itunes.apple.com/no/app/triggermat/id1450756986?l=nb&mt=8

Pålogging
Åpne appen og logg deg på med følgende brukernavn og passord:

Brukernavn: XXXX
Passord: XXXX

Registrering av mat og drikke

- Klikk på "Registrer matinntak"
- Søk etter matvaren du ønsker å registrere eller velg klokkeikonet for å se de siste registrerte matvarene
- Klikk på én eller flere matvarer i listen
- Klikk på "Fortsett"
- Klikk på dato, dersom du vil endre tidspunktet for når matvaren er inntatt
- Klikk på "Registrer"

Registrering av symptom

- Klikk på "Registrer symptom"
- Endre styrken på symptomet på en skala fra 1 til 5

- Klikk på dato, dersom du vil endre tidspunktet for når symptomet startet
- Klikk på "Registrer"

Se på historikk

- Klikk på "Historikk"
- Dersom du vil slette en matvare eller et symptom, dra elementet til venstre og klikk på ikon med søppelspann til høyre. Klikk på "Slett".

Praktisk informasjon

Hva skal du registrere?

- Matvarer og drikke som inntas.
    - Tidspunktet er når du startet å innta matvaren.
- Symptomer som oppleves. Et symptom kan være oppblåsthet, smerter i magen, diaré, forstoppelse eller andre ubehag knyttet til mage og fordøyelsessystem.
    - Styrke skal oppgis relativt utifra hvor stort ubehag du selv føler, der 1 er litt ubehag og 5 er stort ubehag.
    - Tidspunktet er når symptomet startet.

Hva hvis du ikke finner matvaren i appen?
Dersom matvaren du har inntatt ikke finnes i appen, kan du registrere en tilsvarende matvare i samme kategori.

Eksempel: Du søker på "Pepsi Max", men får ikke noe resultat. Søker du derimot på "brus", vil resultatet "Brus, cola, kunstig søtet, lett" være et naturlig valg.

Dette lar seg gjøre siden vi vil se på sammenhengen mellom symptomer og ulike matkategorier, i tillegg til de spesifikke matvarene.

Når skal du registrere?
Det anbefales på det sterkeste å registrere umiddelbart når du starter et måltid. Det er fort gjort å glemme hva man har spist, og det kan bli mye å registrere i ettertid dersom du venter for lenge.

Hvor lenge skal du registrere?
Oppstart er så snart som mulig. Registreringen skal pågå i 14 dager fra den dagen du starter.

Spørsmål?
Dersom du er usikker på hvordan appen fungerer eller har andre spørsmål rundt prosjektet, ikke nøl med å kontakte meg på e-post eller telefon 48 60 67 11.

Instruksjonene ligger også tilgjengelig på https://sites.google.com/view/triggermatapp/home.

Lykke til!

Mvh
Thomas Akselberg Hatlebrekke
Masterstudent, NTNU
Webutvikler, Helse Vest IKT

C Final e-mail

The following e-mail was sent after the patients had completed their logging period:

Hei,

Takk for at du har deltatt i forskningsprosjektet! Det settes stor pris på. Dataene blir analysert og vil forhåpentligvis kunne gi nyttige resultater.

Avslutningsvis håper jeg at du vil svare på et par spørsmål angående registreringen:

1. Opplevde du mobilapplikasjonen som brukervennlig og enkel å bruke? Hva kunne eventuelt vært forbedret?
2. Følte du at det var krevende å registrere alt av mat, drikke og symptomer? Synes du at 14 dager var for lenge, eller kunne du fortsatt med registreringen i en lengre periode?
3. Andre kommentarer rundt din opplevelse av registreringen eller deltakelsen i forskningsprosjektet?

Mvh

Thomas Akselberg Hatlebrekke
Masterstudent, NTNU
Webutvikler, Helse Vest IKT

D REK approval

This appendix includes the approval from "REK - Regional committees for medical and health research ethics".

REK
REGIONALE KOMITEER FOR MEDISINSK OG HELSEFAGLIG FORSKNINGSETIKK

Region:	Saksbehandler:	Telefon:	Vår dato:	Vår referanse:
REK vest	Jessica Svärd	55978497	13.02.2019	2016/1098/REK vest
			Deres dato:	Deres referanse:
			16.01.2019	

Vår referanse må oppgis ved alle henvendelser

Birgitte Berentsen
Medisinsk avdeling

2016/1098 Mage-tarmskolen på internett og mobilapplikasjon

Forskningsansvarlig: Helse Bergen HF, Helse Bergen HF - Haukeland universitetssykehus
Prosjektleder: Birgitte Berentsen

Vi viser til søknad om forhåndsgodkjenning av ovennevnte forskningsprosjekt. Søknaden ble behandlet av Regional komité for medisinsk og helsefaglig forskningsetikk (REK vest) i møtet 06.03.2019. Vurderingen er gjort med hjemmel i helseforskningsloven (hforsknl) § 10.

Prosjektomtale

Vurdering

I vedtak datert 12.11.2018 for prosjektendring innsendt 26.10.2018 ba REK vest om tilbakemelding med begrunnelse for utvidelse av antall deltakere til 2000 og redegjørelse for rekrutteringsprosedyre. Videre ba REK vest om et oppdatert informasjonsskriv der REKs mal for informasjonsskriv er brukt.

Tilbakemelding fra prosjektleder:
Rekrutteringen av pasienter skjer i klinikken etter endt diagnostisk utredning 1) gjennom samtale i poliklikk med lege som et oppfølgingstilbud etter at pasienten har mottatt diagnosen, 2) pasienter på venteliste til standard IBS-skole (gruppekurs) ved læring og mestringssenter blir oppringt per telefon av forskningssykepleier (90% takker ja til et elektronisk tilbud som kan gjøres i egen tid på kvelden fremfor å ta fri fra jobb eller at de har for mye plager til å delta fysisk ved læring og mestringsenter), 3) Henvisning direkte fra fastlege som kjenner til Mage-tarmskolen på internett.

Begrunnelse for utvdelse av antall deltagere til studien (2000) personer er basert på våre preliminære resultater: Pasienter får tilgang til Mage-tarmskolen og jobber seg gjennom 5 moduler over en periode på 8-12 uker. Intervensjonskomponentene involverer eksponeringsterapi (prinsipper fra kognitiv adferdsterapi) og lavFODMAP-diettveiledning ved klinisk ernæringsfysiolog. Pasientrapporterte data i form av spørreskjema blir fylt ut ved oppstart, 3 og 6 måneder.

Nyere litteratur viser til hvordan IBS er en paraplydiagose for multiple sub-kategorier av diagnosen. Dessverre så er frafallet ved denne studien stort da Mage-tarmskolen oppleves som et lavterskeltilbud ved IBS. Så langt har kun 68 pasienter som besvart alle spørreskjema både ved baseline, 3 og 6 måneder. Fullstendige data gjør det mulig å utføre korrelasjonsstudier og identifisere hvilke faktorer ved Mage-tarmskolen som hadde effekt på hvilken undergruppe, av pasienter.

Besøksadresse:
Armauer Hansens Hus (AHH), Tverrfløy Nord, 2 etasje. Rom 281. Haukelandsveien 28

Telefon: 55975000
E-post: post@helseforskning.etikkom.no
Web: http://helseforskning.etikkom.no/

All post og e-post som inngår i saksbehandlingen, bes adressert til REK vest og ikke til enkelte personer

Kindly address all mail and e-mails to the Regional Ethics Committee, REK vest, not to individual staff

Gjennomsnittsalder er 38 år, 19 % menn og 81 % kvinner. Vi ønkser å øke antall studiedeltagere slik at vi får en høyere antall (n) menn slik at det ved statistiske analyser er mulig å justere for alder, kjønn og/eller se forskjell mellom kjønn og alder.

I gjennomsnitt ble IBS-symptomskår (IBS-SSS) ble signifikant redusert etter 3 måneder og forbedringen var vedlikeholdt etter 6 måneder (IBS-SSS, oppstart = 269±81, 3 måneder= 234±79, p = 0,024; 6 måneder = 213±37 p = 0,000). Gjennomsnittsreduksjonsskår var 54 poeng (ansees som klinisk signifikant symptomlette hvis >50 poeng). Responder vs non-responder ble delt ved 50 poeng symptomlette. Responder: 58 % av deltagerne fikk 110 poeng reduksjon ved IBS-SSS. Noe som er meget bra og blandt de beste resultatene vi har hatt innen denne pasientgruppen. Vi anser derfor dette arbeidet som svært viktig for å kunne gi pasientene et kvalitetsikret, mest mulig optimalt helsetilbud. Responder-gruppen rapporterte også endring i livskvalitet i form av bedret kroppsbilde, matunngåelse, helsebekymringer, dysfori, aktivitetsnivå, relasjoner, sosial adferd, og seksuell aktivitet (p>0.000).

Det er stor varianse innen pasientgruppen som er svært heterogen. Våre preliminære resultater viser at ca 20% får en placeboforebedring etter 3 måneder, men ingen vedvarende endring etter 6 måneder, samt ca 20% rapporterer en forverring av tilstanden etter Mage-tarmskolen på internett. For å investigere dette bedre og gjøre korrelasjonsanalyse ved multiple underkategoriser og linerær regressjon basert på symptomskår forventet effect size (f2) medium (0.20) og en statstisk power på 0.8 (80%) med et sannsynlighetsnivå på er kravet n=60 i hver gruppe. Hvis 30% av deltagerne ikke besvarer 6 månedersspørreskjema (=1400 deltagere), og 20% av disse er non-responders som opplever en forrverring av symptomer (=280 deltagere). Våre spørreskjema tilsvarer over 200 variabler per deltager, noe som gir behov for signifikant økt antall n for å kunne feste lit til statistiske funn blandt denne heterogene gruppen. Sub-katergosiering av studiedeltagerne vil tillate muligheten for bedre persontilpasset opplæring og behandlingstilbud. Det er svært viktig å indetifisere hvilke phenotyper/kvaliteter som kategoriserer en non-responder hvis tilstand forverres i løpet av Mage-tarmskolen. Å identifisere og kunne predikere en "klassisk" responder-pasient i forkant av ihversettelse av helsetiltak kan bidra til innsparing og ekkeftivisering av helseresurser i poliklinikken ellers.

I 2019 ønsker vi å tilby pasienter fra hele Helse Vest som er henvist til HUS, å være med i denne studien. Det er viktig for å identifisere mulige regionale forskjeller og legge grunnlaget for mulig fremtidig elektronisk behandling av pasienter i hele Norge.

Revidert informasjonsskriv er lagt ved.

Vurdering

REK vest vurderer tilbakemeldingen som tilfredsstillende og godkjenner prosjektendringen.

Vedtak
REK vest godkjenner prosjektendringen.

Klageadgang
Du kan klage på komiteens vedtak, jf. forvaltningsloven § 28 flg. Klagen sendes til REK vest. Klagefristen er tre uker fra du mottar dette brevet. Dersom vedtaket opprettholdes av REK vest, sendes klagen videre til Den nasjonale forskningsetiske komité for medisin og helsefag for endelig vurdering.

Med vennlig hilsen

Marit Grønning
Prof. Dr. med.
Komitéleder

Jessica Svärd
rådgiver

Kopi til:postmottak@helse-bergen.no

E Association rules of format (X ⇒ y)

Figure 13 to 17 list the top 20 generated rules from experiment 2, 3 and 4 of the format (X ⇒ y). This is how the rules look with no limitations set regarding number of foods as the antecedent, before being filtered to the format (x ⇒ y).

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Milk and milk based beverages), (Margarine and butter), (Porridge), (Egg, prepared), (Cheese, full fat), (Bread, rolls, etc, home-made), (Fruit and berry products), (Water, coffee, tea))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
2	((Bread, rolls, etc, home-made), (Legumes), (Juice, fruit drink, soda etc), (Margarine and butter))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
3	((Milk and milk based beverages), (Bread, rolls, etc, home-made), (Egg, prepared), (Margarine and butter))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
4	((Egg, prepared), (Bread, rolls, etc, home-made), (Margarine and butter), (Porridge))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
5	((Egg, prepared), (Bread, rolls, etc, home-made), (Margarine and butter), (Vegetables, raw and frozen))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
6	((Egg, prepared), (Bread, rolls, etc, home-made), (Margarine and butter), (Water, coffee, tea))	(Had symptoms)	0.050562	0.410112	0.050562	1.0	2.438356
7	((Milk and milk based beverages), (Bread, rolls, etc, home-made), (Egg, prepared), (Mayonnaise, dressing etc))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
8	((Egg, prepared), (Bread, rolls, etc, home-made), (Mayonnaise, dressing etc), (Water, coffee, tea))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
9	((Egg, prepared), (Meat products, sandwich meats), (Bread, rolls, etc, home-made), (Water, coffee, tea))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
10	((Milk and milk based beverages), (Bread, rolls, etc, home-made), (Egg, prepared), (Porridge))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
11	((Milk and milk based beverages), (Bread, rolls, etc, home-made), (Egg, prepared), (Water, coffee, tea))	(Had symptoms)	0.033708	0.410112	0.033708	1.0	2.438356
12	((Egg, prepared), (Bread, rolls, etc, home-made), (Water, coffee, tea), (Porridge))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
13	((Margarine and butter), (Porridge), (Egg, prepared), (Cheese, full fat), (Water, coffee, tea))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
14	((Egg, prepared), (Bread, rolls, etc, home-made), (Vegetable products), (Water, coffee, tea))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
15	((Bread, rolls, etc, home-made), (Fatty fish, prepared), (Fruit and berry products), (Margarine and butter))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
16	((Bread, rolls, etc, home-made), (Fatty fish, prepared), (Fruit and berry products), (Water, coffee, tea))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
17	((Bread, rolls, etc, home-made), (Fatty fish, prepared), (Margarine and butter), (Water, coffee, tea))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
18	((Bread, rolls, etc, home-made), (Fruit and berries, raw/fresh), (Juice, fruit drink, soda etc), (Margarine and butter))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
19	((Bread, rolls, etc, home-made), (Fruit and berries, raw/fresh), (Juice, fruit drink, soda etc), (Water, coffee, tea))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
20	((Bread, rolls, etc, home-made), (Fruit and berries, raw/fresh), (Vegetables, raw and frozen), (Legumes))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356

Figure 13: Top 20 association rules of the format (X ⇒ y) from experiment 2.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Bread, gluten-free, white, with water, home-made))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
2	((Sesame seeds, without shell), (Butter), (Sunflower seeds), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
3	((Mayonnaise, full fat, 80 % fat), (Butter), (Cheese, hard, rich, Norvegia), (Water, tap))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
4	((Mayonnaise, full fat, 80 % fat), (Butter), (Milk, semi-skimmed, 1,2 % fat), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
5	((Butter), (Linseeds, flax seeds, crushed), (Cheese, hard, rich, Norvegia), (Water, tap))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
6	((Butter), (Sunflower seeds), (Linseeds, flax seeds, crushed), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
7	((Sesame seeds, without shell), (Butter), (Linseeds, flax seeds, crushed), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
8	((Egg, organic, boiled), (Butter), (Sunflower seeds), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
9	((Sesame seeds, without shell), (Butter), (Cheese, hard, rich, Norvegia), (Egg, organic, boiled))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
10	((Egg, organic, boiled), (Butter), (Linseeds, flax seeds, crushed), (Cheese, hard, rich, Norvegia))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
11	((Sesame seeds, without shell), (Bread, semi-coarse (25-50 %), industrially made), (Sunflower seeds), (Linseeds, flax seeds, crushed))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
12	((Tea, black, infusion), (Water, tap), (Porridge, prepared with oats and milk), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
13	((Tea, black, infusion), (Water, tap), (Coffee, infusion), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
14	((Water, tap), (Coffee, infusion), (Porridge, prepared with oats and milk), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
15	((Tea, black, infusion), (Coffee, infusion), (Porridge, prepared with oats and milk), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.022472	0.410112	0.022472	1.0	2.438356
16	((Water, tap), (Sunflower seeds), (Tea, black, infusion))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
17	((Sunflower seeds), (Soft drinks, cola drinks, artificially sweetened), (Water, tap))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
18	((Water, tap), (Sesame seeds, without shell), (Tea, black, infusion))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356
19	((Sunflower seeds), (Sesame seeds, without shell), (Water, tap))	(Had symptoms)	0.044944	0.410112	0.044944	1.0	2.438356
20	((Sunflower seeds), (Sesame seeds, without shell), (Tea, black, infusion))	(Had symptoms)	0.028090	0.410112	0.028090	1.0	2.438356

Figure 14: Top 20 association rules of the format (X ⇒ y) from experiment 3.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Sweets mix, without chocolate))	(Had symptoms)	0.023256	0.22093	0.023256	1.0	4.526316
2	((Water, tap), (Bread, 2/3 wholemeal flour, water, home-made))	(Had symptoms)	0.023256	0.22093	0.023256	1.0	4.526316
3	((Soft drinks, cola, with sugar), (Margarine, fat spread, vegetable fat and butter, 40 % fat, Brelett))	(Had symptoms)	0.023256	0.22093	0.023256	1.0	4.526316

Figure 15: All association rules of the format (X ⇒ y) from experiment 4. Patient 1.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Beef, minced meat, pan-fried, without fat))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
2	((Corn starch), (Water, tap), (Salt, table, iodized))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
3	((Cheese, hard, rich, Norvegia), (Coffee, infusion), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
4	((Cheese, hard, rich, Norvegia), (Egg, boiled), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
5	((Water, tap), (Cheese, hard, rich, Norvegia), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
6	((Coffee, infusion), (Bread, gluten-free, coarse, with water, home-made), (Egg, boiled))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
7	((Coffee, infusion), (Bread, gluten-free, coarse, with water, home-made), (Margarine, soft, spread with sterols, 35 % fat, Vita Pro aktiv))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
8	((Water, tap), (Coffee, infusion), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.054945	0.351648	0.054945	1.0	2.84375
9	((Water, tap), (Egg, boiled), (Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
10	((Water, tap), (Bread, gluten-free, coarse, with water, home-made), (Margarine, soft, spread with sterols, 35 % fat, Vita Pro aktiv))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
11	((Cheese, hard, rich, Norvegia), (Coffee, infusion), (Egg, boiled))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
12	((Water, tap), (Cheese, hard, rich, Norvegia), (Egg, boiled))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
13	((Water, tap), (Egg, boiled), (Coffee, infusion))	(Had symptoms)	0.032967	0.351648	0.032967	1.0	2.84375
14	((Water, tap), (Coffee, infusion), (Margarine, soft, spread with sterols, 35 % fat, Vita Pro aktiv))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
15	((Corn starch), (Salt, table, iodized), (Tomato ketchup))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
16	((Corn starch), (Water, tap), (Tomato ketchup))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
17	((Bread, gluten-free, coarse, with water, home-made))	(Had symptoms)	0.054945	0.351648	0.054945	1.0	2.84375
18	((Water, tap), (Salt, table, iodized), (Tomato ketchup))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
19	((Corn starch), (Salt, table, iodized), (Tomato ketchup), (Beef, minced meat, pan-fried, without fat))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375
20	((Corn starch), (Water, tap), (Salt, table, iodized), (Beef, minced meat, pan-fried, without fat))	(Had symptoms)	0.021978	0.351648	0.021978	1.0	2.84375

Figure 16: Top 20 association rules of the format (X $\Rightarrow$ y) from experiment 4. Patient 2

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
1	((Avocado, raw))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
2	((Cheese, ripened, Brie))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
3	((Egg, organic, boiled), (Linseeds, flax seeds, crushed), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
4	((Egg, organic, boiled), (Linseeds, flax seeds, crushed), (Sunflower seeds), (Butter))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
5	((Mackerel, warm smoked), (Egg, organic, boiled), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
6	((Mackerel, warm smoked), (Egg, organic, boiled), (Sunflower seeds), (Butter))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
7	((Egg, organic, boiled), (Sunflower seeds), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.045455	0.25	0.045455	1.0	4.0
8	((Mackerel, warm smoked), (Linseeds, flax seeds, crushed), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
9	((Mackerel, warm smoked), (Linseeds, flax seeds, crushed), (Sunflower seeds), (Butter))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
10	((Soft drinks, cola drinks, artificially sweetened), (Linseeds, flax seeds, crushed), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
11	((Sweet pepper, green, raw), (Linseeds, flax seeds, crushed), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
12	((Soft drinks, cola drinks, artificially sweetened), (Linseeds, flax seeds, crushed), (Sunflower seeds), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
13	((Sweet pepper, green, raw), (Linseeds, flax seeds, crushed), (Sunflower seeds), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
14	((Mackerel, warm smoked), (Sunflower seeds), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
15	((Soft drinks, cola drinks, artificially sweetened), (Sunflower seeds), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
16	((Sweet pepper, green, raw), (Sunflower seeds), (Sesame seeds, without shell), (Butter))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
17	((Mackerel, warm smoked), (Egg, organic, boiled), (Cheese, hard, Jarlsberg), (Linseeds, flax seeds, crushed))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0
18	((Egg, organic, boiled), (Linseeds, flax seeds, crushed), (Sesame seeds, without shell), (Cheese, hard, Jarlsberg))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
19	((Egg, organic, boiled), (Sunflower seeds), (Cheese, hard, Jarlsberg), (Linseeds, flax seeds, crushed))	(Had symptoms)	0.034091	0.25	0.034091	1.0	4.0
20	((Mackerel, warm smoked), (Egg, organic, boiled), (Sesame seeds, without shell), (Cheese, hard, Jarlsberg))	(Had symptoms)	0.022727	0.25	0.022727	1.0	4.0

Figure 17: Top 20 association rules of the format (X $\Rightarrow$ y) from experiment 4. Patient 4.

NTNU
Norwegian University of
Science and Technology