Dissertation

Activation of Patients with Chronic Diseases

Submitted by

Magdalena Holter

BSc. MSc.

for the Academic Degree of

Doctor of Medical Science

(Dr. scient. med.)

at the

Medical University of Graz

Institute for Medical Informatics, Statistics and Documentation

under the Supervision of

Univ.-Prof.[in] Dipl.-Ing.[in] Dr.[in] techn. Andrea Berghold

Univ.-Prof. Dr.med.univ. Andreas Wedrich

Research Prof. Priv.-Doz. Mag.rer.nat. Dr.rer.nat. Alexander Avian

2024

Declaration

I hereby declare that this thesis is my own original work and that I have fully acknowledged by name all of those individuals and organisations that have contributed to the research for this thesis. Due acknowledgement has been made in the text to all other material used. Throughout this thesis and in all related publications I followed the "Standards of Good Scientific Practice and Ombuds Committee at the Medical University of Graz".

Magdalena Holter
04.04.2024 Graz

Disclosure

Part of this thesis will be published in: Holter M, Avian A, Weger M, Strini S, Michelitsch M, Brenk-Franz K, Wedrich A, Berghold A. The utility of the Patient Activation Measure administered in an interview setting. Quality of Life Research. In press.

Andrea Berghold and Alexander Avian are affiliated with the Institute of Medical Informatics, Statistics and Documentation of the Medical University of Graz. Andreas Wedrich, Martin Weger, Sanja Strini and Monja Michelitsch are affiliated with the Department of Ophthalmology of Medical University of Graz. Katja Brenk-Franz is affiliated with the Institute of Psychosocial Medicine, Psychotherapy and Psychooncology, Jena University Hospital, Jena, Germany.

Acknowledgments

This journey would not have been possible without the skills, knowledge and confidence of countless individuals and I take this opportunity to express my gratitude.

First and foremost, I am deeply grateful for my supervisors, whose expertise and mentorship have been very valuable throughout all the years of my doctoral studies. Special thanks go to Andrea Berghold for generously providing all kinds of different support throughout the years. Thank you Alexander Avian, for the endless discussions. Thank you Andreas Wedrich, for making this cooperation possible and providing support even in difficult times. Thank you all for your valuable insights, constructive feedback, and encouragement in this journey.

Regarding organisational and financial support, I want to thank the Doctoral School Sustainable Health Research (DS SHR) of the Medical University of Graz.

My gratitude extends to all the staff members from the Department of Ophthalmology, whose dedication and hard work made this study feasible in such a magnitude.

Moreover, I am thankful for all the patients who participated in this study, without whom this thesis would have not been possible. Their cooperation and willingness to contribute in the study have been crucial to the success of this research.

Last but not least, I am grateful to my friends and family, who supported and inspired me throughout the years and provided valuable insights from different perspectives.
The journey of my doctoral studies has been shaped by the contributions, whether big or small, from everyone involved. The support, encouragement, and belief in this research by all of you have made all the difference and I am forever grateful for each one of you.

"*Change is always given to the searchers who explore*"

"Goods", iamamiwhoami, 05.05.2012, ©TWIMC

Contents

Abbreviations

AICc: Akaike information criterion corrected
BMI: Body mass index
CFI: Comparative fit index
CI: Confidence interval
CTT: Classical test theory
DIF: Differential item functioning
DME: Diabetic macular edema
EM: expectation-maximization
GPCM: General partial credit model
ICC: Item characteristic curves
IIC: Item information curves
IRT: Item response theory
LR: Likelihood ratio
MML: Marginal maximum likelihood
NCDIF: Non-compensatory differential item functioning
RMSEA: Root mean square error of approximation
RVO: Macular edema due to retinal vein occlusion
SABIC: Sample adjusted Bayesian information criterion
SES: Socioeconomic status
SRMSR: Standardized root mean square residual
TLI: Tucker-Lewis index
PAM®: Patient activation measure
PROMs: Patient-reported outcome measures
WHO: World Health Organisation

Nomenclature

PARAMETERS	UNITS	DESCRIPTION
ρ	0 to 1	Guessing parameter
j	1 to ∞	Individual answering item
α	0 to 1	Internal consistency (reliability)
i	1 to ∞	Item
b	$-\infty$ to ∞	Item difficulty
a	$-\infty$ to ∞	Item discrimination
θ	$-\infty$ to ∞	Person ability
g	1 to ∞	Response category
τ	$-\infty$ to ∞	Threshold parameter

List of Figures

List of Tables

Kurzzusammenfassung

Einleitung

Patientenaktivierung beschreibt die Zuversicht, die Fähigkeit und das Wissen im Umgang mit der eigenen Krankheit. Diese kann mit dem Patient Activation Measure ® (PAM) erfasst werden. Das Hauptziel der Dissertation ist die Untersuchung der psychometrischen Eigenschaften des deutschen PAM-Fragebogens bei chronisch kranken Patient*innen.

Methode

Ambulante Patient*innen mit Makulaödem nahmen an dieser fragebogenbasierten Querschnittsstudie teil. In Interviews wurden Patientenaktivierung, selbst eingeschätzter Gesundheitszustand, Selbstwirksamkeit, Lebensqualität, allgemeine Stimmung und Suche nach Gesundheitsinformationen gemessen. Die psychometrischen Eigenschaften wurden mittels probalistischer Testtheorie (IRT), Cronbachs α und Trait-Trait-Korrelationen untersucht.

Ergebnisse

554 Patient*innen wurden in die finale Analyse einbezogen. Das Durchschnittsalter betrug 69 (IQR: 62.0-76.0) Jahre. Die Stichprobe wies einen mittleren Aktivierungswert von 74.1 (SD:13.7) auf. Alle Items zeigten Deckeneffekte. Die empirische Reliabilität des IRT-Modells und Cronbachs α lagen bei 0.75. Der PAM® zeigte eine Spearman-Korrelation von 0.54 ($p<.001$) mit Selbstwirksamkeit, 0.51 ($p<.001$) mit Lebensqualität, 0.34 ($p<.001$) mit allgemeiner Stimmung und 0.27 ($p<.001$) mit Suche nach Gesundheitsinformationen.

Diskussion

Patientenaktivierung kann mit dem PAM® reliabel und valide bei Patient*innen im klinischen Alltag erfasst werden. Zusätzlich ist eine objektive Messung im Rahmen eines Interviews möglich. Patient*innen mit geringer Patientenaktivierung lassen sich durch den Fragebogen gut identifizieren. Die Messqualität verschlechtert sich bei Patient*innen mit hohem Aktivierung, da die PAM®-Items in diesem Bereich wenig Information bieten. In dieser Studie lag eine hohe Patientenaktivierung vor, was bereits darauf hindeutet, dass sie sich mit ihrer eigenen Krankheit auseinandersetzen.

Abstract

Introduction

Patient activation describes the confidence, skills and knowledge in managing ones own disease. It can be assessed by the Patient Activation Measure® (PAM). The main objective of the thesis is to investigate the psychometric properties of the German PAM® survey in patients with chronic diseases.

Methods

Outpatients with macular edema participated in this questionnaire-based cross-sectional study. Through interviews, patient activation, self-rated health, self-efficacy, quality of life, general mood and health information seeking were assessed. Psychometric properties were investigated using item response theory (IRT), Cronbach's α and trait-trait correlations.

Results

554 patients were included in the final analysis. The median age was 69 (IQR: 62.0 – 76.0) years. The study sample showed a mean overall activation score of 74.1 (SD: 13.7). All items showed ceiling effects. Empirical reliability from the IRT model and Cronbach's α were 0.75. The PAM® survey showed a Spearman correlation of 0.54 ($p<.001$) with self-efficacy, 0.51 ($p<.001$) with quality of life, 0.34 ($p<.001$) with general mood and 0.27 ($p<.001$) with health information seeking.

Discussion

The PAM® survey reliably assesses patient activation in an everyday clinical setting, and it is also a valid measurement tool. Objective assessment in an interview setting with the PAM® survey is possible. Patients with low patient activation are identified with the questionnaire. Measurement quality deteriorates for those with high activation ability level, as items offer little information to distinguish between them. In general, patients were highly activated, indicating they are already engaged in the management of their health.

Chapter I

Introduction

1.1 Aim and structure of the thesis

Disease and health problems are omnipresent experiences that challenge individuals to strive for physical fitness. For a long time, it was overseen that additionally, psychological and social factors are important aspects to reach overall well-being. An emerging psychological factor in health research is *patient activation*, describing the knowledge, skills, and confidence in managing one own's disease. The concept of patient activation emphasizes to not only support individuals in achieving physical fitness, but also to make them realize they are the person responsible for their health status. This is especially important for patients suffering from a chronic disease.

In this thesis, firstly the concept of patient engagement and its measurement in various cultural and demographic settings will be explored. Specifically, the history of health and patient engagement will be discussed, leading to a concept of patient activation. Secondly, the measurement of patient activation in distinct populations around the world are described. Thirdly, the methodology and results of a cross-sectional study carried out within the scope of this thesis is described. Fourthly, the results of this study are presented, including an analysis of the measurement of patient activation. Fifthly, the results of the cross-sectional study will be compared to former studies. The goal is to contribute to the discussion on how we can better support individuals in their pursuit of well-being, with a focus on patient activation.

1.2 Background

Chronic disease is long lasting and can develop slowly over time and may worsen gradually over an extended period, ranging from months to years, depending on the type of disease. The World Health Organization (WHO) defines four main types of chronic diseases: (1) cardiovascular diseases like coronary heart disease, (2) cancer, (3) chronic respiratory diseases like chronic obstructed pulmonary disease and asthma, and (4) diabetes [1]. These chronic diseases are typically not self-resolving, are generally not (completely) cured and are the main causes of death in the world [2]. Alarmingly, the number of patients suffering from a chronic disease is steadily rising worldwide [3]. In Austria it is estimated that 38% of the population suffer from a chronic disease (Statistik Austria, 2019) [4]. Importantly, for individuals with a chronic disease comes greater responsibility, thus it is substantial that patients can manage their health [5]. On the one hand, this involves the treatment of their chronic disease. Usually this includes complex treatment plans, multiple yearly visits at medical doctors and/or hospitals and changes in life style [3]. On the other hand, psychosocial factors have found to be very important in managing treatments and everyday life for patients suffering from a chronic disease to reach a better health outcome [6]. Speaking about diseases, the question arises how '*health*' is defined.

1.3 Health models

For a long time, health was defined as the opposite from disease: it described a state where an individual is free from disease or malady [7]. Nowadays, the WHO defines health as '*Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.*' [1]. This definition marked a shift from a purely biomedical perspective, acknowledging the connection of health with broader aspects of life. Subsequent frameworks, such as the Ottawa Charter for Health Promotion, emphasized the importance of social, economic, and environmental factors in health [1]. A dynamic perspective on health emerged, viewing health as the ability to adapt [8]. Further, health was described as a dynamic state of well-being, highlighting individuals' responsibility in managing their health and the role of treatment strategies in optimizing potentials [9]. Reflecting on these contributions, a reconsidered 2020 definition emerged: '*Health is the dynamic balance of physical, mental, social, and existential well-being in adapting to conditions of life and the environment.*' This definition underscores the dynamic, multidimensional nature of health, emphasizing balance, adaptation, and continuous evolution [10].

Since health and disease were regarded from different perspectives and defined distinctively over the progress of humanity, several theoretical models emerged, describing the etiology and nature of diseases [7]. The predominant model influences the perception in the society and can have many important implications. Depending on the underlying health perspective or model that is present in society, the available treatments for a particular health condition may vary. Attitudes and cultural beliefs towards diseases are influenced by the predominant model as well. [7]. There are several theoretical models stating different risk and protective behaviors and factors for one's health. For centuries, the biomedical model was predominant [11]. It views health as the absence of disease and focuses on diagnosing and treating physiological malfunctions. Disease is always caused by pathogens. The healing process is solely a treatment of symptoms. The medical doctor has the knowledge and skill to cure the disease. The patient merely has to follow his/her instructions [11]. Over centuries, the nature of diseases changed and it became more inefficient to explain them solely with the biomedical model [7]. More and more chronic diseases emerged, instead of infectious diseases troubling mankind for centuries [3]. An extended model was proposed, including psychological and social factors, the biopsychosocial model [11]. Health was no longer defined as the absence of disease, but as a state of well-being. Biological, psychological and social facets and their interaction determine physical health and disease. An unbalance in those facets can cause a disease. Healing includes more than just treating symptoms: all facets have to be regarded and changed, if necessary. The medical doctor helps the patient in this process. The patient himself is responsible for a good health outcome [11]. Resources in the biological facets are for example a healthy and balanced diet, exercise and enough sleep [7], whereas psychological resources for health are self-efficacy and positive health behavior [12]. Social support is a resource regarding the social facet [13]. Known biological risk factors are e.g. older age and high body weight [4]. Depressiveness [14] and addictive behavior [4] are examples for psychological risk factors. In the social facet, socioeconomic status (SES) is a known risk factor, a low SES predicts a poor health outcome [15]. These risk and resource factors are only some examples of impacts on our health.

1.4 Healthcare model

However, since even more chronic diseases are present, new challenges arose, leading to the development of a framework for delivering healthcare that focuses on the management of chronic diseases. To improve the care of patients with chronic diseases, the chronic care model [16] was proposed. It is designed to address the challenges faced by healthcare systems in delivering high-quality care to individuals with chronic diseases. The model identifies six key elements that are necessary to improve the management of chronic diseases: health system, health care organization, delivery system design, clinical information systems, decision support, and self-management support. It emphasizes the importance of a coordinated and integrated approach to care delivery of the health system, health care organization and healthcare personnel. According to that model, patients need to be equipped with the knowledge, skills and resources necessary to manage their conditions effectively. This may include education about their disease, goal-setting and support in making behavioral changes. Informing patients and assisting them in self-care is a critical step toward achieving an improvement in health status. In the case of the chronically ill, a favorable health status can be maintained even in the absence of continuous medical care [16]. Comparing health and healthcare model, the biopsychosocial model focuses on understanding the patient's experience of disease, while the chronic care model focuses on the management of chronic diseases.

1.5 Patient engagement

Since the predominant health model has shifted away from the biomedical model towards the biopsychosocial and in the area of healthcare models the chronic care model gained importance, more active involvement of patients is demanded [5]. This involves chronically ill patients having certain knowledge about their disease and abilities to manage it [16]. Moreover, patients need confidence to live with the disease [12]: everyday life may be different than before, due to changes required in routine behavior, which in the best case becomes a normal part of life [16].

Different terms are used to refer to that one concept of patients themselves being an active part in their health care [17]. It is called patient *engagement*, *participation* or *empowerment*. The latter is more associated to the patients' process of taking responsibility of their own health. *Participation* is more associated to the patients' active participation and involvement in decisions about their healthcare [17]. For patient *engagement* there are many definitions, but they all have one thing in common: the facilitation and strengthening of individuals using services to build health, as well as health care policy and practice [18]. Substantially, patient engagement, participation and empowerment refer to the same characteristics of patients' involvement in their own health care. In this study, the term '*patient engagement*' will be used.

As mentioned before, the definition of patient engagement is not unique, it differs across different studies. Most definitions of patient engagement include self-determination and control over his/her own health and healthcare, these are not determined by the medical doctors [18]. Moreover, most definitions include personal control and self-efficacy as crucial characteristics and also focus on the capability of patients in making decisions about their health and behavior [18]. That includes everyday life facets of chronically ill patients influencing their health [18]. However, definitions of patient engagement vary in the extent a patient is involved in his/her own care to count as engaged [19]. Moreover, definitions of patient engagement also vary in the 'time perspective': sometimes it is seen as a short term characteristic, other times as long term [19]. Furthermore, patient engagement sometimes is more described as a process and other times more as an outcome [18].

In order to reach clarity about the definition of patient engagement, a concept analysis was conducted. Four different key aspects of patient engagement were identified: personalization, access, commitment and therapeutic alliance [17] *Personalization* refers tailoring the interventions as close as possible to the patient's unique desires and circumstances. Shared decision making should be included in that process as well as modifying information and resources to the patient's level of responsiveness depending on interests, abilities and life circumstances. *Access* refers to the patient's ability to receive guidance and consistent quality care. It also refers to gaining relevant health information. *Commitment* refers to cognitive and emotional factors of a patient to participate in treatment activities. *Therapeutic alliance* refers to medical doctor-patient relationship, including their communication. Patient engagement is seen as a process and behavior [17].

In theory, more engaged patients should help improve their health and well-being by making thoughtful, informed healthcare decisions. Moreover they should need healthcare services less and cause lower costs [18]. To gain insight if patient engagement is a protective factor for patients with a chronic disease, a systematic review was conducted [20]. Within it is scope, studies including an intervention aimed to improve engagement, vs. no such intervention were compared. Nearly all studies included in the systematic review reported an increase in patient engagement. Moreover, it was shown that self-reported health status improved with all patient engagement interventions. In some of these studies even improvement of clinical markers were reported: for people suffering from diabetes, the long term blood sugar value HbA1c decreased. Patient engagement was also found to improve blood pressure.

The influence of patient engagement was investigated further: different interventions involving patient engagement vs. disengagement in their own treatment were compared. Clearly, a preference for patient involvement was shown to lead to improved health outcomes in patients with chronic diseases [21]. Improvement of depression symptoms and better self-rated health was demonstrated in more engaged chronic disease patients [21]. So, in general a positive effect of more engaged patients on health outcomes was shown. Furthermore, the impact of teaching patients effective communication skills as part of engagement in healthcare interactions was investigated as a part of patient engagement [22]. Most communication interventions impacted patients' active participation in a positive way in healthcare interactions. However, the psychosocial well-being or health outcome were not always effected [22].

Moreover, gaining insight about patients' engagement in their health management gives the opportunity to design preventive targeted interventions to optimize treatment regimes, a critical factor in enhancing the quality of care [23]. Interestingly, for judging the success of a treatment, having a desirable health status mostly relies on physical health outcomes, but other outcomes like engagement are important as well [18]. For example, usually the biggest fear of patients with an eye disease is loosing sight. They are not so worried about their intraocular eye pressure as is their eye doctor, they want to be able to drive their car [24].

1.6 Measurement of patient engagement

One important point not mentioned yet while talking about the effect of engaged patients on health outcomes, is how patient engagement can be assessed. Usually, the aim in primary care is to measure health status, health engagement and health perceptions [25]. A systematic review concluded that no questionnaire exists, measuring all three aspects [25]. Most of them assess the current state and some aim to measure changes in one of those three aspects directly [25]. Especially, measurement of patient engagement is important, because relationship and interactions between patients and medical doctors may be altered by knowing more about the patient's struggles [26]. Originally, Patient-Reported Outcome Measures' (PROMs) [1] emerged to assess the efficacy of healthcare interventions [27]. In primary care, there are several PROMs that can be used for specific diseases and issues, however, many generic PROMs exist as well, which can be used in multiple populations and diseases [25]. Searching through literature, there are many different questionnaires assessing patient engagement, defined in the same or distinct way. [18, 23, 28].

The 'Patient Activation Measure' (PAM®) measures knowledge, skills and confidence in managing one own's disease [12]. Similar, the 'Effective Consumer Scale' assesses the main skills, attitudes and knowledge individuals need to effectively manage and participate in their healthcare [29]. The 'Altarum Consumer Engagement Measure' broadens the concept of patient engagement in health and healthcare decisions through including modern information sources, such as online health resources [30]. It assesses confidence, ability, and the perceived role in managing one's health as well as using health related information and the intention to engage in one's treatment decisions.

A different perspective of patient engagement is captured by the 'Patient Health Engagement Scale' [23]. It measures cognitive, emotional, and behavioral attitudes of patient's engagement. Furthermore, another angle is taken by the 'Health Care Empowerment Inventory': it measures multiple perspectives of engagement in health care, including individual background factors, personal resources and emotional processes [31]. Similar, the 'Patient Engagement Index' assesses influences of patient's psychological, physical, and socioeconomic conditions on patient engagement [32].

Moreover, there are questionnaires developed for specific patient populations. The 'Patient Assessment of Chronic Illness Care' was designed especially for individuals with a chronic disease, based on the chronic care model [28, 16]. The goal was to support and improve the patient-centered care. It assesses patients' involvement in decision-making, setting specific goals, possible barriers in adhering to treatments and follow-up appointments. This questionnaire was extended by the '5As' (assess, advise, agree, assist, arrange) [28] investigating behavioral change after behavioral counseling. Another questionnaire developed for chronic disease patients is the 'Health Education Impact Questionnaire'.

[1] PROMs are usually self-report questionnaires. Their goals is to assess certain aspects of the patient's health.

It was developed to assess the impact of any patient education intervention on patient engagement [33]. It measures self-management skills important for people suffering from a chronic disease like everyday and health-directed behavior, emotional characteristics and cognitive resources.

Furthermore, ‚Patient Empowerment in Long-Term Conditions' was developed specifically for patients with long-term conditions in the primary care setting [34]. It assesses positive attitude and sense of control, knowledge and confidence in decision making, and enabling others. The ‚Barriers to Self-Care in Multiple Long-Term Conditions' takes the opposite perspective: barriers in being engaged in his/her own health management are regarded [35].

Moreover, there are many questionnaires developed for patients suffering from a specific disease. The 'Diabetes Empowerment Scale' was designed especially for patients suffering from diabetes, assessing patient engagement as psychosocial self-efficacy [36]. The 'Empowerment Scale' was developed for mental healthcare , includes self-efficacy and feelings of power [37]. The 'Patient Empowerment Scale' was designed for cancer patients and assesses the use of coping resources, an aspect of personal control [37]. The 'Genetic Counselling Outcome Scale' was developed for individuals with genetic conditions, assessing perceived personal control, hope and emotional regulation [18]. The 'Patient Motivation Questionnaire for Patients with Athritis' assesses motivation and engagement particularly in those patients, since the nature of the disease is different than most other chronic diseases and other scales are not meaningful for this population [37].

As shown, there are many different measures to assess patient engagement - and the questionnaires described are only a few. There are so many more questionnaires and scales aiming to assess patient engagement, in particular for specific diseases [37]. However, in practice, not every questionnaire is useful [38]. There are several criteria a questionnaire should fulfill, to be judged as a good assessment of patient engagement. First, since there is no unique definition of 'patient engagement', it should be clearly defined prior to designing the questionnaire so that the attributes to be captured are not random. It should capture all important factors of patient engagement. Secondly, the questionnaire should measure patient engagement precisely and consistently over time (reliability). Thirdly, it should capture patient engagement and not any other characteristic (validity). A questionnaire fulfilling these three criteria is needed to measure how patients suffering from a chronic disease cope with their challenges in daily life and health treatments. Of all the questionnaires described, only a few meet this criteria: most are lacking a sound definition of patient engagement and/or quality of assessment is low [25, 26]. Notable positive questionnaires are the 'Patient Health Engagement' scale [23], 'Patient Assessment of Chronic Illness Care', [28], 'Patient Engagement Index' [32] and 'PAM® survey' [12]. Since the PAM® [12] is the main focus of this study, it is development will now be described. This was done in four steps.

1.7 Patient Activation Measure®

1.7.1 Introduction of patient activation

In a first step, the authors from the U.S. aimed for an appropriate definition [12]. A literature search was done, focusing on publications concerning self-care, self-management, medical doctor-patient communication and the usage of different information in health related decisions. Overall, six different characteristics of individuals emerged on having a positive effect on health status: ability to self-manage symptoms; engaging in activities to preserve physical abilities and to decrease health deterioration; involvement in choices of treatment and diagnostics; cooperation with health care providers; selecting health care providers due to accomplishments or quality; and how to use the medical system. These six characteristics were used in an expert consensus process and in patient focus groups. Additionally, these patients should think of three criteria: beliefs, knowledge and skills and their requirement for an efficacious management of a chronic disease. The experts were consulted regarding the importance of the six characteristics having a positive effect on health status in combination with the three criteria (beliefs, knowledge, skills) and to establish an order according to their importance. The experts were able to make further additions to these characteristics and criteria. The original criteria to be concerned of beliefs, knowledge and skills was extended to include availability of emotional support. The same six characteristics having a positive effect on health status were explored by two groups of patients with a chronic disease. They were also able to make further additions. Similar results were found compared to the experts. However, emotional support in health management was not seen as important as by the experts [12].

Taking the results of the literature, experts and focus groups into account, a conceptual definition for patient engagement was developed and called '*patient activation*' [12]. It describes the *confidence*, *knowledge* and *skills* of individuals *managing their health.* Highly activated patients believe in their crucial part in self-management, in cooperation with health care providers in preserving their health. They know how to control their health status, engage in activities to preserve physical abilities and to decrease health deterioration. Moreover, highly activated patients possess the skills and behavioral repertoire to manage their disease, cooperate with health care providers, preserve physical abilities and select health care providers due to accomplishments or quality [12].

Patient activation is a part of patient engagement [39]. The later is a more broader construct including interventions and their impact as well, while activation is more about thoughts, feelings and believes. The major point is the willingness and ability of the patient taking an active role in their health care: activation is about the actual patient's behavior. Further, patient activation is more about patients interaction with the healthcare system, while patient engagement includes their lifes outside the healthcare system as well: it is more about integrating the disease in their everyday life [40]. Being activated can be seen as a process that has to be made before a patient can be engaged.

1.7.2 Questionnaire development

In a second step, the definition of patient activation was used to develop a questionnaire assessing patient activation [12]. By operationalizing this definition, an item pool of 80 items was developed. After 20 patients with a chronic disease answered it in 'thinking along aloud' mode, they were reduced to 75 items. 100 patients with different chronic diseases answered those items in a pilot study through a telephone interview. Demographic data and health status were assessed as well. These data were analyzed by a Rasch-model, which aims to explain the answer pattern on items by the underlying patient activation of participants. Items itself can be investigated, how well they fit and perform. With these results the number of items were reduced to 21. The questionnaire was found to be unidimensional, meaning that only one latent factor, presumably patient activation, explains the variance in the answer patterns. The model reliability was 0.87 and Cronbach's α 0.87, indicating good reliability. 30 participants were part of assessing retest reliability which could be confirmed by retest activation scores close to the original ones. To gain insight about criterion validity, experts interviewed patients with high and low activation and classified them on the PAM® questionnaire: 83% (25 of 30) classifications were correct [12]. The Rasch-model also assesses the difficulty of items, representing the effort or the level of patient activation needed from patients to endorse it. While designing this questionnaire, the aim was to evaluate a wide range of patient activation. However, the items in this questionnaire were ordered by difficulty and it was discovered that they did not cover the lowest and highest levels of patient activation.

1.7.3 Questionnaire refinement

In a third step the questionnaire was extended and modified [12]. New items were created to better capture activation in the lower and higher ability range. The resulting 22 itmes were answered by N=486 individuals with or without a chronic disease. Despite the items being modified, the same order arose again regarding difficulty. From the results believes, knowledge and skills clearly emerged as important factors influencing patients in being active in their health management. Taking into account the insights from the literature search and the questionnaire results, four levels of activation were defined: The *first* is about believing the magnitude of his/her influence on one's own health status. Patients don't engage in the management of their own health. The *second* level is about gaining the required confidence and knowledge to take action. Application of medication is familiar as well as healthy lifestyle changes. Patients are confident in talking to their medical doctors and when to look for assistance. Some confidence and knowledge is still missing. The *third* level is about taking action. Health is maintained or improved, further symptoms are avoided and they can be handled on one's own. Some confidence is still lacking. The *fourth* level is about maintaining a healthy lifestyle even during stressful periods. Problems can be handled by patients on their own and their health problems are not impeding their life [12].

1.7.4 National questionnaire application

In a fourth step the questionnaire was tested in a US national sample [12]. The goal was to ensure the performance for different populations and to investigate construct validity of the questionnaire. N=1515 individuals older than 45 years were interviewed per telephone. Same results as in step three were obtained: items maintained the same order of difficulty and the same developmental hierarchy for those with and without chronic disease. Good reliability and construct validity were found in this sample. Individuals who were highly activated showed better health status, less visits at medical doctors and emergency rooms and less hospital stays [12].

1.7.5 Development of a short-form questionnaire

One further aim while developing the questionnaire was that it is easy applicable in a daily clinical setting [41]. In clinical practice patient education usually is done in a general way, all individuals get the same information [12]. However, a patient who has no insight in his or her disease, is not receptive for health actions. To be able to quickly capture the level of patient activation, the number of items in the questionnaire were further reduced. This was done using the same data as in the questionnaire development earlier described under the fourth step. Items were selected through content-related and statistical considerations. One item was deleted at a time and the impact on the overall measurement precision and reliability was investigated. Reduction of items was stopped, as a further deletion of items led to a worse measurement precision. Using the same data set, the newly developed 13 item version was compared to the 22 items. The short version demonstrated a slightly lower reliability in some subgroups, but it was still acceptable. Women, younger individuals, more educated participants and higher self-reported health status showed higher PAM® scores. Ethnicity showed an influence as well [12].

Since its development, the PAM® is mainly used in 3 distinct settings: first to distinguish the population and determine risk factors for diseases, secondly to customize the support to the level of patient activation and thirdly to assess the impact of an intervention [39]. The reliability and validity of the PAM® survey was endorsed by multiple studies using different psychometric methods [42, 43, 44, 45]. It is extensively tested, showing good reliability, validity and is sensitive to capture change in patient activation [25]. Furthermore, the PAM® survey was translated into several languages around the world.

1.7.6 PAM® survey translations

Overall, the PAM® survey exists in over 35 languages (see `https://www.insigniahealth.com/pam/`). In Europe, for example, there is a German [45], Italian [42], Danish [46], Swedish [47], Norwegian [48], Dutch [49], Spanish [50], Portuguese [51] and Hungarian [52] version. To examine these translated versions of the PAM® survey, they were used in a range of different healthcare settings.

To examine the German version [45] of the PAM®, it was applied once in a very broad patient population in primary care in a multi-center study including Germany, Austria and Switzerland. A sample size of N=508 was reached. Principles of classical test theory were used to investigate the psychometric properties. A different order of items regarding their difficulty compared to the original version was found. Reliability was good. Activation showed positive association with self-efficacy, which reinforces the success of the translation by showing convergent validity.

The Italian [42] and European Spanish [50] version of the PAM® survey, were applied to a sample of patients with chronic diseases. Similar, to examine the Dutch version [49], it was applied to people with chronic disease or disability. To examine the Hungarian version [52], it was applied to a sample of the general Hungarian population $\geq$ 40 years, most of them chronically ill. The Swedish [47] and Norwegian [48] versions were applied in a hospital setting. To examine the Danish version [46], it was applied to patients with dysglycaemia, meaning that the blood sugar is either too low or high. To examine the European Portuguese version [51], the questionnaire was applied to diabetes patients.

Overall, the translations of the PAM® survey across Europe succeeded and are comparable in their reliability and validity. In most studies, reliability was investigated using internal consistency. The translated versions showed good reliability [45, 42, 50, 47, 46, 51]. One study investigated test-retest reliability [48], which was also found to be good. Two studies investigated both, internal consistency and test-retest reliability [52, 49]. Reliability regarded as internal consistency was good in both studies, but test-retest reliability was found to be moderate [52, 49].

To gain indices about the validity, besides examining if the translation is unidimensional [45, 42, 50, 52, 47, 46, 51] as the original version, associations with other characteristics were used. Overall, higher activation was associated with higher self-efficacy [45, 50, 53] and physical quality of life [50]. Further, patients with higher education showed more patient activation [47, 49]. Additionally, it was found that male, and healthier participants showed higher patient activation [49]. Moreover, participants with a chronic disease showed lower activation than healthy ones [54]. Furthermore, patient activation showed a weak to moderate association with health literacy [49]. More activated patients showed less lifestyle risks, seeking more health information and participating more in patient education programs [52].

A different order of item difficulty than in the original US version was found in several countries: German [45], Italian [42], Danish [46], Norwegian [48], Dutch [49], Swedish [47], European Spanish [50] and Portuguese [51]. For the Hungarian version [52] this information is not available. It is important to consider that there may be cultural differences between countries in Europe and the US, as well as differences in the healthcare systems. As a result, patients with chronic diseases in different countries may face different challenges when it comes to managing their health and maintaining their health status.

There are further translations in different parts of the world. The Hebrew version [54] of the PAM® was answered in telephone interviews by a nationally representative sample of Israeli adults. It was found to have acceptable reliability. The Turkish version [55] was applied to patients with different chronic diseases. Good internal consistency and high retest reliability was found. They found a different order of item difficulty than in the original US version. The Korean version [56] of the PAM® was examined in patients with osteoarthritis. Good reliability was found. The Malay version [57] was examined in patients with metabolic syndrome. Good reliability was found and it was fairly stable over time. The Chinese version was examined by applying it to patients with hypertension and/or diabetes [58]. Reliability was excellent.

Overall, the translations of the PAM® survey yielded many applicable questionnaires in multiple languages. However, not only different languages and cultures can influence the nature of patient activation, but also different populations.

1.7.7 Patient activation in different populations

A systematic review was conducted featuring chronically ill patients assessing relationship between patient activation and health care behavior [59]. If they were in level 1 or 2 of patient activation, they were hospitalized more often than in level 3 or 4. For example, this effect was shown for diabetic and cardiovascular disease patients. Patients in level 1 or 2 are more likely to use the emergency room. Regarding medication adherence, no conclusion on the association with patient activation could be made. Results from different studies were contradictory, maybe because different methods were used to assess medication adherence [59]. Recently, a similar systematic review and meta-analysis was conducted investigating the relationship between patient activation and healthcare resources [60]. It was found that patients with chronic disease, who have a higher level of patient activation, have a lower risk of hospitalization or requirement of the emergency room. However, in the included studies most of the patients were in activation level 3 or 4 [60]. These findings are consistent to another study, where it was shown that if patients with a chronic disease are less activated, they visit their main health care provider more often [61]. Furthermore, if chronic disease patients are more activated, they tend to use more self-management services and show more self-management behavior [62]. They were also found to be more compliant in taking their medication. Highly activated patients are more satisfied with their care and show more life satisfaction [62]. In addition, they believe in being in control of the management of their own health [63]. A systematic review and meta-analysis of patient activation in people living with chronic diseases showed that patients taking an active role in managing their health, can lead to better use of healthcare services and improved self-care [64]. Studies have found that interventions involving an interdisciplinary team of healthcare providers tend to be more effective at increasing patient activation. To be most effective, interventions should be tailored to the patient's level of activation. While many interventions have shown positive results in increasing patient activation and promoting healthy behaviors like exercise and glucose monitoring, the meta-analysis of randomized controlled trials did not find conclusive evidence to support these findings [64].

Lately, a systematic review and meta-analysis was conducted investigating the impact of behavioral interventions on patient activation in adults with hypertension [65]. Randomized clinical trials evaluating the impact of self-management interventions on patient activation in adults with hypertension were included. An inclusion of a self-management training lead to higher patient activation scores than usual care . The home-based activation intervention increased the PAM® score the most, followed by a community self-management program, and patient activation and motivational interviewing strategies [65].

Furthermore, a systematic review and meta-analysis was conducted, investigating patient activation and self-management in patients with diabetes [66]. The results showed that improving patient activation levels led to significant improvements in self-management and clinical outcomes, including a reduction in HbA1c levels. Interventions that were delivered by peer coaches in community settings and had longer follow-up periods tended to be more effective in improving patient activation levels, self-management, and clinical outcomes. Based on their findings, they conclude that patient activation is a reliable trait for improving self-management and clinical outcomes. The PAM® survey was also used to test the impact of an intervention for people suffering from diabetes, targeting optimal dietary carbohydrate management [67]. Although patients showed generally high levels at baseline, most patients' activation level rose one or two level up after the intervention and showed an increase in diabetes empowerment.

A closer look was taken on patients with different ocular diseases [68]. Patient who missed their scheduled appointments more often had lower patient activation. It was also lower if patients had caregivers than if they had none. A multi-center study including North America, Europe and Australia investigated patient activation in individuals with wet age-related macular degeneration [69]. The majority of participants were at patient activation level 3 or 4, except for Japan where 70% (N=30) were in level 1. Overall, participants were already adopting the necessary behaviors for enhanced treatment outcomes.

Multimorbid older adults were interviewed in-person, including the PAM® survey [70]. Patients with more activation also showed a higher functional status, health care quality, and were more consequent in positive health behaviors.

A systematic review was done, investigating the effectiveness of patient activation interventions on chronic obstructive pulmonary disease self-management outcomes [71]. It was found that patients with higher levels of activation were more successful at self–managing chronic obstructive pulmonary disease and had better clinical outcomes. The interventions improved patient activation levels moderately. A closer look was taken on cancer patients [72]. A strong association was found between perceived patient participation and activation levels, with higher activated patients showing more participation behavior. Moreover, more activated patients reported higher quality of life. The focus of another study were patients with chronic kidney disease [73]. Patients with high activation had fewer symptoms and reported higher quality of life. Further examples of different populations the PAM® survey was applied to, are a sample of patients with osteoarthritis [74], neurological patients [75] and mental health patients [76]

Overall, the PAM® survey is useful in assessing patient activation [25]. Although, the PAM® survey works differently than intended in some languages and populations. To quickly identify when patient's answers start to disagree and what challenges in his/her health management he/she is currently facing [41], the items are ordered according to their difficulty. This ordered structure of the items regarding their difficulties could not be replicated in every sample e.g. [75, 51, 49]. However, by looking at all answers patient activation still can be identified in many languages and populations. It is important to have a closer look at different patient populations and their activation to gain insight about their struggles and to be able to help patients cope with their disease.

1.7.8 Growth of patient activation

Therefore, it is crucial to understand the development of patient activation. Socioeconomic characteristics explain about 5-6% of variance in the PAM® score [77]. A systematic scoping review was conducted investigating psychosocial and psychological factors associated with patient activation [78]. There is little information about how patient activation arises and what factors influence this process. Since mainly cross sectional studies were found, this question could no be solved. 21 psychosocial or psychological factors were found to have a significant relationship with patient activation. Often, self-efficacy, hope, health status, and depression were associated with the PAM® score. The latter is the only factor showing a negative association. Overall, five different clusters could be found: (1) depression; (2) self-assessed health status including quality of life; (3) social context including social support; (4) cognition including self-efficacy and (5) health literacy; and emotions, including emotional feelings and worries about the future.

Furthermore, the relationship between patient engagement, patient activation, treatment adherence, and emotional factors was examined [79]. Suggesting a cluster of factors associated with patient activation: first and foremost, the level of patients' elaboration of their disease, but also the quality of the patient-medical doctor relationship, and positive emotional attitudes towards their health conditions. Higher patient activation was associated with more engagement, better treatment adherence, more positive feelings and a higher perceived quality of the patient-medical doctor relationship. According to another study, self-management regarding self-care behaviors predicted 26% of patient activation [80]. In this study, it was further concluded that this prediction is regardless of sociodemographic or health-related data.

1.8 Research objective

The concept of patient activation has rarely been studied in Austria and Germany, which raises questions about potential cultural differences with other countries. The goal is to gain further insight how a specific questionnaire, the PAM® survey, can be used in a clinical setting in patients with a chronic disease. The German version of the PAM® survey has not yet been analyzed using item response theory (IRT), which is a method for confirming construct validity. IRT allows for comparison of item parameters across different languages and also enables criteria-based interpretation of the results. Furthermore, reliability over different patient activation levels will be investigated.

This study will focus on patients with macular edema, a chronic disease affecting the eye and sight. Patients with this condition are able to impact their health status through their behavior, making them a suitable population for investigating the research question at hand.

The PAM® survey has been used in both self-administered and interview settings [12, 54, 70]. A recent study found [68] that patients with vision loss had normal levels of patient activation when they answered the questionnaire. However, it is not yet clear how the PAM® survey performs when read aloud in a face to face interview setting and not by telephone, which is particularly important for patients with chronic eye diseases. Further research is needed to understand the psychometric properties of the PAM® survey in this context. Additionally, there is interest in exploring the interactions between patient activation and various risk and resource factors across biological, psychological, and social facets.

Chapter II

Methods

2.1 Study design

This is a cross-sectional questionnaire study, consisting of an ad-hoc sample of the population of outpatients suffering from a macular edema, at the Department of Ophthalmology of the Medical University of Graz. After receiving informed consent, each included patient was asked to fill out the questionnaires once. Data collection was carried out from March 2020 until the end of February 2022. After one week of data collection, at the beginning of March, the study had to be paused because of COVID-19. In June 2020 the study could be resumed again under safety conditions. The study was approved by the ethical committee of the Medical University of Graz (ethical committee number: 32-101 ex 19/20).

The implementation of the study was piloted with N=10 patients in February 2020. Overall, feasibility of the study procedures was shown. To assess the duration of completing all questionnaires while they are read out loud, time was measured during these interviews. The overall duration of the interview was around 25 minutes (IQR: 21.2 - 27.2), which seemed acceptable for patients. All intended questionnaires remained in the study.

2.2 Study population

This study focused on patients suffering from macular edema. It is a condition in which fluid accumulates in the macula, the central part of the retina at the back of the eye. This is obstructing vision and can lead to severe visual impairment, if not treated properly.

Since this is a heterogeneous disease with different possible etiologies, the focus was on two types: macular edema due to diabetes or retinal vein occlusion. Both types of macular edema have similar risk factors due to vascular diseases [81]. In people with diabetes, macular edema can be caused by changes in the blood vessels of the retina due to high blood sugar levels. These changes can cause the blood vessels to leak fluid, leading to a swelling in the macula. Other factors that can contribute to the development of macular edema in people with diabetes include high blood pressure, high cholesterol, and poor blood flow to the retina [82]. Diabetic macular edema develops slowly and occurs mostly bilateral. The onset of retinal vein occlusion is sudden and most frequently only one eye is affected. Retinal vein occlusion occurs when one of the veins that carries blood away from the retina becomes obstructed. This can be caused by a number of factors, including inflammation, arteriosclerosis (hardening of the blood vessels), or other conditions that cause blood clots to form. When the vein becomes occluded, blood and fluid can accumulate in the affected area, leading to swelling of the retina and macular edema [81].

Treatment for macular edema typically involves medications such as corticosteroids to reduce inflammation and laser treatment to reduce the swelling. Most patients with a macular edema are treated with intravitreal injections, wherefore multiple visits at the hospital are mandatory. The number of visits depends on the severity of the disease. Macular edemas due to diabetes and retinal vein occlusion cannot only be treated by medications, but also by the patients' health behavior.

Inclusion and exclusion criteria

In this study, patients with macular edema due to diabetes or retinal vein occlusion were included anytime after the diagnosis.

Further inclusion criteria:

- Adult patient (at least 18 years)
- Willingness to participate in the study (signed informed consent)
- Good German language skills (able to complete questionnaires)

Exclusion criteria:

- Macular edema due to a different reason (not diabetes or retinal vein occlusion)
- Hearing difficulty (verbal communication impossible)

- Cognitive impairment
- Part of piloting sample

2.3 Questionnaires

In this study, the following questionnaires were used:

Patient Activation Measure (PAM® survey) [41]

The PAM® survey assesses patients' confidence, skills and knowledge to manage their own chronic illness. Items are scored on a Likert-type scale with four response categories ranging from "Disagree" to "Agree strongly". Every item corresponds to a specific level of patient activation: *level 1* items 1-3; *level 2* items 4-8; *level 3* items 9-11 and *level 4* items 12-13. The answers to items are summarized and transformed to a scale from 0 – 100 [41]. Sufficient reliability of the German version of the PAM® survey was shown through internal consistency with a Cronbach's α of 0.84. Indications of validity were given by factorial structure and a trait-trait correlation of r=0.43 between the score of the PAM® score and general self-efficacy [45].

The PAM® survey is licensed by Insignia Health, a company based in Portland, Oregon, USA,(`www.insigniahealth.com`). To be able to use it in this study, a research license was obtained. One condition to use the questionnaire without a fee for academic purpose is to forward them study relevant data in anonymized form (for more details see 'Data handling', section 2.5). Insignia Health states various other criteria as well, some of them are: the PAM® survey is not to be used in a commercial way, a sample size of at least 75 patients is required, and the plan to publish the results. By applying for a PAM® survey research license, Insignia Health got the rights to use our data for product improvement and validation efforts.

Questionnaire for Perceived Social Support (F-Sozu) [83]

To measure self-perceived social support, the F-Sozu with 22 items was used. They are rated on a Likert-type scale with five response categories ranging from "Disagree" to "Agree strongly". It is possible to build one general score ('social support') but also to create three subscale scores (emotional support, practical assistance and social integration). Sufficient reliability of the 22 item version was found in earlier studies (Cronbach's

$\alpha = 0.91$ for the general score, Cronbach's $\alpha = 0.72 - 0.87$ for the subscale scores) [83, 84, 85]. Validity for the F-Sozu is given through the factorial structure. In a factor analysis one factor 'social support' was found. Further validity is given through correlations with external criteria, such as self-confidence and life satisfaction. The subscale scores of the 22 item version differ in the size of the correlation with these external criteria, providing more information about social support than the general score [83].

This questionnaire is licensed by Hogrefe (`https://www.hogrefe.com/at/`). Authorization was provided for the inclusion of the F-Sozu. Since this a non-funded research study, a modest charge per completed questionnaire was stipulated.

Trait Well-Being Inventory Mood Level Scale [86]

This scale assesses an individual's general mood and quality of life through 13 items using a Likert-type scale with 6 response options ranging from "Agree not at all" to "Agree exactly." It consists of two subscale scores: general mood measured with six items and quality of life with seven items. The mean scores for each of the two scales were calculated, with internal consistency reliability coefficients of 0.83 and 0.87, respectively. The scale has been found to have factorial validity, with a single factor identified for the general mood scale and a further model containing the factor "habitual subjective well-being" constructed using the two subscale scores. The questionnaire also demonstrated construct validity through strong positive correlations with a life satisfaction scale [86].

Goals Associated with Health Information Seeking (GAINS) [87]

This is a questionnaire about one's motivation to seek information about health topics. It consists of 16 items which are rated on a Likert-type scale with five response categories ranging from "Totally disagree" to "Fully agree". The items are summed up to form 4 subscale scores: 'Understanding', 'Reassurance', 'Action planning' and 'Hope'. All items are summed up to gain a score for the general information seeking requirement [87]. Internal consistency of this overall scale was found to be acceptable with a Cronbach's α of 0.68 [88]. Those values were 0.80 for 'Understanding', 0.85 for 'Action Planning', 0.90 for 'Hope', and 0.85 for 'Reassurance', showing very good internal consistency. Validity was given through the factorial structure.

General Self-Efficacy Scale [89]

This questionnaire evaluates the individual's confidence in their ability to manage difficult situations independently. It's 10 items are scored on a Likert-type scale with four responses, beginning with "Disagree" and ending with "Agree". The overall score is between 1 and 4. Research done on German samples gave a Cronbach's α of 0.80 - 0.90. The questionnaire was found to be unidimensional, meaning that answers on the items can be explained by one underlying factor. The validity of the scale is further demonstrated by correlations with other variables, such as negative correlations with depression, anxiety, and burnout [89].

The Trait Well-Being Inventory Mood Level Scale [86], Goals Associated with Health Information Seeking (GAINS) [87] and General Self-Efficacy Scale [89] are part of the "Open Test Archive" (`http://www.testarchiv.eu/`) [90]. The Open Test Archive is a German repository providing a collection of questionnaires across various fields in psychology and related disciplines. This questionnaires are freely available for download and use in research. The Open Test Archive operates under a Creative Commons license that allows test developers to share their copyrighted questionnaires, including norm tables, scoring aids, and other related documents. The mean score over item answers were built for all questionnaires used in this study, except for the PAM® survey.

2.3.1 Additional measures

Current self-perceived health status was assessed by the question "How is your health in general?" with 5 answer options from "Very poor" to "Very good" [91]. Represented in numbers, lower values imply better health. Furthermore, demographic data like sex, age, and education were assessed. The net income was asked for, divided into five categories based on the income quintiles of elderly individuals in Austria reported in the year 2017 by 'Statistik Austria' (`www.statistik.at`) [92].

Information about relevant ICD-10 diagnosis for characterizing study participants were gathered from the hospital information system "openMedocs". These are: atherosclerosis, cataract senilis, diabetes including type I or II, unspecified diabetes mellitus with ophthalmic complications, essential hypertension, glaucoma, iridocyclitis, retinal vein occlusion and retinopathy including diabetic retinopathy.

2.4 Data collection

Overall, there were four different data collectors: three psychology students VW, AK, JG and MH. A standard interview procedure was established to ensure the objectivity and comparability of each survey. Before collecting data, interviewers were properly trained by MH. The interview procedure was practiced and study materials were familiarized. Possible barriers and difficulties while interviewing were discussed and appropriate behavior rehearsed. A pseudonym was created for each participant via a web-based pseudonymization tool called 'iPSN' [93]. This tool was developed at the Institute for Medical Informatics, Statistics and Documentation and ensures safe data handling. A survey on LimeSurvey [94] was prepared before the start of the study, containing the items of the questionnaires and questions about demographic data. For pseudonymization and data collection a tablet was used. Participans had the choice of being interviewed or to fill out questionnaires by themselves. When the questions were read out loud, the interviewer checked the answers on a tablet on LimeSurvey. If the patients wanted to fill out the questionnaires on their own, they were able to choose between paper and pencil or using the tablet. For the pseudonymization tool and LimeSurvey, a stable internet connection was mandatory. If it didn't work, paper and pencil were used for data assessment. If an interview was done with the paper and pencil procedure, the data were entered into LimeSurvey by the assessing interviewer afterwards. Before data entry, a training with each data collector took place to ensure quality of data entries.

At the Department of Ophthalmology of the Medical University of Graz, an outpatient clinic for patients with a macular edema due to diabetes is held two times per week and for patients with a macular edema due to retinal vein occlusion one time per week. These patients were approached about participating in the study. A day of data collection started with the interviewers checking the lists of incoming patients. Diagnoses of the patients were investigated and if the diagnosis fitted to the study criteria, it was checked if the patients had already done the interview. All eligible patients were communicated to the medical doctors in the outpatient clinic. The medical doctors explained the study purpose and gained informed consent. If a patient was willing to participate in the study, the interviewer searched for a room to conduct the survey. Before COVID-19, a tiny room was used. After COVID-19 appeared, larger alternatives had to be searched for. There were several options: a free examination room, the lecture hall, the laser room or a room next to "IVOM" treatment. The data assessments were usually done while patients waited in-between medical routine eye-examinations or after they had finished examinations. Because of these examinations, patients were not able to read the questionnaires by themselves. Thus, they were read out loud to the patients by the interviewers. The answer scales for each questionnaire were printed out in big font, those were set up with safety distance between the participant and the interviewer. A pseudonym was created in iPSN and copied into LimeSurvey, which was used to store the answers of the participants. Instructions how to answer every questionnaire and their scales were explained. Items were read out loud in an appropriate speed and volume matching the participants condition. They could either answer vocally or point on the printed out scale. If participants answered in sentences or stories, they were politely inquired to sum that up in one of the answer scale options. After completion of the survey, the interviewers lead the patient back to the waiting room or wherever required.

Another matter arose during data collection: after the occurrence of COVID-19, on most days, there was only one or two interviews possible due to a decreased number of patients in the outpatient clinic. Therefore,"IVOM patients" were included in the study by the end of October 2020. They come from the same patient population as the other participants. When they visit the clinic, they only get an intravitreal injection and are not examined by the doctors in the outpatient clinic. An examination is only done after a certain number of injections.

2.5 Data handling

Only authorized persons had access to the pseudonymization translation list. LimeSurvey was hosted on a server of the Medical University of Graz. Data was stored there as well. No names were entered in the online system. Data from the hospital information system are kept in an pseudonymized form on a server at the Institute for Medical Informatics, Statistics and Documentation. Informed consents of participants are safely locked away. Participants were informed that data concerning the PAM® survey and study relevant data will be forwarded to Insignia Health in anonymized form. This includes questionnaire based data collected during the study and demographics of the participants. Age was forwarded as a number, no birth dates were given. A random number was generated as ID for each participant. As mentioned before, forwarding these data is one condition to use the questionnaire without a large fee for academic purpose.

2.6 Statistical analysis

2.6.1 Sample size

The main objective of this study is to analyze the PAM® survey with an appropriate IRT model. In general, sample size depends on the chosen IRT model: a larger sample size is necessary with growing model complexity. Further, the more response options the questionnaire provides, the more participants are required, as more item parameters must be estimated. A minimum sample size of 500 is recommended for a two-parameter model [95]. The purpose of the analysis is influencing the required sample size as well: The aim of this study is to obtain precise and stable estimates on item characteristics and person scores. A sample size of at least 500 patients is recommended, requiring a bigger sample size if more items are analyzed [96].

There is one suggestion to calculate the power of a Rasch model via differential item functioning (DIF), assuming that there is a criterion dividing the participants into at least two groups [97]. DIF describes how an item works in different subgroups and if it measures the same abilities for members of those subgroups. Simulations were done which showed that with decreasing number of items and increasing number of participant's, power rises and a DIF can be detected more easily. For example if the questionnaire consists of 20 items and there are two groups with the same mean and the same DIF for two items, with a sample size of 50 a power of 43% would be achieved, whereas with a sample size of 100 a power of 79% would be reached. No simulations with a higher number of participants were done, which would be more adequate for the underlying study, but it is clearly visible that with a larger sample size the power to detect the same DIF increases.

Since the PAM® survey consists of 13 items with 4 response options and item characteristics and person scores have to be estimated, the aim was to include at least 500 patients in this study, to reach stable model estimations [95]. Given previous findings indicating DIF across various subgroups, such as sex and age, the plan was to analyze several subgroups. Consequently, a sample size of 700 participants was deemed necessary to ensure adequate statistical power for detecting potential DIF in the German version of the PAM® survey. Notably, the determination of this sample size was not based on direct estimation or calculation, but rather deduced from the outcomes of a simulation study [97]. Every year, approximately 1200 patients suffer from a newly diagnosed macular edema caused by diabetes or retinal vein occlusion and attend the outpatient clinic in

Graz. The aim to include 700 patients in one year seemed possible at the time the study was planned.

2.6.2 Descriptive analysis

Categorical data are presented as absolute and relative frequencies, while continuous data are presented as means and standard deviations or medians and interquartile ranges, as appropriate. According to the algorithm developed by Insignia Health, the answer category "Not applicable" was transformed into missing values, and the raw scores of the PAM® survey were summed up and transferred to a scale of 0-100 (`https://www.insigniahealth.com/products/pam`).

As a first step of the psychometric analyses category answer frequencies for every item were investigated. Floor and ceiling effects were defined as >15% of responses in the lowest or highest answer category, respectively [98].

2.6.3 Psychometrics

Since it is not possible to directly observe certain attributes (e.g. patient activation), questionnaires are used [38]. They assess the attribute in an indirect way. This is also called latent construct, trait or variable. Several models have been formulated to connect observable phenomena like answers on an item to the theoretical attribute of interest or latent trait, like patient activation. The goal of these models is to ensure that the attribute in question is measured and not something else, as well as that it is measured precisely. The focus is the objective measurement of latent constructs that cannot be directly observed. In general there are two approaches to this issue: the classical test theory (CTT) and the item response theory (IRT) [99].

2.6.3.1 Classical test theory

The traditional method to predict an individual's location on a latent trait based on an observed total score of a questionnaire is CTT [100]. It assumes that every observed score from a questionnaire consists of a true and an error part. The true score describes the attribute of interest, whereas the error score depicts unsystematic (e.g. random) error. CTT assumes that random errors are uncorrelated with the true score: there is no systematic relationship between the true score of a person and if the person gets a higher or lower test score than their true ability on the latent trait. Within CTT, the level of the latent trait and the observed score are predicted by the true score. If there were no errors in measurement, the true score of a person would be gained by solely completing the questionnaire. However, the true score can never be assessed, only the observed score. The location on the latent trait of an individual is determined by comparing the observed test score to the reference group scores. Therefore, a representative sample is needed to establish the so called norms [38].

In CTT, the test itself is the unit of analysis. Reliability and validity of a questionnaire are investigated by its items. Reliability is regarded as the consistency of a questionnaire. It is considered reliable if consistent results over time and across various administrations are achieved. The more items a questionnaire has, the more reliable it is. More reliable means less measurement error. To measure internal consistency of a questionnaire, Cronbach's α is used. It is a measure of the internal consistency of a questionnaire and is calculated by the mean correlations between all pairs of items. Values above 0.7 indicate acceptable reliability [101]. The so called retest-reliability assesses the association between the same questionnaire filled out at two different time points by the same individual. Validity describes if a questionnaire measures what it is intended to measure. In general, there are three different types: content, criterion and construct validity [38]. Content validity describes if the questionnaire really or sufficiently accurately measures the intended construct and cannot be tested by any statistical analysis. Criterion validity describes the correlation of questionnaire performance with one or more criteria with which the test should correlate due to its measurement claim [38]. Construct validity describes if the questionnaire measures the trait or ability that it is supposed to measure [38]. It includes three different types: factorial, convergent, and divergent validity. Factorial validity describes if questionnaire data can be explained by one underlying factor, the desired trait. To examine factorial validity, a confirmatory factor analysis can be used, to identify the underlying structure of a questionnaire. Convergent validity describes the correlation of the new questionnaire with an existing one, which aims to measure the same trait. Large correlations are expected. Divergent validity describes the correlation of the new questionnaire with an existing one, who aims to measure a similar trait. Small to medium correlations are expected [38]. For example, an association is expected between a questionnaire of patient activation and self-efficacy, since both constructs measure confidence of ones' own behavior.

In CTT, items themselves can be investigated regarding two properties: difficulty and discrimination. Item difficulty describes how hard it is to answer the item correctly. In CTT, difficulty is defined by the item mean score. Items with lower difficulty are solved more easily. For hard items it is more challenging to find the correct answer. Since items measuring personality attributes like patient activation have no right or wrong answer, endorsement behavior of individuals is investigated. Within this context, the difficulty parameters can be interpreted similar: it describes how hard it is to endorse an item (e.g. select "Totally Agree"). If "Agree" is represented by the highest value, a higher mean score implies an easier item. Items with lower difficulty are endorsed more easily, whereas hard items are challenging to endorse. Item discrimination describes an item's measurement quality, specifically how well it distinguishes between individuals of high and low ability. Items with high discrimination are more effective in distinguishing between individuals with different levels of ability. Usually it is expressed as the correlation between a given answer on a particular item and the overall questionnaire score [38].

CTT assumes unidimensionalty of the measured underlying construct and local stochastic independency. However, it fails to provide a method for verification within it's own theory. Moreover, traditional validation methods used in CTT such as exploratory factor analysis and Cronbach's α coefficient may overlook measurement issues such as item fairness.

2.6.3.2 Item response theory

IRT generally consists of a collection of statistical models that explain the connection between the questionnaire items and person ability 'θ' (e.g., patient activation), which is regarded as a latent trait [102]. IRT assumes that the responses to each item are influenced by both the latent trait and random error. The models estimate the latent trait and the item parameters like item difficulty and discrimination. All given answers on the items provide an indirect observation of the person ability θ. It ranges from $-\infty$ to ∞, but is usually displayed in logits between -3 to 3 [38]. IRT models describe the probability of selecting a certain response category of an item with a specific person ability. Higher categories (e.g., "Agree strongly") are more likely chosen with higher person ability 'θ'. The person ability is estimated on a continuous scale, rather than creating particular reference groups, as in CTT [102].

In IRT, the items are the unit of analysis. The answers to each item are assumed to be a noisy measure of the latent trait. In contrast to CTT, more items do not imply higher reliability. In IRT, the measurement error differs depending on the individual's underlying ability. Usually, questionnaires assess a trait with more error in the area of high and low abilities than in the area of medium ability. In IRT, the measurement error can be not only random, but also systematic: answer behavior can be influenced by a different factor than the attribute of interest [103].

Depending on the IRT model, different parameters are estimated for each item. The item difficulty parameter 'b' can be interpreted as in CTT, and indicates how easily people agree on the item. Item difficulty is estimated by the model as a latent variable. The discrimination parameter 'a' has the same meaning as in CTT as well and describes the capability of an item to differentiate between individuals with different levels of the ability being measured. Item discrimination is estimated by the model as a latent variable as well and is affected by both item difficulty and the spread of item category thresholds [102]. These item thresholds are estimated by the model as well. They represent the points on the latent trait scale, if prior items were endorsed, where one answer category becomes more likely to be picked than the preceding one. In other words, they represent the value where an individual's latent trait level is considered to be 'high enough' to be classified into a particular category. Thresholds are defined as the level of ability where the probability of an correct or incorrect response is the same (50%). In other words, item threshold is the value of the ability parameter that separates individuals who are likely to endorse the item from those who are likely to disagree it. The threshold between "Agree" and "Agree totally" might be estimated to be at a latent trait level of 1.8. This means that individuals with latent trait levels above 1.8 would choose "Agree totally", while those below 1.8 would choose "Agree". To put it another way, the threshold represents a boundary between the categories of "Agree" and "Agree totally" and individuals with latent trait levels above the threshold are more likely to be classified into the "Agree totally" category [38].

In IRT, empirical reliability can be estimated. It describes how well the model predicts the probability of endorsing an item, given a certain level of ability. This is usually done by comparing the estimated item parameters obtained from the model to the underlying study data. Like Cronbach's α the values vary between 0 to 1, with 1 meaning that the model perfectly predicts the observed data [102].

The performance of the questionnaire can be evaluated and any problems with item difficulty or scale reliability identified, once an IRT model is fitted. A clear picture of the performance of each item is given as well as the overall functioning of the questionnaire for measuring the attribute of interest in the study population [38]. In the next part, different IRT models will be described. In general, there are models for dichotomous and polytomous responses.

2.6.3.2.1 One-parameter logistic model

The simplest model is the one-parameter logistic model (1PL) [104] or Rasch model for dichotomous data. It can be used in case of items with binary responses e.g. "Yes" or "No". The 1PL assumes that the response to each item is influenced by the latent trait, random error and item difficulty. The ability of the participants and the difficulty of the test items are estimated. Discrimination is assumed to be equal across all items. A single difficulty parameter for each item on the scale is estimated by an individuals probability of endorsing an item:

$$P(X_{ij} = 1 \mid \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \tag{2.1}$$

where j stands for the individual, i for the item, θ is the latent trait of the individual, and b is the difficulty parameter of the item [38]. Item thresholds vary from item to item. This value can be used to evaluate how "easy" an item is, depending on the position of the threshold in dependence to the person ability distribution. An easy item would have a low threshold, it can be answered correctly by individuals with a wide range of ability levels [38]. For example, if the threshold for a particular item would be -0.4, individuals with a greater ability than -0.4 would endorse the item. Individuals with a latent trait of -0.4 would have a 50% chance of endorsing it.

2.6.3.2.2 Two-parameter logistic model

A further model to analyze dichotomous data is the two-parameter logistic model (2PL) [102] or Birnbaum model, an extension of the 1PL. The 2PL assumes that the response to each item is influenced by the latent trait, random error, item difficulty and, additionally discrimination. The ability of the participants, the difficulty and discrimination of the test items are estimated using the equation:

$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i) = \frac{e^{(a_i(\theta_j - b_i))}}{1 + e^{(a_i(\theta_j - b_i))}} \tag{2.2}$$

where a is the discrimination parameter of the item [38]. The discrimination parameter indicates the extent to which the probabilities of endorsing an item changes as a function of the underlying trait [38].

2.6.3.2.3 Three-parameter logistic model

An extension of the 2PL is the three-parameter logistic model (3PL) for dichotomous data [102]. It assumes that the response to each item is influenced by the latent trait, random error, item difficulty, discrimination and, additionally guessing behavior. The ability of the participants, the difficulty, discrimination and guessing parameter of the test items are estimated by the equation:

$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i)\frac{e^{(-a_i(\theta - b_i))}}{1 + e(a_i(a - b_i))} \tag{2.3}$$

where c is the guessing parameter [38]. Due to this guessing parameter, each individual always has a probability higher than zero to solve an item. For example, in a multiple choice test one could simply guess the correct answer by chance, even if the underlying ability level is lower than the item difficulty [38].

The models described are only examples for IRT models for analyzing dichotoumos data. There exist more models, e.g. the four-parameter logistic model [38], which is an extension of the 3PL and additionally assumes that the response to each item is influenced by the probability to never respond correctly to an item. This way an individual with a high ability can falsely answer to an item as he had low ability, without his overall ability estimate degrading drastically. As well as for dichotomous data, there are several different IRT models to describe polytomous data, some of which will be described in the next part.

2.6.3.2.4 Partial credit model

In case of items with polytomous responses e.g. "Agree", "Undecided", "Disagree", data can be analyzed by a partial credit model (PCM) [105]. It is designed for polytomous items, but different answer categories throughout the questionnaire are possible. The ability of the participants and the difficulty of the questionnaire items are estimated. The PCM assumes that the response to each item is influenced by the latent trait, random error and item difficulty. Discrimination is constrained to be equal across all items. The probability of answering in a specific item category P_{i_g} is given by

$$P_{i_g} = \frac{e^{\sum_{g=0}^{l}(\theta - b_{i_g})}}{\sum_{h=0}^{m} e^{\sum_{g=0}^{h}(\theta - b_{i_g})}} \tag{2.4}$$

where i stands for the item, θ is the latent trait of the individual, b is the difficulty parameter and g is the response category modeled with $h = 0, 1, . . . , m$ [103].

This model enables partial credit to be given for partially correct answers. This is useful in situations where there are multiple correct answers or where items can be partially correct. For example, a multiple-choice item could have more than one correct answer. In the PCM, a participant who chooses all correct answers of an item would receive full credit, while a participant who chooses only some of the correct answers would receive partial credit based on the distance between the correct answers and the selected ones. Item thresholds are estimated as the difference in latent trait scores between two adjacent response categories, as in dichotomous models. Thresholds represent the values of the latent trait where the probability of choosing both answer categories is the same. For example, if the response categories are "Agree", "Undecided", "Disagree", the thresholds would be estimated as the latent trait scores that correspond to the decision between "Agree" or "Undecided", and "Undecided" or "Disagree". Item thresholds steps are estimated separately for all items and can vary between them. Since item answer categories only have to be polytomous but not ordered, thresholds themselves don't have to be ordered [103]. For example, for an item the decision between "Undecided" or "Disagree" could go along with a lower person ability score than the decision between "Agree" or "Undecided". Once the thresholds have been estimated, they can be used to calculate the amount of partial credit that is given for a partially correct response [102].

2.6.3.2.5 Generalized partial credit model

Another model is the generalized partial credit model (GPCM) [105], which is an extension of the partial credit model. The ability of the participants, the difficulty and, additionally the discrimination of the test items are estimated. It assumes that the response to each item is influenced by the latent trait, random error, item difficulty and item discrimination and that item responses consist of two or more ordered categories and is estimated by using the equation:

$$P_{i_g} = \frac{e^{a_i[\phi_g(\theta - b_i) + \sum_{g=1}^{l} \tau_g]}}{\sum_{h=1}^{m} e^{a_i[\phi_h(\theta - b_i) + \sum_{g=1}^{h} \tau_g]}} \tag{2.5}$$

where ϕ_i is the standard scoring function, equaling the count of answer options, for the specific category g being modeled; and τ_{i_g} is the threshold parameter representing the category boundary locations relative to the item location parameter [103].

This means, item discrimination can vary across items. Some items may be more efficient in differentiating individuals with high and low abilities than others. Item thresholds are equally defined as in the partial credit model.

2.6.3.2.6 Rating scale model

The rating scale model (RSM) [106] is a further model of the IRT family to analyze questionnaire data. The RSM assumes that the response to each item is influenced by the latent trait, random error and the same difficulty over all items. Discrimination is assumed to be equal across all items. The ability of the participants and the difficulty of the test items are estimated using the equation [103]:

$$P_{i_g} = \frac{e^{\sum_{g=0}^{l}[\theta - (b_i + \tau_g)]}}{\sum_{h=0}^{m} e^{\sum_{g=0}^{h}[\theta - (b_i + \tau_g)]}} \tag{2.6}$$

The first or last category has to be the lowest (e.g. "Disagree") or highest (e.g. "Agree"). Answer options have to be the same for every item. Item thresholds are estimated as the latent trait scores that correspond to each response category. One threshold between each pair of adjacent categories is estimated. In the RSM, item thresholds steps are equal across all items since during the model estimation they are fixed to be equal for all items. Therefore, the distances between adjacent step difficulties are the same across all items. In the RSM, participants either receive full or no credit for an item answer. For example, if the threshold for "Undecided" is 0.8 and the threshold for "Agree" is 1.4, an individual who selects "Agree" and has a latent trait score of 1.5 would receive full credit, while a respondent who selects "Undecided" and has a latent trait score of 0.7 would receive no credit.

2.6.3.2.7 Graded response model

A further model is the graded response model (GRM) [107]. It is made for polytomous ordered responses. Different answer categories throughout the questionnaire are possible. The ability of the participants, the difficulty and discrimination of the test items are estimated. The graded response model assumes that the response to each item is influenced by the latent trait, random error, item difficulty and item discrimination using the equation [103]:

$$P_{i_g} = \frac{e^{a_i(\theta - b_{i_g})}}{1 + e^{a_i(\theta - b_{i_g})}} \tag{2.7}$$

The discrimination parameter is the same within an item for every category, but can differ between items. Item difficulty can vary for each response category. In this model thresholds are estimated differently than in the models before: they represent the value of the latent trait where at least a specific category is chosen. One step has to be made or rather one level reached before the next one can be attained. Threshold parameters are not the probability of choosing between two adjacent categories. They represent the probability of an individual choosing a higher category than a given one [103]. For example, consider that response categories are "Agree", "Undecided", "Disagree". 2 thresholds between the adjacent categories are estimated. The first threshold describes the probability of choosing "Undecided" over "Disagree" with a specific ability. Thresholds are always ordered, e.g. the value in the latent trait of choosing "Disagree" is always lower than choosing "Undecided".

2.6.3.3 Item response theory analysis

As this study includes measures of personality traits without right and wrong answers, IRT will be explained in the context of endorsing items and not answering correctly from this point on. There are several properties that can be examined, once the IRT model is estimated.

2.6.3.3.1 Assumptions

In general, for estimating IRT models, there are four different assumptions. The assumption of (1) unidimensionality can be investigated before a model is estimated, (2) monotonicity, (3) local independence and (4) item invariance should be checked in the final model.

Unidimensionality describes that the model assumes that one latent trait is measured and that this trait is responsible for the selected answers of an individual. This is checked by a confirmatory factor analysis [38]. For example, regarding the PAM®, one underlying latent factor, patient activation, is assumed to explain the variance in the data. Factor loadings are the correlation between every item and the estimated latent factor. Communalities are gained by squaring the factor loading. They describe the percentage of variance of the item explained by the latent factor. For example, a factor loading of an item of 0.8 would correspond to a communality of 0.64. This means that 64% variance of the item is explained by the latent factor [38]. Root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMSR) were used to evaluate absolute fit of the confirmatory factor analysis. RMSEA describes the difference between the hypothesized model and a perfect model. A RMSEA < 0.05 indicates a good fit and < 0.08 suggests a reasonable fit [108]. SRMSR describes the average standardized residual covariance of the fitted model. A SRMSR ≤ 0.08 indicates a good fit [108]. To assess incremental fit of the model, the Tucker-Lewis index (TLI) and comparative fit index (CFI) were used. They both are incremental fit indices and describe the difference between the hypothesized model and the baseline model. The CFI and TLI values also range from 0 to 1, with larger values indicating a better fit [108]. For both indices, a value above 0.95 indicate a good model fit [108].

The IRT assumption of *monotonicity* is satisfied, if with a higher ability level, the probability of endorsing an item response also increases. This is investigated by item characteristic curves. Monotonicity can be further evaluated by the so called 'item-trace-line' plots, where the expected and observed probability values over the whole ability range is displayed for every category, respectively [38].

The IRT assumption of *local independence* describes that associations between items are only due to the latent trait, individual items in a questionnaire are independent from each other. In- and outfit statistics provide relevant information to conclude local independence [102].

The IRT assumption of *item invariance* describes that the measurement of the latent trait is independent of the sample characteristics within a population. Invariance is satisfied if item parameters are constant across different characteristics and not influenced ('biased') by an external factor. This is investigated with differential item functioning [109].

2.6.3.3.2 Fit of IRT models

RMSEA, SRMSR, TLI and CFI can be used to assess model fit in IRT. Additionally, to investigate relative model fit, the sample adjusted Bayesian information criterion (SABIC) and the Akaike information criterion corrected (AICc) can be used. They can take any positive value, and can be used to compare the relative goodness of fit of different models. Smaller values indicate a better model fit [38]. Infit and outfit statistics are calculated to determine how well the items fit the model that described the underlying construct. If the item and respondent are close on the level of the latent trait, infit statistics are more sensitive. Outfit statistics are more sensitive when items and person ability are distant to each other. The range of 0.5 - 1.5 is effective for measuring [110]. If data are too predictable, suggesting that a different variable is influencing the response patterns, values are $\leq$ -2.

2.6.3.3.3 Model selection

Regarding different IRT models, the question arises which one is fit to analyze the data. In general, the goal is to find the simplest model to describe the data in a good way [103]. Less parameters mean an simpler model, e.g. the PCM is simpler than the GPCM, because the discrimination parameter is not estimated. The RSM, GPM and GPCM models are similar to another. The GPCM is the most general model. Reducing the estimation of the discrimination by fixing it to a specific value leads to the GPM. Additionally constraining the threshold parameters to be equal results in the RSM [103]. Since the basic structure is the same and only parameter specifications differ between them, they are nested within each other. Furthermore, these three models belong to the Rasch model family.

However, there is no definitive process or strategy for selecting the best model for a given dataset. On the one hand statistical considerations can be done, on the other hand content-wise deliberations must be considered. Different models can be compared by fit indices and there are also some statistical tests available. There are two common tests for IRT model comparison: the likelihood ratio (LR) test for nested models and the

Vuong-test for nested and additionally non-nested models [111].

The LR test is a statistical test used to compare the fit of two models estimated with the same data set [111]. The log-likelihoods of the two models are compared and it is examined whether the improvement in fit of one model over the other is statistically significant. For example, a LR test can be performed to compare the RSM, PCM and GPCM.

Further, estimated models can be compared by a Vuong test [111]. This is a statistical test that compares the fit of nested or non-nested models estimated with the same data set. In contrast to the LR test, the Vuong test is non-parametric and non-nested models can be compared. For example, a RSM and GRM can be compared with this test. The Vuong test is generally considered more robust than the LR test. As in the LR test, the log-likelihoods of the models are compared and examined. In the first step it is analyzed if the two models are distinguishable. In the second step it is determined, which model fits best.

Theoretical considerations should be taken into account as well, when deciding on a final model to describe a specific data set. Different IRT models have different implications in their interpretation. Specifically, the inclusion of the discrimination parameter into the model is an essential question. Moreover, the kind of estimation of thresholds are influencing the results and therefore the interpretation.

2.6.3.3.4 Item characteristic curve

Item characteristic curves (ICC) are used to investigate answer patterns, an example see figure 2.1. It is a plot of the probability of choosing a specific answer of a questionnaire item as a function of an individual's ability. Every category is displayed as a curve over the ability range of θ in logits. There are several different shapes that ICCs can take, depending on the difficulty of the test item and the amount of discrimination it provides [38]. The difficulty influences whether the curve is more to the left or the right. Further, for example, a high discriminating item will have a steep ICC, meaning that the probability of endorsing the item increases rapidly with higher ability. A less discriminating test item will have a more flat ICC, indicating that the probability of endorsing an item increases more slowly with higher ability.

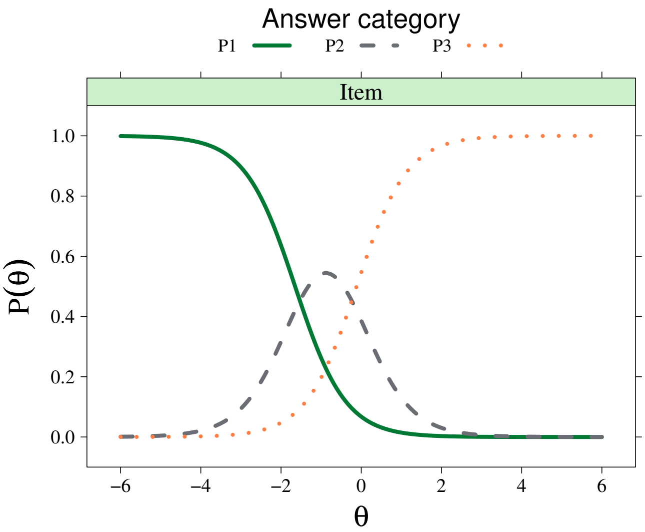


Figure 2.1: Item with three answer categories. Probability (0 to 1) for choosing a specific answer category over the whole ability range displayed in logits (-6 to 6). P1 = "Disagree"; P2 = "Undecided", P3 = "Agree".

2.6.3.3.5 Wright Map

Another graphical helpful presentation used in IRT, is the Wright Map [112]. This graph depicts the probability of selecting between item answer categories and patients' ability expressed in logits, see figure 2.2. It consists of two parts: on the top the ability distribution of the sample is displayed as a histogram, on the bottom item thresholds are visualized. This part of the Wright Map consists of a series of parallel lines, with each line representing a different test item. The position of a line on the map reflects the difficulty of the test item. On these lines, thresholds for choosing between answer categories are displayed as well. Their position relative to the person's ability provides information about the value of the latent ability needed to endorse an item. It can be used used to evaluate the overall quality of a questionnaire and to identify areas where the test may be lacking discrimination [112].

2.6.3.3.6 Item and test information

In addition, another graphic display in IRT, the item information curves (IIC) should be examined [102], an example is displayed in figure 2.3. They represent the amount

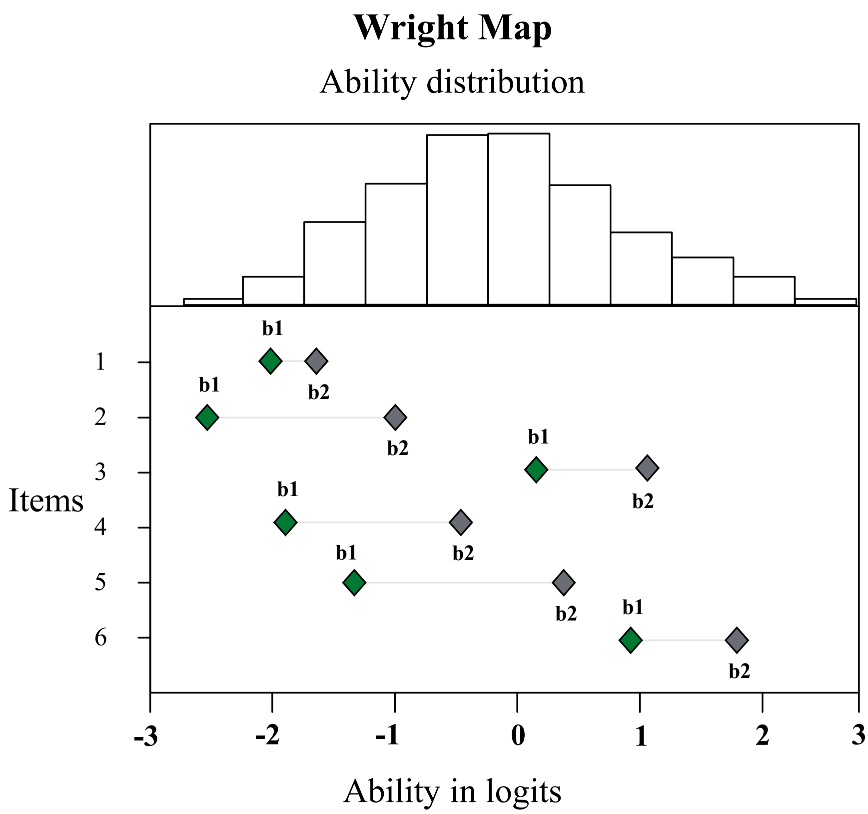


Figure 2.2: Top: ability distribution in logits. Bottom: distributions of thresholds per item. *b1* and *b2* stand for first and second threshold, respectively. For example, the first threshold *b1* could describe the probability of choosing between “Disagree” or ”Undecided”, the second threshold *b2* between “Undecided” or ”Agree”.

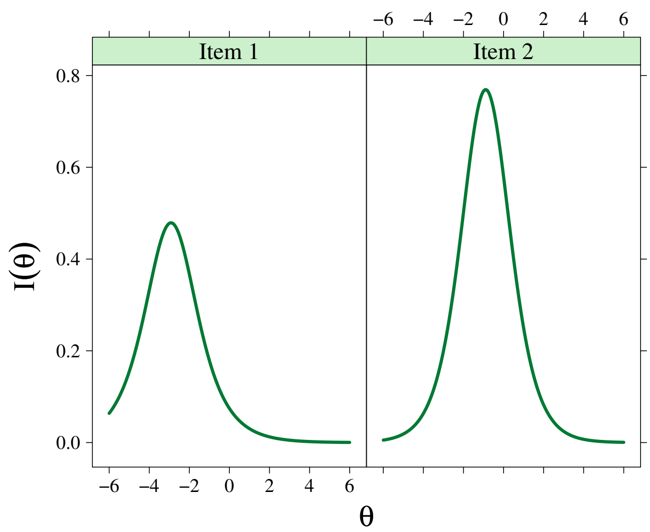


Figure 2.3: Item information for two items. Each curve displays the amount of information given in the corresponding area of the latent trait in logits

of information provided by a questionnaire item about the person's ability. The IIC is a curve that plots the item information, the amount of information provided by the item, as a function of an individual's latent trait. The shape of the IIC depends on the difficulty of the test item and the amount of discrimination it provides. The higher the curve, the more information the item provides in the particular ability range. A low discriminating questionnaire item will have a flat IIC, meaning that it provides relatively little information about an individual's latent trait or rather in differentiating between individuals with low and high ability. A highly discriminating test item will have a steep IIC, indicating that it provides a greater amount of information [102].

All item information's can be merged into the total test information. This can also be displayed graphically in one curve only, illustrating how much of the desired ability is measured over the whole ability range [102], see figure 2.4. The shape of the test information curve depends on the characteristics of the questionnaire items and the distribution of the latent trait in the test population. A questionnaire that contains easy, difficult, and highly discriminating items and has good coverage of the latent trait will generally have a high test information curve, indicating that it provides a large amount of information about an individual's latent trait. A questionnaire that contains mostly easy or difficult and weakly discriminating items, or that has poor coverage of the latent trait, will have

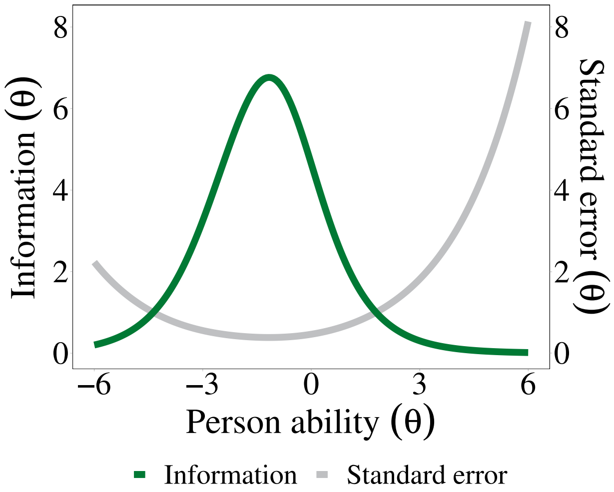


Figure 2.4: Test information (green line) and standard error (gray line) over the whole ability range in logits.

a lower test information curve, indicating that it provides less information [102]. By item characteristic curves, Wright Map, item information, test information and standard error, the quality of test items can be evaluated and items identified that are either too easy or too difficult for a given population [103].

2.6.3.3.7 Reliability

In IRT, reliability can be seen as the consistency of an individual's classification on the latent trait [102]. From the test information, the standard error is used as an indicator of test reliability. The standard error is a measure of the precision of an estimate of an individual's latent ability. It is typically represented as a curve and interpretation can be made inversely to the test information curve. A questionnaire with good coverage of the latent trait will generally have a low standard error curve. Moreover, models estimating fewer parameters will also have a lower standard error than models estimating more parameters [103]. Therefore, latent trait estimates and their corresponding standard errors are used to estimate empirical reliability.

The separation index is a further measure of reliability. It describes the amount of "true" standard deviation of the average measurement error [112]. The separation index itself is an indicator in how many theoretical groups individuals can be classified by a questionnaire. It is also possible to calculate the number of empirical distinguishable groups by the separation index from the estimated model [112]. The results indicate how many different groups (e.g. low and high activation) the study population can be put in by the questionnaire data. The separation index can be transformed into the number of distinguishable groups by the formula: (4*separation index + 1)/3.

2.6.3.3.8 Differential item functioning

Differential item functioning can be investigated to see if answer behavior of an item of individuals on the same ability level in different subgroups is influenced by certain characteristics [109]. Examples for such subgroups are gender and educational status. In the case of uniform DIF, the item is easier to endorse for one group compared to the other. Even if two individuals have the same ability level, in one group they are more likely to agree to the statement than in the other group. The difficulty parameter b varies in those groups [109]. For example, an item could be easier to endorse for women vs. men. Non-uniform DIF involves a difference in both the discrimination a and difficulty b parameters. Depending on the level of ability, the item is either easier to endorse for the one or the other group. For example, an item could be easier to endorse for younger individuals in the low ability range, but in the high ability range it is easier endorsed by older individuals. In case of non-uniform DIF two cases can appear: balanced or unbalanced DIF. It is balanced, if the overall item difficulty is identical in the two groups. In other words, the sum of the item difficulty over the whole ability range is the same for the groups, but the item is easier to endorse in different ability levels for one of the groups [109]. For example, one item is easier to endorse for younger individuals in the low ability range, but in the high ability range it is easier endorsed by older individuals. In the low ability range the item is more likely to be endorsed by younger ones with a specific probability. In the high ability range it is easier endorsed by older ones with the same specific probability. Unbalanced DIF exists, if the overall item difficulty differs between groups [109]. The sum of the item difficulty over the whole ability range differs in the groups, the item is always easier to endorse for one group compared to the others. Across the ability range, the difference in item difficulty varies between the groups. For example, in the low ability range the difference in endorsing an item could be small, in the middle range large and in the higher range small again.

The likelihood-ratio χ^2 test can be performed to detect uniform and non-uniform DIF. McFadden's pseudo-R^2 can be used as a measure of DIF magnitude, if DIF is found [109].

For example, a value of pseudo-R^2=0.044 means that 4.4% of the overall item variance are explained by the factor causing DIF. Moreover, non-compensatory differential item functioning (NCDIF) as a measure of DIF magnitude can be used [113]. The square root of NCDIF provides the overall impact on the score, e.g., a NCDIF of 0.16 means that an average absolute difference of 0.4 of a point of an item is influenced by the factor causing DIF.

In this study, the aim was a comprehensive analysis by using both CTT and IRT approaches. By emphasizing the strengths of both methodologies, the goal was to achieve an in-depth understanding of the psychometric properties of the German version of the PAM® survey.

2.6.4 Validity

Trait-trait correlations between PAM® scores and other questionnaires should provide evidence for divergent validity [38]. According to literature, a moderate negative association was expected between patient activation and self-rated health status [70, 114, 115]. Moderate positive associations were expected between patient activation and self-efficacy [45], quality of life [73, 74, 116], general mood [74, 54, 117], and goals associated with health information seeking [52]. A weak correlation was expected with perceived social support [117, 118]. Due to the ordinal nature of the questionnaires used, Spearman correlations were performed. The strength of the correlations coefficients were judged as small if those were >0.10, as medium if >0.30 and as large if >0.50 [119].

2.6.5 Missing data

As in many studies, the problem of missing data occurs. Generally, this means loss of information and therefore bias could be introduced, leading to incorrect statistical conclusions [120]. Moreover, some analysis methods may only be performed with non-missing data. IRT models can be estimated with missing data, however their fit indices cannot. There are different solutions to this problem: analyzing only what is there, excluding cases or imputing data [120]. For data imputation different methods exist. For example, missing values can be replaced with the mean, median, or mode of the non-missing values in the same item. Furthermore, data can be imputed by a regression model, which predicts missing values based on other item answers in the dataset [120]. In IRT exists the possibility of imputating missing data by model results [121]. Firstly, the ability level on the latent trait for every participant is estimated. Secondly, these estimations are further used to estimate the missing values. Since they can vary, this step should be done several times [121].

2.6.6 R packages

R offers several different packages for IRT analysis. The mirt package [121] was chosen because of it's many implemented functions. It offers the possibility to estimate several dichotomous and polytomous models, including (multidimensional) item response theory, exploratory and confirmatory factor analysis. The most common IRT models are included, e.g. rating scale and generalized partial credit model.

The basis for estimating the unknown model parameters is the marginal maximum likelihood (MML) estimation. It is determined by maximizing the marginal likelihood of the observed data. It allows estimation of the parameters of IRT models using full information maximum likelihood methods. This means that it takes into account all available data, even in the presence of missing data. However, the marginal likelihood is often unfamiliar since the distribution of the unobserved latent trait is unknown in the beginning. The mirt package offers various MML estimation methods, including expectation-maximization (EM), Metropolis-Hastings Robbins-Monro, and quasi-Monte Carlo EM. These methods efficiently and precisely estimate complex models. For example, the EM algorithm seeks to find the MMLE of the likelihood in several iterations by applying two steps: expectation and maximization. The expectation step defines the expected value of the unobserved latent trait, estimating a log-likelihood function. The maximization step estimates the parameters, maximizing this expected log-likelihood function. These parameter estimates are used in the next expectation step to resolve the distribution of the latent trait.

The lordif package [122] offers estimation of DIF using iterative hybrid ordinal logistic regression and Monte Carlo simulations. The package provides a flexible and robust framework for detecting various types of DIF, including uniform and non-uniform DIF. Moreover, the package can handle dichotomous and polytomous items. The possible influencing characteristic on answer behavior can be dichotomous (e.g. sex: male or female) or have more categories (e.g. comorbidity: hypertension, diabetes or both). The package performs an ordinal logistic regression DIF analysis using IRT θ estimates as the conditioning variable. The package provides default criteria and thresholds, which can be adjusted. This includes Mantel-Haenszel chi-square test, logistic regression Wald test, logistic regression likelihood ratio test, SIBTEST, standardized mean difference and effect size. Items exhibiting DIF are adjusted within specific subgroups to compute trait assessments that consider DIF effects. These steps are repeated until the same items are identified with DIF in successive iterations. The package also provides several graphics to evaluate the impact of DIF, including item characteristic curves by group and item response functions.

Statistical analysis was done using R studio version 4.1.1 [123]. Tidyverse was used to clean and prepare data [124]. The package psych was used for descriptive analysis [125], crosstable [126] and flextable [127] were used to create tables. Ggplot2 was mainly used to plot data [128]. The packages mirt [121] and lordif [122] were used for the IRT analysis.

Chapter III

Results

3.1 Study participants

707 patients were screened for study eligibility, see figure 3.1. Most of the patients were willing to participate in the study, only N=39 (6%) declined study participation. N=81 were not suitable because they did not speak German well enough, were mentally not fit, had hearing difficulties or were too frail. N=587 gave informed consent, 6 (1%) withdrew their consent before the interview was started. 5 (1%) patients did the interview twice. 12 (2%) interviews had to be terminated, because patients turned out to be not suitable, e.g. because of hearing difficulties. 564 interviews were completed, however, in 10 cases at least two questionnaires were not completed. N=554 finished all questionnaires except GAINS. Some participants (N=51) did not search for health relevant information themselves. Therefore, the GAINS questionnaires could not be filled out, resulting in N=503 fully completed interviews. The main analysis focuses on the N=554 patients.

3.1.1 Descriptive statistics

Demographic variables describing the patient sample are in table 3.1. Slightly more men (N=321, 58%) took part in the study. Median age was 69.0 (IQR: 62.0 – 76.0). 317 (57%) patients had a diabetes-related macular edema, 224 (40%) had retinal vein occlusion, and 13 (2%) had both diagnoses. N=358 (65%) patients suffered from diabetes and N=425 (77%) from hypertension. Regarding the extracted study relevant ICD-10 diagnoses from the health information system, patients had 4.0 (median, IQR: 3.0 - 5.0) documented diseases. The frequency of different diagnoses can be found in table 3.1.

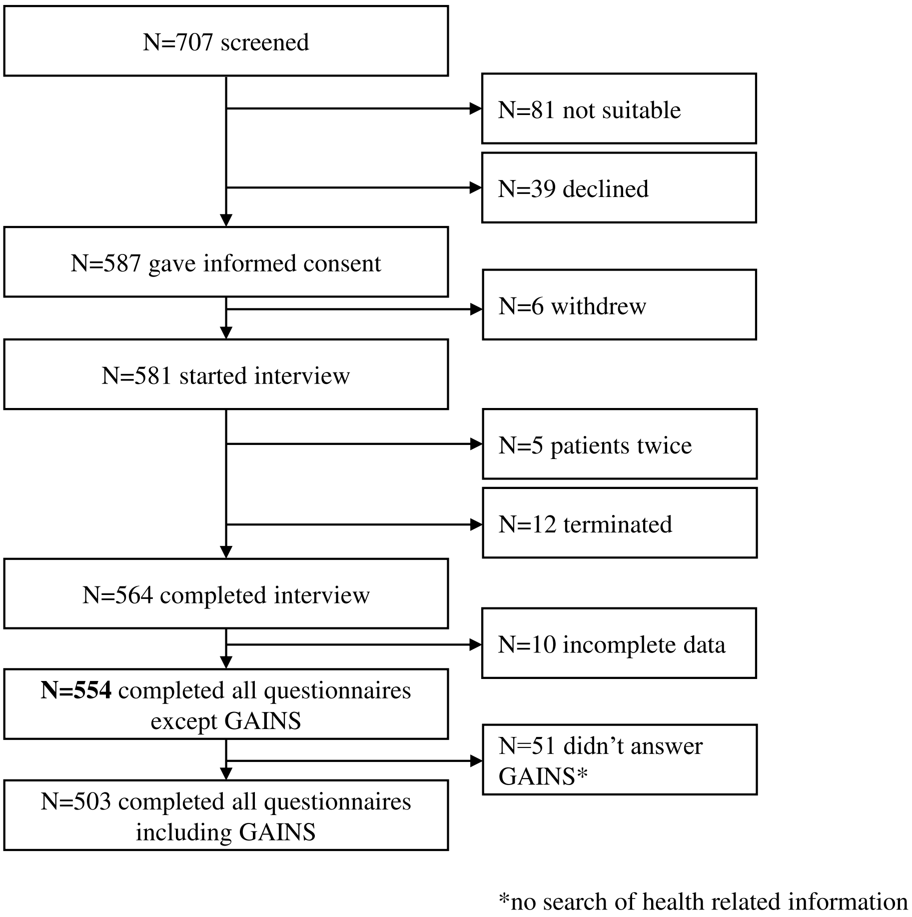


Figure 3.1: Flow-chart of study participants.

Table 3.1: Demographics

N = 554	Total
Age in years	69.0 (62.0 -76.0)
Sex	
Male	321 (58%)
Female	233 (42%)
Education	
Basic education	411 (74%)
High school	68 (12%)
Higher education	75 (14%)
Working status	
Working	66 (12%)
Other [1]	41 (7%)
Retired	447 (81%)
Monthly net income	
<800€	72 (13%)
<1125€	91 (17%)
<1500€	112 (21%)
≤ 1950€	82 (15%)
>1950€	184 (34%)
Missing	13 (2%)
BMI[2]	25.8 (24.7 – 30.0)
Type of macular edema	
Diabetic	317 (57%)
Retinal vein occlusion	224 (40%)
Both	13 (2%)
Comorbidity	
Diabetes	129 (23%)
Hypertension	196 (35%)
Both	229 (41%)
ICD-10 Diagnoses	
Retinopathy (H35 & H36)	445 (80%)
Cataract senilis (H25)	368 (66%)
Vessel occlusion (H34)	231 (42%)
Glaucoma (H40)	63 (11%)
Atherosclerosis (I70)	41 (7%)
(E14.3)[3]	9 (2%)
Iridocyclitis (H20)	5 (1%)

Data are presented as N (%) or median (25th – 75th percentiles)
1: Category 'Other' includes jobless, studying and homemaker
2: BMI = Body mass index.
3: Unspecified diabetes mellitus with ophthalmic complications

Table 3.2: PAM® score for relevant patient characteristics

N = 554	**Total**	**PAM® score**
Sex		
Male	321 (58%)	72.9 ± 13.4
Female	233 (42%)	75.7 ± 13.9
Education		
Basic education	411 (74%)	74.9 ± 13.8
High school	68 (12%)	71.1 ± 14.0
Higher education	75 (14%)	72.2 ± 12.3
Monthly net income		
<800€	72 (13%)	75.2 ± 13.1
<1125€	91 (17%)	71.7 ± 14.5
<1500€	112 (21%)	76.3 ± 13.7
≤1950€	82 (15%)	74.0 ± 12.7
>1950€	184 (34%)	73.5 ± 13.5
Missing	13 (2%)	74.4 ± 18.1
Type of macular edema		
Diabetic	317 (57%)	73.4± 13.7
Retinal vein occlusion	224 (40%)	75.1± 13.7
Both	13 (2%)	72.0± 11.3
Comorbidity		
Diabetes	129 (23%)	74.6 ± 12.6
Hypertension	196 (35%)	75.1 ± 13.5
Both	229 (41%)	72.9 ± 14.4
Health status self-rated		
Very good	61 (11%)	83.3 ± 11.1
Good	236 (43%)	76.6 ± 11.7
Moderate	222 (40%)	70.6 ± 13.8
Bad	29 (5%)	60.8 ± 14.1
Very bad	6 (1%)	72.7 ±15.4

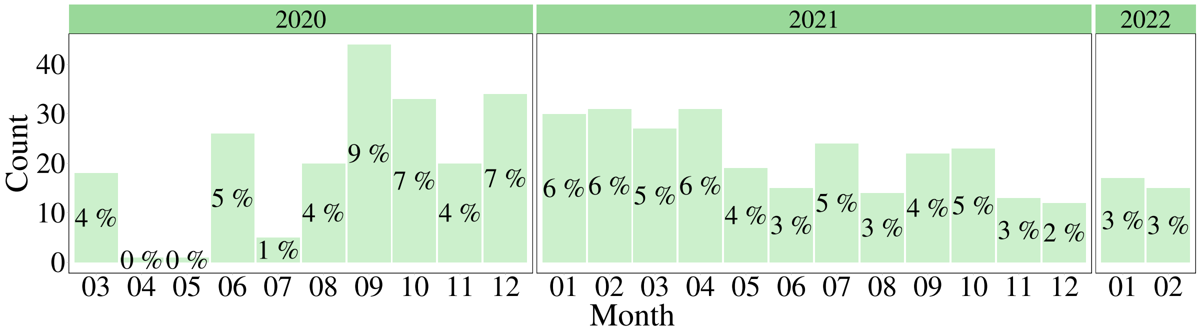


Figure 3.2: Conducted interviews over study period per month and year. In April and May 2020 no interviews were done.

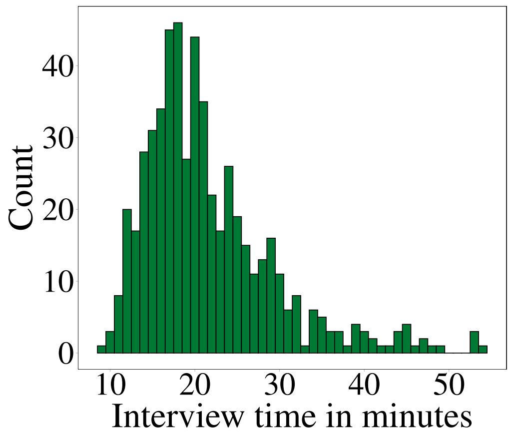


Figure 3.3: Duration of interviews in minutes

3.1.2 Interviews

None of the participants chose to fill out the questionnaires by themselves. Data assessment solely was made through interviewing patients. 63% (N=349) of those were conducted in the diabetes outpatient clinics. Each month between 5 and 44 interviews were made, see figure 3.2. One interview lasted approximately 20 (IQR: 16.4 – 24.8) minutes, see figure 3.3. The PAM® survey took about four (IQR: 2.7 – 4.7) minutes. Overall, there were four different interviewers. Interviewer MH did 185 (33%), interviewer VW 159 (29%), interviewer AK 114 (21%) and interviewer JG 96 (17%) of all interviews. All four interviewers were women.

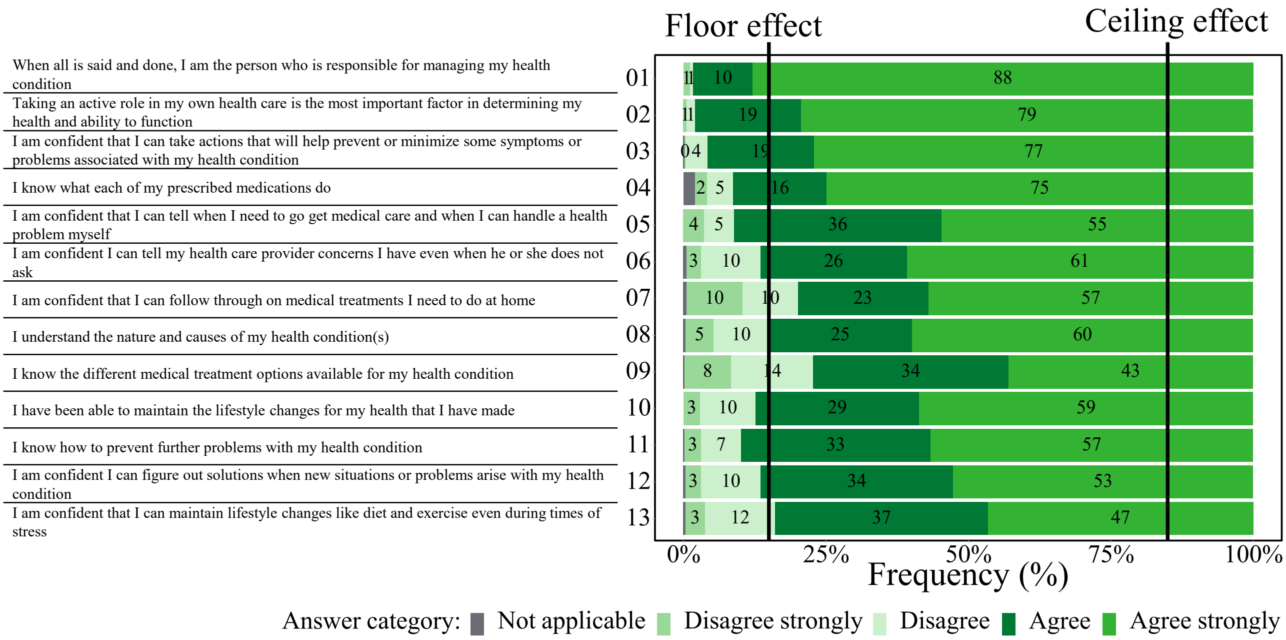


Figure 3.4: Category answer frequencies for the PAM® survey

3.2 Description of Patient Activation Measure®

Due to the interview setting where an answer to every item was inquired, initially there were no missing answers for the PAM® survey. The answer "Not applicable" was transformed into missing values. It was selected most frequently for item 4, occurring in 2% of all cases. "Agree strongly" was the most frequent response category across all items, which was selected between 43% (item 9) and 88% (item 1) of patients (see figure 3.4). "Agree" was selected by 10% - 38% of participants over all items. The least chosen response category appeared to be "Disagree strongly", for most items not exceeding 3% of the answers. Only for item 7 and 9 this category was chosen by more than 5% of the patients. The mean response scores representing CTT difficulty varied between 3.1 (SD 0.9) for item 9 to 3.9 (SD 0.5) for item 1, see table 3.5.

The study population showed a mean activation score of 74.08 (SD: 13.67), see 3.5. Their corresponding activation levels according to patients' overall scores were 19 (3%) level 1, 37 (7%) level 2, 91 (16%) level 3, and 407 (73%) level 4. The PAM® score was highest for patients who self-rated their health as 'Very good'. For patients with different types of macular edema, the PAM® score was similar, see table 3.2.

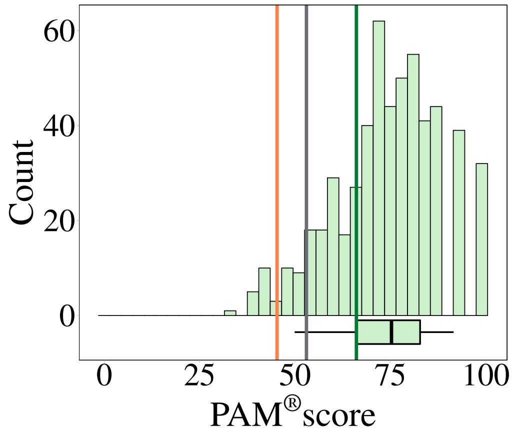


Figure 3.5: PAM® score and cut-offs for different activation levels

3.3 Psychometrics

3.3.1 Assumptions

Before the start of the IRT analysis the assumption of unidimensionalty was investigated. This criteria was met sufficiently with one factor extracted from the PAM® data in a confirmatory factor analysis, explaining 0.295 variance of the PAM® data, more details see table 3.3. Fit indices for the confirmatory factor analysis were acceptably with a RMSEA of 0.062 (90% CI: 0.052 - 0.072), a SMSR of 0.064, a TLI of 0.894 and a CFI of 0.912. The remaining three assumptions were evaluated in the final model and will be discussed later on, see section 3.3.4. In this section, the data imputation is described as well. After data imputation, the confirmatory factor analysis was redone, see table 3.3. Fit indices for the confirmatory factor analysis of the imputed data were similar with a RMSEA of 0.062 (90% CI: 0.052 - 0.072), a SMSR of 0.064, a TLI of 0.905 and a CFI of 0.900.

3.3.2 Model estimation

At first analysis was done with the initially four answer categories. Four different IRT models for polytomous items were estimated, see appendix table 5.1. However, the lowest two answer categories “Disagree strongly” and “Disagree” were chosen rarely, leading to estimation problems in the models. For item 3, only one participant has chosen "Disagree strongly". ICCs were investigated, revealing problems for several items. For some

Table 3.3: Confirmatory factor analysis

	Original data		**Imputed data**	
	Factor loading	**Communality**	**Factor loading**	**Communality**
Item 01	0.54	0.29	0.52	0.27
Item 02	0.54	0.29	0.53	0.28
Item 03	0.52	0.27	0.51	0.26
Item 04	0.54	0.29	0.54	0.29
Item 05	0.58	0.34	0.58	0.34
Item 06	0.39	0.15	0.39	0.15
Item 07	0.50	0.25	0.49	0.24
Item 08	0.48	0.23	0.49	0.24
Item 09	0.52	0.27	0.55	0.30
Item 10	0.55	0.30	0.55	0.30
Item 11	0.64	0.41	0.63	0.40
Item 12	0.68	0.46	0.68	0.47
Item 13	0.54	0.29	0.54	0.30

items, the two middle answer categories "Disagree" and "Agree" never had the highest probability of getting selected. An example is displayed in figure 3.6.

Therefore, the two categories “Disagree strongly” and “Disagree” were merged into one, named “Disagree strongly & Disagree”. The same four IRT models suitable for polytomous items were estimated. The model fit increased. Fit statistics and other model related information are in table 3.4. Since the PAM® survey consists of 13 items and three categories, 39 parameters were estimated for the RSM: one slope and two threshold parameters per item are estimated. The RSM exhibited ordered and equidistant thresholds due to model characteristics. The PCM was estimated with 27 parameters, the discrimination of items set to 1 and item thresholds estimated separately for each item. The GPCM was estimated with 39 parameters, including two threshold parameters and one discrimination parameter per item. The PCM and the GPCM showed disordered thresholds for item 4 and 7. The GRM was estimated with 39 parameters, with two threshold parameters and one discrimination parameter per item. The GRM had no disorderes thresholds due to the model specifications, but they were not equidistant. All thetas of the estimated polytomous IRT models showed a Pearson correlation of 0.97 - 0.98 with each other.

Since the category answer frequency analysis revealed ceiling effects and polytomous

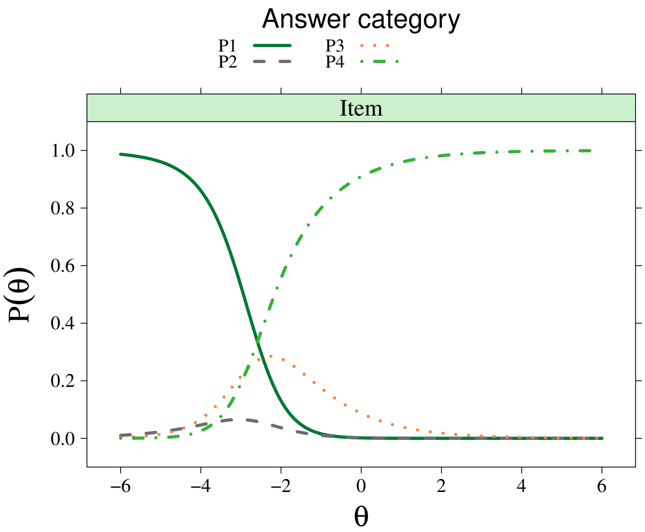


Figure 3.6: Item characteristic curve for item 1 estimated by a GPCM with initial 4 answer categories

IRT models revealed low probabilities for the middle categories to be chosen, two different dichotomous models were estimated as well. In the first one "Disagree strongly" and "Disagree" were merged into one category and also "Agree" with "Agree strongly". In the second one, merging of the answer categories was done by statistical considerations: "Disagree strongly", "Disagree" and "Agree" were merged into one category, leading to "Agree strongly" as the second answer category in the model. Fit indices of those model indicated a good model fit, see table 5.1. However, reliability was insufficient. In addition, more importantly, these dichotomous models are not meaningful enough from a theoretically point of view.

3.3.3 Model selection

The PAM® survey consists of 13 items and 4 ordered answer categories. Four different IRT models were estimated for this data set: RSM, PCM, GPCM and GRM.

An LR-test was used to compare the fit of the RSM, PCM and GPCM models. The results of the comparison indicated that the GPCM had a superior fit when compared to both the RSM ($p<0.001$) and PCM ($p<0.001$). For the sake of completeness, the PCM outperformed the RSM ($p<0.001$) in the pairwise comparison.

A pairwise comparison of all four different polytomous IRT models, RSM, PCM,

Table 3.4: PAM® survey IRT analysis: Fit indices of polytomous IRT models estimated with three answer categories (N = 531), except final model (N=554).

Parameter	**Rating scale model**	**Partial credit model**	**Generalized partial credit model**	**Graded response model**	**Generalized partial credit model final model**
Degrees of freedom	101	77	65	65	65
RMSEA (90% CI)	0.069 (0.060 - 0.077)	0.062 (0.053 - 0.072)	0.058 (0.048 - 0.068)	0.066 (0.056 - 0.076)	0.062 (0.052 - 0.072)
SRMSR	-	0.075	0.064	0.063	0.064
TLI	0.882	0.903	0.915	0.891	0.905
CFI	0.866	0.904	0.929	0.91	0.921
SABIC	11488.1	11439.7	11431.1	11399.2	11496.6
AICc	11504.3	11467.6	11469.6	11437.7	11458.1
Test information	26.0	26.0	22.55	23.12	22.60
Standard Error: Mean ± SD	0.43 ± 0.06	0.43 ± 0.06	0.50 ± 0.08	0.49 ± 0.07	0.50 ± 0.08
Empirical reliability	0.73	0.73	0.74	0.75	0.75
Marginal reliability	0.78	0.78	0.73	0.75	0.74
Correlation with PAM® score (95% CI)[1]	0.98 (0.98 – 0.98)	0.98 (0.98 – 0.98)	0.98 (0.97 - 0.98)	0.97 (0.96 – 0.97)	0.97 (0.97 – 0.98)

RMSEA=root mean square error of approximation; SRMSR= standardized root mean square residual TLI=Tucker-Lewis index; CFI=comparative fit index; SABIC=sample adjusted Bayesian information criterion; AICc = Akaike information criterion corrected;

1: Pearson correlation of estimated model theta with PAM® score (95% CI)

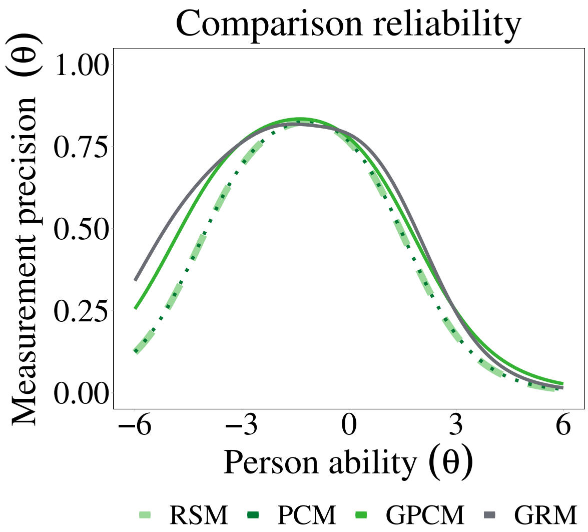


Figure 3.7: Empirical reliability for different estimated models

GPCM, and GRM, was conducted using a Vuong test. The first step revealed that all models are distinguishable from another. The GRM revealed to have the best fit compared to all three other models ($p<0.001$ against RSM and PCM; $p<0.002$ against GPCM). The GPCM showed the second best fit ($p<0.001$ against RSM; $p<0.002$ against PCM). The PCM revealed to have a better fit than the RSM ($p<0.001$).

The results of the LR and Vuong tests indicate that the discrimination parameter should be included in the model to describe the PAM® survey. This parameter is estimated in the GPCM and GRM model. A comparison of the empirical reliability estimated by the different models is visualized in figure 3.7. Theoretical considerations regarding the model characteristics of different estimated models and their implication were taken into account as well. For example, in the GRM a specific structure in the underlying data is assumed: the thresholds are always ordered. In the GPCM thresholds can be unordered. Therefore, ICCs differ between the models. Since the interest lies in the investigation of the nature of the items of the PAM® survey, ordered thresholds do not have to be assumed. While analyzing the PAM® data, unordered thresholds appeared in the PCM and GPCM. Since the study aim was to find a model, recognizing the underlying pattern in the data, the GPCM was chosen as the final model to analyze the PAM® data.

3.3.4 Final model: GPCM

In a first step, the GPCM was estimated with all available data. Fit indices can be seen in table 3.4. Since the answer category “Not applicable” of the PAM® survey was transformed into missing values [41], an incomplete data set was available for data analysis. Model selection was made with this data set. After the model was selected to evaluate PAM® data, imputations for missing values (N=23, 0.3% of all values) were performed. Firstly, the ability level on the latent trait for every participant was estimated. These estimations of the ability level on the latent trait were used to estimate the missing values. For the final value replacing the former “Not applicable” answer, the mean of 1000 estimations was taken. The final model estimated latent ability patient activation based on the answer patterns of the 13 items from 554 participants, as well as the difficulty and discrimination of each item. The fit indices RMSEA: 0.062 (90% CI: 0.052 – 0.072), SRMSR: 0.064, TLI: 0.905, CFI: 0.921, SABIC: 11496.6 and AICc: 11458.1 indicated a good model fit. A Spearman correlation of 0.97 ($p<0.001$) of the patient activation ability score calculated from the GPCM was found with the PAM® overall score.

Mean msq item infit was 1 (SD: 0.1), ranging from 0.9 to 1.1. Mean msq item outfit was 0.9 (SD: 0.04), ranging from 0.8 to 1. All items were within the range of 0.5 - 1.5, indicating accurate measurement of patient activation [110]. Mean z item infit was -1.0 (SD: 0.9), ranging from -2.6 to 0.4. Mean z item outfit was -1.7 (SD: 0.8), ranging from -3.3 to -0.5. Most items lie in the area between -1.9 to 1.9, where data have reasonable predictability [110]. Items 9 and 12 demonstrated bad in- and outfit. Items 5, 10, 11 and 13 revealed bad outfit.

Item difficulties ranged from -2.9 (item 1) to -0.5 (item 9), detailed information can be found in table 3.5. Item discrimination varied between 0.5 to 1.3, with a mean of 0.9 (SD: 0.2).

3.3.4.1 Item characteristic curves

Item characteristic curves are displayed in figure 3.8. In item 4 and 7, the middle category "Agree" never had the highest probability of getting chosen. Participants in the lower ability range had the highest probability of choosing "Disagree strongly/Disagree". In the middle to higher latent trait area, "Agree totally" had the higher probability to be selected.

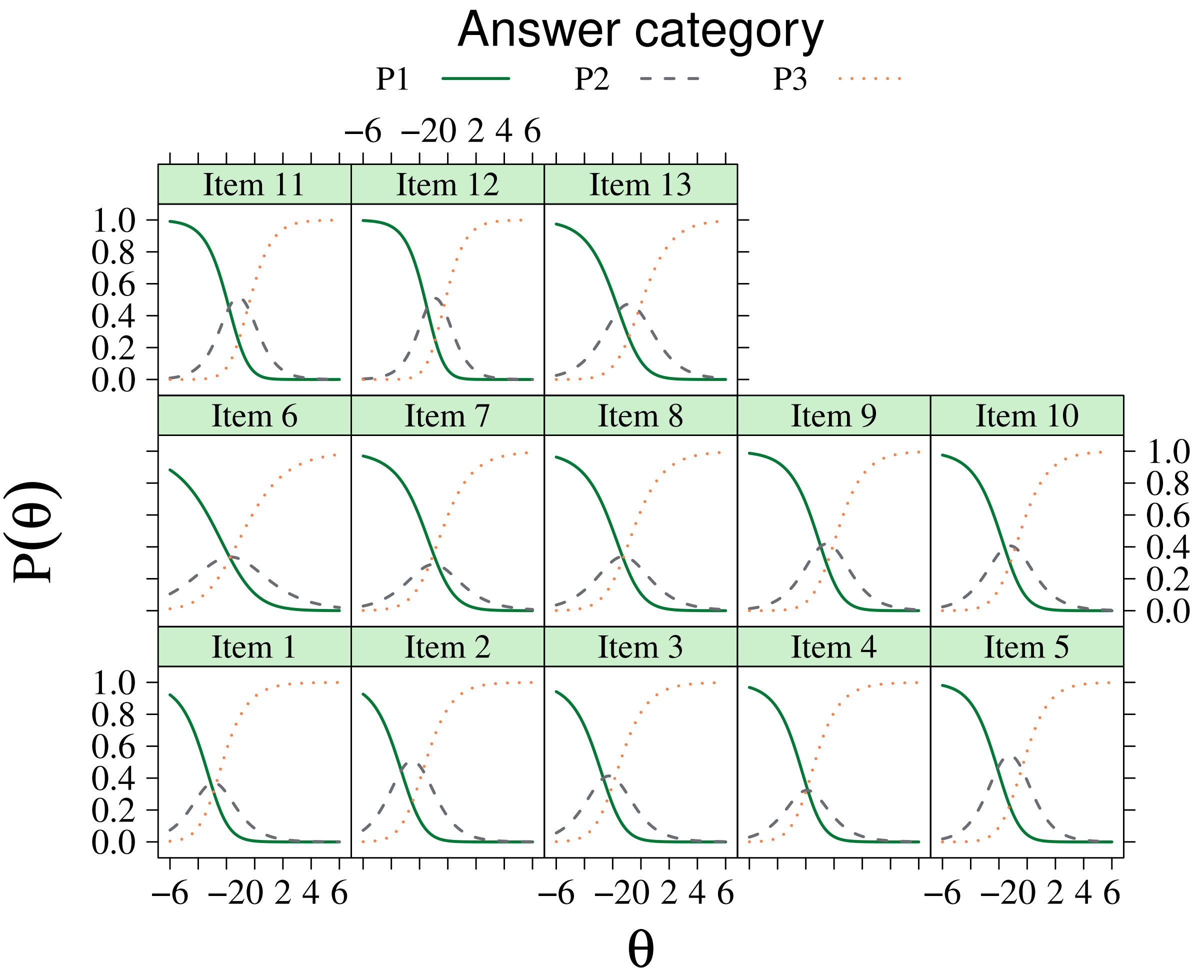


Figure 3.8: PAM® survey item characteristic curves estimated by a GPCM with three answer categories. Probability (0 to 1) for choosing a specific answer category over the whole ability range displayed in logits (-6 to 6). P1 = "Disagree strongly & Disagree"; P2 = "Agree", P3 = "Agree strongly".

Table 3.5: PAM® item statistics

PAM® items	Order by IRT diff. [1]	CTT difficulty	IRT difficulty	IRT discrimination	Infit mean squared	Infit z	Outfit mean squared	Outfit z
Item 01	1	3.9	−2.9	0.9	1.1	0.4	0.9	−0.5
Item 02	2	3.8	−2.5	1.0	1.0	−0.1	0.9	−0.7
Item 03	3	3.7	−2.3	0.8	1.0	0.0	0.9	−1.0
Item 04	4	3.7	−1.9	0.8	1.0	−0.2	0.9	−1.3
Item 05	7	3.4	−1.2	1.0	0.9	−1.4	0.9	−2.2
Item 06	5	3.5	−1.7	0.5	1.0	−0.5	1.0	−0.6
Item 07	10	3.3	−1.0	0.7	0.9	−1.4	0.9	−1.9
Item 08	6	3.4	−1.3	0.7	1.0	−1.0	0.9	−1.6
Item 09	13	3.1	−0.5	0.9	0.9	−2.4	0.9	−2.7
Item 10	8	3.4	−1.2	0.8	0.9	−1.3	0.9	−2.1
Item 11	9	3.4	−1.2	1.1	0.9	−1.5	0.8	−2.7
Item 12	11	3.4	−0.9	1.3	0.9	−2.6	0.8	−3.3
Item 13	12	3.3	−0.9	0.8	0.9	−1.6	0.9	−2.0

1: Item sequence by IRT difficulty, ordered from easiest to hardest.

3.3.4.2 Wright Map

Looking at the distribution of patient activation, participants showed abilities between -3 and 2 logits, see figure 3.9. The least difficult item was number 1. Several items demonstrated thresholds in the same ability range. Some areas of the latent trait are barely represented by the thresholds. They are not spread over the whole ability range. Thresholds started at an ability level of -3. No item exhibited a threshold for latent trait values higher than 0 logits. Disordered thresholds were found for Item 4 and 7, indicating a violation of the IRT assumption of monotonicity.

3.3.4.3 Item and test information

In figure 3.10 item information estimated by the GPCM for every item of the PAM® survey are displayed. Several items provide information in the same ability range, e.g., item 1 to 5. Little information is given by item 6. Item 12 contributes a lot in recognizing the patient activation ability of patients.

The total test information of the PAM® was estimated to be 22.55. Figure 3.11 displays the distribution of the test information and the standard error over the ability range. In the ability area of -3 to 1 logits, most information is given. For highly activated patients, little information is given by the PAM® survey items. The assessment of patients in the lower ability range is conflicted with less error than medium to highly activated

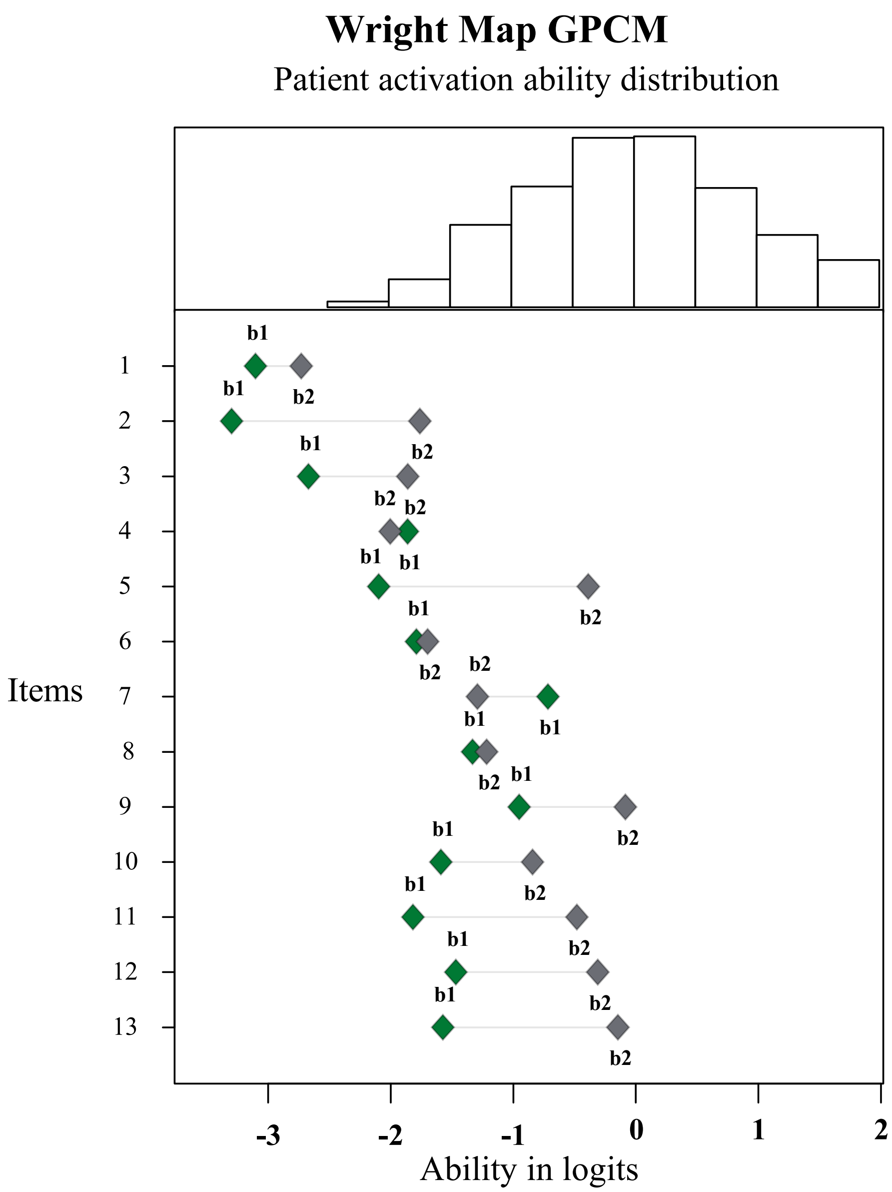


Figure 3.9: Wright map estimated by a GPCM with PAM® survey data. Top: distribution of the sample in activation scores in logits. Bottom: distributions of thresholds per item. The first threshold b1 describes the probability of choosing between “Disagree strongly & Disagree” or ”Agree”, the second threshold b2 between “Agree” or ”Agree strongly”.

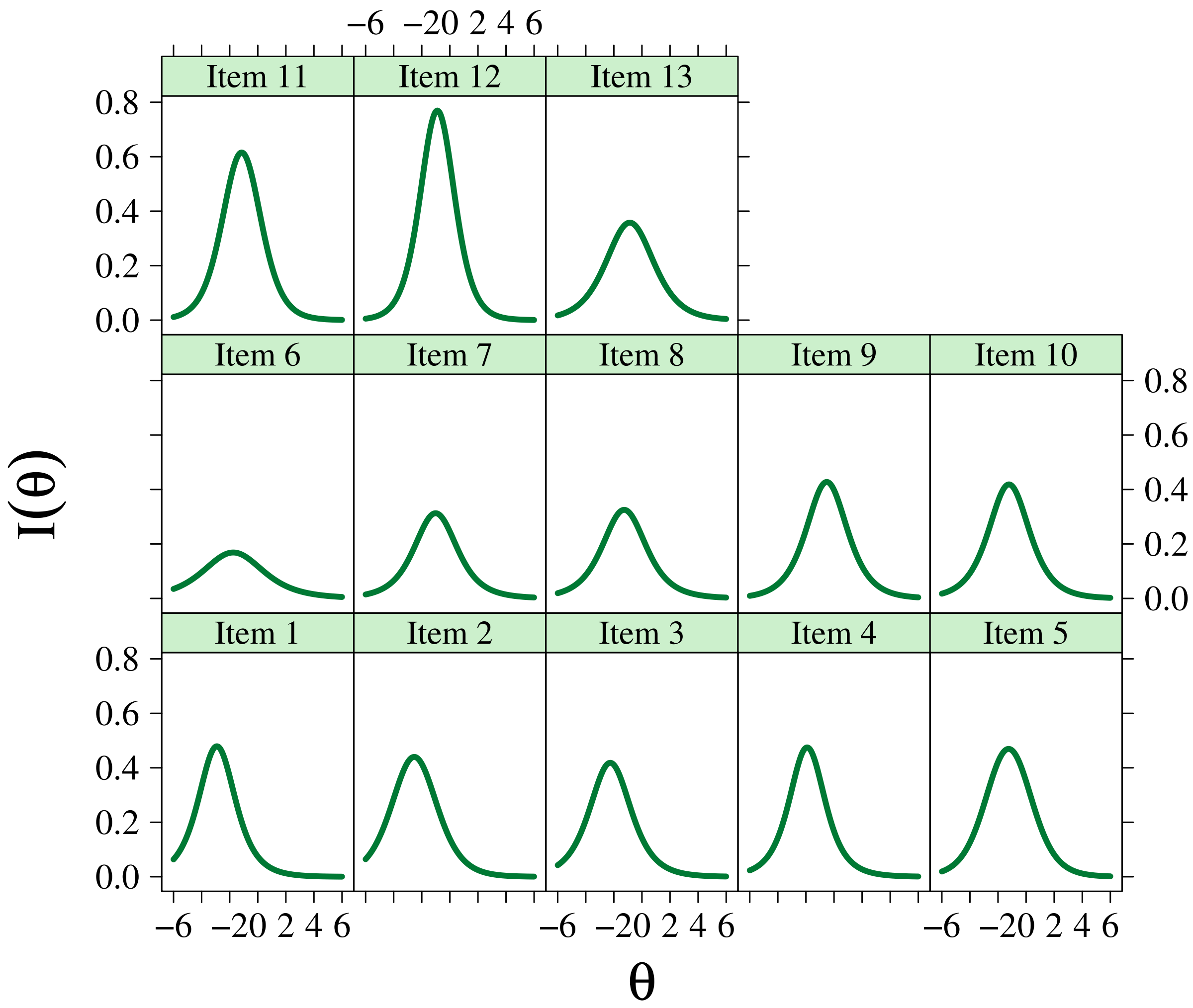


Figure 3.10: Item information for each PAM® survey item. Each curve displays the amount of information given in the corresponding area of the latent trait (patient activation) in logits

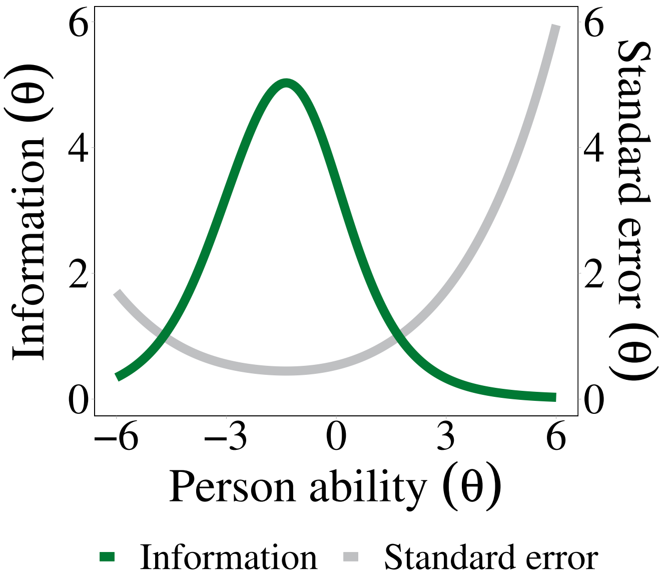


Figure 3.11: PAM® test information (green line) and standard error (grey line) over the whole ability range in logits.

patients.

3.3.4.4 Reliability

Using the model, the empirical reliability was calculated. It was estimated to be 0.75, indicating moderate reliability, see figure 3.12. It was also investigated using CTT methods. A Cronbach's α of 0.75 was estimated, indicating acceptable internal consistency (reference value $>$0.7 [101]).

The estimated person separation index was 1.39. According to that value, two different groups (e.g., low and high activation) are distinguishable by the questionnaire.

3.3.4.5 Differential item functioning

In this study, subgroups were determined by age groups (younger $<$70, older $\geq$ 70), sex (male vs. female), education (basic education, high school and higher education), health status (very good to good vs. moderate to bad), type of macular edema (diabetic vs. retinal vein occlusion), comorbidity (hypertension, diabetes or both) and assessing interviewer (number 1, 2, 3 or 4). DIF was estimated by ordinal logistic regressions, which allows detection of uniform and non-uniform DIF [109].

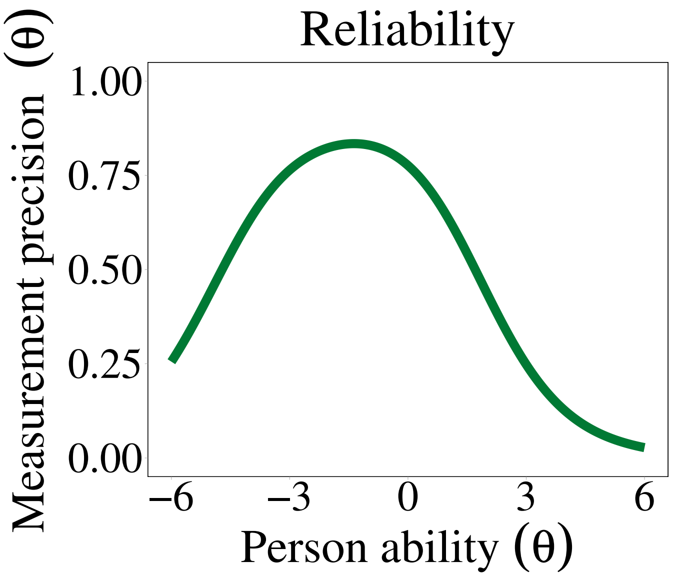


Figure 3.12: PAM® empirical reliability over the whole ability range in logits.

DIF regarding interviewers could not be estimated as initially intended. Due to the answer category patterns, a comparison between all interviewers at once was not possible. Therefore, the results of one interviewer against results from the other three interviewers taken together was compared. Uniform DIF was detected for interviewer VW compared to all the other interviewers for item 7 (p<.001 R^2=0.028, NCDIF=0.123, see appendix figure 5.1). This item was easier to endorse for patients if interviewer VW administrated it.

Non-uniform unbalanced DIF was found for the three education subgroups (basic education (reference), high school and higher education) for item 7 (p=0.018, R^2=0.007, NCDIF1: 0.066, NCDIF2: 0.103, see appendix figure 5.5). In the low ability range, the item was easier for higher educated patients. In the high ability range, item difficulty was similar for all educational groups. No DIF effects were found for sex. There was significant uniform DIF for age, if split in two groups (younger <70, older $\geq$ 70) for three items: DIF was found for item 10 (p=0.005, R^2=0.010, NCDIF=0.030, easier for older patients, see appendix figure 5.2), item 11 (p=0.004, R^2=0.011, NCDIF=0.022, easier for younger patients, see appendix figure 5.3) and item 13 (p<0.001, R^2=0.013, NCDIF=0.044, easier for older patients, see appendix figure 5.4).

Uniform DIF was found for health status (two groups: very good to good vs. moderate to bad): Item 10 (p<0.001, R^2=0.012, NCDIF=0.046, see appendix figure 5.12), was easier to endorse for patients with good health. It was significant for type of macular edema, if split in two groups (DME: diabetic macular edema, RVO: macular edema due to retinal vein occlusion, in case both types were recorded, patients were put in the group according to witch outpatient clinic they visited at study inclusion) for three items: DIF was found for item 5 (uniform, p=0.004, R^2=0.008, NCDIF=0.018, easier for RVO, see appendix figure 5.6), item 8 (uniform, p<0.001, R^2=0.026, NCDIF=0.095, easier for DME, see appendix figure 5.7) and item 10 (non-uniform, p<0.001, R^2=0.019, NCDIF=0.041, low ability range easier for DME, middle to high ability range easier for RVO, see appendix figure 5.8). Regarding the three subgroups of comorbidity (hypertension (reference), diabetes or both) DIF was found for item 2 (uniform, p=0.006, R^2=0.018, NCDIF1=0.01, NCDIF2=0.012, easier for hypertensive patients, see appendix figure 5.9), item 8 (uniform, p<0.001, R^2=0.034, NCDIF1=0.103, NCDIF2=0.154 more difficult for hypertensive patients, see appendix figure 5.10) and item 10 (non-uniform unbalanced, p=0.001, R^2=0.013, NCDIF1=0.043, NCDIF2=0.036, see appendix figure 5.11). In the low ability range, item 10 was more difficult for patients with hypertension and both diseases, in the high ability range, the item was easier for patients with hypertension only.

3.4 The other questionnaires

Patients showed high self-efficacy with a mean of 3.3 (IQR: 3.0 - 3.6), as can be seen in table 3.6. They also rated their quality of life high as well as their general mood. Moreover, patients reported their social support and their seeking for health information to be high.

PAM® score demonstrated a Spearman correlation of -0.34 (95% CI: -0.41 to -0.26 p<.001) with health status, 0.54 (95% CI: 0.47 to 0.59, p<.001) with self-efficacy, 0.51 (95% CI: 0.45 to 0.57, p<.001) with quality of life, and 0.34 (95% CI: 0.26 to 0.41, p<.001) with general mood. Further, Spearman correlations between the PAM® score with perceived social support 0.40 (95% CI: 0.33 - 0.47 , p<.001), practical assistance 0.34 (95% CI: 0.24 to 0.32, p<.001), emotional support 0.38 (95% CI: 0.38 to 0.45, p<.001), social integration 0.33 (95% CI: 0.38 to 0.45, p<.001) and goals associated with health information seeking 0.27 (95% CI: 0.19 to 0.35, p<.001) were found.

	N	Mean	SD	Median	Interquartil range	Possible range
Health status self-rated	554	2.4	0.8	2.0	2.0 - 3.0	1 - 5
Self-efficacy	554	3.3	0.5	3.3	3.0 - 3.6	1 - 4
Quality of life	554	5.2	0.7	5.3	4.9 - 5.7	1 - 6
General mood	554	4.5	1.0	4.7	4.0 - 5.2	1 - 6
Social support	554	4.1	0.6	4.1	3.8 - 4.5	1 - 5
Practical assistance	554	4.2	0.7	4.3	4.0 - 4.8	1 - 5
Emotional support	554	4.1	0.6	4.2	3.8 - 4.6	1 - 5
Social integration	554	3.9	0.6	4.0	3.5 - 4.3	1 - 5
Health information seeking	503	4.0	0.8	4.1	3.5 - 4.6	1 - 5

Table 3.6: Descriptive statistics for questionnaire data. For health status, a lower value indicates better self-rated health, while for other questionnaires, higher values indicate higher ability across various constructs being assessed.

Chapter IV

Discussion

The psychometric properties of the PAM® survey in a sample suffering from macular edema were investigated. The PAM® survey demonstrated to be a reliable and valid measurement tool. The reliability of the PAM® survey is moderate in the lower to medium activation ability level. In the high patient activation area, the assessment is conflicted with more error. In the main person ability range a moderate reliability was observed. Therefore, overall, the PAM® survey showed adequate measurement precision. Associations between patient activation and health status, self-efficacy and quality of life indicated construct validity. Regarding items themselves, all showed ceiling effects: medium patient activation ability is sufficient for a patient to choose "Agree strongly" in all items.

Patients included in this study ended up showing high patient activation, indicating study participants were already engaged in their own health management. The PAM® score can range from 0 - 100 and was found to be 74.1 (SD: 13.7) in this study. 73% of participants were in level 4, indicating a healthy lifestyle is preserved even during stressful periods. This score is higher than usually found, e.g. 56.6 (SD: 12.9) [70]; 57.6 (SD 15.5) [129]; 61.9 [41]; to 66.4 (SD: 15.4) [130]. However, this study is not the first one to find high PAM® scores. In Israel patients with inflammatory bowel disease showed similar high scores with 79% of participants being in PAM® level 4 [114]; even more patients (83%) were at the highest level in a study including older adults in the US [131]. Recently, multimorbid patients in Germany were reported to have scores similar to those in this study with a mean of 76.1 (SD: 16.4) [132].

Differences in patient activation scores can generally be attributed to two causes:

cultural or disease-specific effects. The scores obtained in this study are higher than those typically found in the US, for example, 60.8 (SD: 12.3) [133] or 66.6 (SD: 16.8) [134]. In the US, approximately 41% to 46% fall into level 4, 33% to 37% in level 3, 14% to 15% in level 2, and 7% in level 1 [135].

Furthermore, the mean score of this study is higher than scores previously reported in Europe [43]. The scores ranged from 61.2 (SD: 0.34) in the Netherlands to 67.2 (SD: 0.66) in a German-speaking sample, primarily consisting of Germans. Consequently, in this study, a higher proportion of participants were classified in the highest activation level, level 4. Although a similar finding was reported in Europe, specifically Finland [115], where over half of the participants were in level 4, the percentages observed in our study are unusually high compared to other studies. In Europe, the typical distribution is approximately 31% to 56% in level 4, 25% to 28% in level 3, 9% to 20% in level 2, and 9% to 17% in level 1 [43]. Moreover, a comparison of the PAM® survey in different European countries revealed that German-speaking countries have more individuals in level 4 compared to Danish and Dutch samples [43].

Comparing conditions in different countries and therefore different cultures, accessibility of the health care system and patient satisfaction must be considered, since higher patient activation is associated with accessibility to health care services [12] and satisfaction with the health care system [62]. One reason for the high scores found in this study could be that the health care system in Austria is easy accessible [136]. In Austria, universal health insurance coverage ensures that all residents have access to medical services and treatments, regardless of their income or social status. Medical services and treatment are usually covered by the government. Furthermore, doctors and other people working in health care are well-educated and ensure that patients can receive high-quality care across the country. A comparison of the different health care systems in Europe revealed medium accessibility of health care in Austria, like many other European countries [137]. Other countries in this category are Italy, Norway and Germany. The mean scores in this countries were found to be about 66 [42, 138] in Italy, 51.9 in Norway [76], and 68.3 [45] to 76.1 [132] in Germany. Even though the PAM® scores are slightly different than in this sample, most scores refer to the same level of patient activation. The sample in Norway is distinct to the others regarding age, which may explain their lower scores. Comparing countries judged among the best regarding accessibility of the health care system, most scores are not distinct from countries with medium accessibility, e.g. Denmark (mean score 64.2), However in Spain, more than 50% of participants endorsed all the items [50]. Among countries with the worst accessibility is Turkey, where only 20.2% were in activation level 3 and 6.2% were in activation level 4. However, it is important to note that the mean PAM® score and the distribution across activation levels can differ between populations [43], due to the sample characteristics such as age or disease. Therefore these comparisons just discussed are influenced by more than just the accessibility of the health care system, an explanation for the inconsistent results regarding level of accessibility and PAM® scores.

Further comparing possible cultural influences on patient activation, patient satisfaction must be considered. In contrast to nearly half of the European countries, in Austria most patients are not charged any fee to see a physician. In other countries, patients have to pay a portion of visit expenses themselves [137]. The patient satisfaction with the costs of general practice care were compared in Europe, Portugal and Turkey were among the countries with the lowest satisfaction. Compared to this sample, PAM® scores of 58.5 (SD: 10.1) in Portugal and 6% of Turkish patient being in level 4 are lower as well. Looking at the US, where no universal health insurance coverage exists, and patients have to bear the costs, they are most likely to report access problems related to costs [139]. Additionally, patients in the US generally have lower PAM® scores compared to those in Europe [43].

Moreover, in Austria, the majority believes that the system works quite well, with only minor improvements needed [4]. Approximately two-thirds of men and half of women expressed this opinion. In terms of the quality of medical care, six out of ten individuals aged 15 and above rated it as "excellent" or "very good." However, approximately one-tenth of the population expressed less satisfaction, with women aged 45 to 59 being particularly likely to rate it as "sufficient" [4]. Even though satisfaction with the health care system was not assessed in this study, the general high satisfaction in Austria may partially explain the high scores found in this study.

Additionally, health care interventions showed to increase patient activation [140]. In Austria, for diabetic patients a disease management program is offered by the health insurance provider (`https://www.gesundheitskasse.at/cdscontent/?contentid=10007.890777&portal=oegkportal`). It consists of diabetes education courses and aims to help those affected to understand and cope better with the disease and prevent further progression. The courses include how to manage everyday life, all the physiological and psychological components of diabetes, as well as the side effects. This program is predominantly conducted on a single afternoon. In this study, 30% (N=107) of patients suffering from diabetes were part of a disease management program offered by the health insurance provider. They showed to have a similar score compared to the rest of the sample, with a mean PAM® score of 74.6 (SD: 13.4.) Patients suffering from diabetes who are not part of the disease management program showed a mean PAM® score of 72.8 (SD: 14.1).

Regarding disease specific effects on patient activation, a comparison of scores to a similar population is not possible, since patients with macular edema were not studied yet in this regard. However, most study participants suffer from diabetes and/or hypertension, two diseases already studied using the PAM® survey. Different samples of diabetic patients completed the PAM® survey, their scores being lower with 59 (SD: 12) [141], 57 (SD: 14) [51] and 59 (SD: 10) [142]. Our sample also showed higher scores than patients suffering from hypertension 61 (SD: 12) [133] and 61 (SD: 16) [58]. The scores from a sample having both, diabetes and hypertension 60 (SD: 13) [58] were lower than those found in our study. Since scores in this sample were generally higher, having one of these two diagnoses (diabetes and hypertension) alone does not seem to have a substantial impact on patient activation. One important aspect may be the impairment of vision. The PAM® survey was used to study populations with diseases affecting vision before. In a sample with vision loss the scores were lower than in our study (58.5 (SD: 15.0)) [68]. Patients suffering from wet age related macular degeneration showed to be less activated than our sample with 47% being in activation level 3 and 10% in level 4 [69]. Given the higher scores observed in this sample, impaired vision alone does not appear to impact the level of patient activation.

Differences in self-management tasks between different diseases must be considered when studying patient activation. Looking at severe chronic conditions, in a study with lung cancer patients similarly high activation scores were found (71.7 SD: 18.3) as in our study [143]. Studies on patients with life-threatening long-term illnesses (cardiac diseases or cancer) showed higher activation with a mean score of 68.0 (SD: 12.5) than other samples in the US [144]. When patient groups with different vascular diseases were investigated, the group with the most severe disease showed the highest activation [145]. It may be that if the disease gets really threatening, patients start to engage to prevent the worst outcome.

Additional to cultural and disease specific effects on the measurement of patient activation, higher activation is associated with better physical health [131] and self-perceived health status [115]. In this study, self-perceived health status was only rated as bad or worse in 6% of all cases, indicating an overall good health status of participants. This observation potentially provides a partial explanation for the high scores that were seen in this study.

4.1 Psychometrics

4.1.1 Model selection

To analyze the PAM® survey in greater detail, item response theory (IRT) was used. Firstly, an appropriate model for describing the data was sought. Given the absence of a definitive process or strategy for selecting the best model for a given dataset, various approaches were used. These included content-based considerations, comparisons of fit-indices, statistical comparisons, and taking into account IRT models previously utilized in analyzing the PAM® survey.

Previous analyses of the PAM® using IRT have been conducted in different samples. During its development, the PAM® was analyzed by IRT, although the exact model used was not explicitly stated, nor the process how they come to choose this model, which unfortunately is a common limitation in many studies [12]. However, based on the results, it can be inferred that the item discrimination parameter was not estimated during the development of the PAM®. Other studies have also chosen models where no item discrimination parameter was estimated, such as the Rasch model (possibly the rating scale model) [46, 52, 54, 75, 53, 51], and the Partial Credit Model (PCM) [42, 47]. One study reported that the RSM theoretically fitted better than PCM to analyze the PAM® [50], others do not mention model comparison e.g. [46, 54, 51].

During the model selection process in the present study, statistical comparisons indicated that the discrimination parameter should be included in the model. This parameter provides information about the measurement quality of an item and therefore items that do not perform well can be identified. Content-based considerations further supported the inclusion of the discrimination parameter, as it is logical to assume that each item possesses distinct power in differentiating patients in their level of activation. Items within the PAM® capture different aspects, highlighting the need for consideration of different item parameters.

4.1.2 Final model

Regarding the IRT analysis, out of all models compared, the GPCM was chosen. Fit indices indicated a good model fit for the whole questionnaire. Furthermore, all items showed ceiling effects, indicating that the items of the PAM are generally endorsed easily. This result is comparable to studies, that also found ceiling effects in other European countries [42, 46]. In order to further examine the psychometric properties of the questionnaire, an in-depth analysis of the individual items was conducted. An overview is given in table 4.1. Only items showing malfunctioning will be discussed. DIF will be discussed afterwards.

Table 4.1: Measurement quality of PAM® items

Item	**Agree strongly**	**Agree & Agree strongly**	**Thresholds**	**Discrimination**	**Information**
1	88 %	98 %	narrow	moderate	moderate
2	79 %	98 %	broad	high	moderate
3	77 %	96 %	medium	moderate	moderate
4	75 %	91 %	disordered	moderate	moderate
5	55 %	91 %	broad	high	moderate
6	61 %	87 %	almost disordered	low	low
7	57 %	80 %	disorderd	low	low
8	60 %	85 %	almost disordered	low	low
9	43 %	77 %	medium	moderate	moderate
10	59 %	88 %	medium	moderate	moderate
11	57 %	90 %	broad	high	high
12	53 %	87 %	broad	high	high
13	47 %	84 %	broad	moderate	moderate

For *item 4* ("I know what each of my prescribed medications do"), patients chose “Not applicable” most often (2% of all cases) compared to the other items, as found in other studies e.g. [74]. Despite suffering from a chronic disease, not everyone has to take medication. For patients with a macular edema due to retinal vein occlusion, this can be their first diagnosis, without a history of a chronic disease yet. Interestingly, the middle response option had the highest probability of being chosen for only a small range of ability levels. This indicates that for this item three answer categories may be insufficient to measure patient activation. Furthermore, these results are reflected in the item category thresholds: disordered thresholds were found for item 4, indicating that participants were not using the response options in the expected order according to their underlying patient activation ability level. Specifically, the response options for this item with the highest threshold was not between "Agree" and "Agree strongly". In a well functioning scale there should be no disordered thresholds, because that is counter-intuitive: a higher patient activation ability is needed to choose between "Disagree strongly & Disagree" and "Agree" than to choose between "Agree" and "Agree strongly". Therefore, the validity of the questionnaire can be impacted and the IRT assumption of monotonicity is violated [102]. However, the ordering requirement is not always seen essential: some argue, the thresholds can be arranged in any sequence [146]. Item parameters are defined logically, but they do not have to be associated with the intended order of the response categories. If disordered thresholds are found, this is an implication for a problem with the intended ordering of response categories. Disordered thresholds should be seen as an opportunity for diagnosis and understanding incorrect ordering [146]. This study is not the first one to find disordered thresholds for item 4 [47], highlighting the importance of the need to improve this item in measuring patient activation. Specifically, results indicate that item 4 could be answered with "Yes" or "No" only. Despite the disordered thresholds, this item delivers medium information about patient activation, indicating that it yielded a moderate amount of information regarding the measured construct patient activation.

For *item 6* ("I am confident I can tell my health care provider concerns I have even when he or she does not ask") the middle response option never had the highest probability of being chosen, suggesting that these items may benefit from a reduction in the number of response options. Additionally, item 6 had almost disordered thresholds, with a very small difference in threshold values between adjacent response options. This indicates that it distinguished between patients with different patient activation across a very small range. The difference in the patient activation ability between two individuals, choosing either response category can be irrelevant in management of their own health. Item 6 provided least discrimination and amount of information across all ability levels, indicating that it is not able to distinguish well between patients with high and low patient activation. Furthermore, this item exhibited the lowest loading on factor one, specifically it was below 0.50, a benchmark that should be reached according to [38]. These findings suggest that item 6 may be less useful for measuring patient activation.

For *item 7* ("I am confident that I can carry out medical treatments I may need to do at home") the same pattern was found in the item characteristic curves. These results are reflected in the item category thresholds as well: disordered thresholds were found for item 7, indicating that participants were not using the response options in the expected order according to their underlying patient activation ability level. Consequently, item 7 provides low amounts of information. This item is not able to distinguish patients with low and high activation.

Item 8 ("I understand the nature and causes of my health condition(s)") had almost disordered thresholds, with a very small difference in threshold values between adjacent response options. It provided relatively low amounts of information across all ability levels. Consequently, it is not able to distinguish well between patients with high and low patient activation.

Summarizing the results of malfunctioning items suggests that these four items would benefit from changes in item wording or answer options to contribute more effectively to measuring patient activation.

4.1.2.1 Differential item functioning

This study also aimed to investigate item fairness: if patients are evenly activated, there should be no other factor systematically influencing their item responses. An overview is given in table 4.2.

Item 7 was the only one where an interviewer bias was found: it was easier to endorse for patients with one interviewer (VW). This may be due to the fact that many participants did not understand what the item was asking for; it had often to be further explained by a standardized sentence. For future assessment it may be beneficial to add an example for such a medical treatment to prevent a misunderstanding of this particular item. Moreover, this item showed also DIF regarding educational status: in the lower ability level, more educated patients have an advantage over less educated patients. In the higher ability range, this difference diminishes. Similar results were found before [74], but in another study it was easier for less educated participants [42]. Item 7 might demand a specific level of health-related knowledge or skills, which could be influenced by education (compare [74]).

An influence of age on the answer patterns was found for three items: *10*, *11* and *13*. Items 10 ("I have been able to maintain lifestyle changes, like healthy eating or

Table 4.2: Differential item functioning for PAM® items

Item	Factor Interviewer	Education	Age	Health status	Type of macular edema	Comorbidity
1						
2						X
3						
4						
5					X	
6						
7	X	X				
8					X	X
9						
10			X	X	X	X
11			X			
12						
13			X			

X = Differential item functioning found

exercising") and 13 ("I am confident that I can maintain lifestyle changes like diet and exercise even during times of stress") was easier to endorse for older patients, while item 11 ("I know how to prevent further problems with my health condition") was easier to endorse for younger participants. The influence of age has been found previously, but its contradictory: some items were found to be easier for younger [42], some items have been found to be easier for older people [46, 47]. In the studies where items were easier for older patients, those patients belonged to a similar age group, specifically, they were over 65 years old [46, 47]. The relevance of some PAM® items may differ for younger patients who are still working compared to older, retired adults. Moreover, the influence of multiple conditions and therefore impairments in their daily routine may also be relevant for some items [147].

Furthermore, *item 10* was easier to endorse for patients with better self-rated health, as found before [47]. For example, diabetic people may not be able to exercise or even go for a walk, if their disease affects their legs.

The type of macular edema influenced answer behavior of patients for item *5* ("I am confident that I can tell when I need to go get medical care and when I can handle a health problem myself"), *8* & *10*). For example, item 8 concerns the understanding of its own disease, it seems to be easier for patients with a diabetic macular edema: compared to RVO, they have their diagnosis longer and therefore more time to gain understanding.

Comorbidity had impact on answer behavior of patients, DIF was found in three items (*2* (" Taking an active role in my own health care is the most important factor in determining my health and ability to function"), *8* & *10*) for the comorbid subgroups in this study. When measuring patient activation, it may be crucial to take into account the variations in self-management tasks between diseases and populations [75]. Although having the same PAM® score, this indicates even for the same disease or similar diseases, answer behavior can vary, as the patients have to face slightly different challenges in their daily lives or to understand their disease. For example, DIF was found previously within patients with kidney disease for patients in different stages [148].

Overall, small DIF effect sizes and small impacts were found, mostly for the same items. The overall influence of DIF on the overall score is considered negligible. DIF results indicated no meaningful difference in the overall PAM® score between subgroups but they could help to understand the different challenges that arise, for example, from different primary diseases or different health status.

4.1.2.2 Summary of IRT analysis

Summing up the in-depth analysis of the individual items, ceiling effects found in the descriptive analysis are reflected by multiple results in the IRT analysis. Most items demonstrated infit and outfit values within a range, effective for measuring. Infit values found indicate the PAM® survey does not differentiate much between patients. Although analyses support a unidimensional construct, six items showed a poor outfit, which could indicate that another factor influences the response behavior. An explanation may be that all items concerned (*5*, *9*, *10*, *11*, *12* & *13*) rank among the most difficult ones. As known so far, this was not found in other studies [54, 51, 68], however, they did not report the z-outfit, where the misfit occurred in our study. The same items (*4*, *6*, *7*, *8*, *10*) showed malfunctioning throughout the psychometric analysis. Overall, our findings highlight the importance of thoroughly examining the performance of individual items within a measure, and using multiple methods to identify areas for improvement and refinement. As found in other studies, the “Disagree strongly” response category was chosen in $\leq$ 10% of the responses [45, 46] and had to be merged with the “Disagree” category for analysis [75, 47].

Generally, item thresholds are distributed in the lower and medium ability range and there are no thresholds in the higher ability range. Thus, the items cannot differentiate patients with high patient activation. From the medium ability level on, "Agree strongly" has the highest probability of getting chosen. These results suggest that items *4*, *6*, *7*, and *8* may benefit from further investigation and potential revision to improve their psychometric properties. Although the response categories "Disagree strongly & Disagree" were already merged for this analysis, these results provide useful insights for the use of the PAM® survey, and highlight the importance of considering the optimal number of response options for each item.

According to the original design of the questionnaire, the items are getting more difficult as by their sequence. In this study, this was only true for the first four items. Similar results were found before in the German version of the PAM® survey [45], as well as in other translations [42, 49] or in a sample of diabetic people [51]. These findings including ours could be due to specific disease and cultural factors. Although the original goal was for item difficulty to increase over the course of the questionnaire, the fact that this is not found in many studies does not pose a problem for the interpretation of the overall PAM® score. The endorsement of items was created to belong to a specific activation level. However, one should be careful evaluating patients' health behavior on a few items, as the items do not measure the originally intended patient activation level.

Although general item information curves all look similar and thus information about patient activation ability level is given for the same range of construct, the amount of information the items provide varies. Consequently, resulting from the item information curves, the overall test information is narrow. The PAM® survey provides most information to assess patient activation in the ability range of -3 to -1, in other words in the lower to middle ability range.

Moreover, the classification into four different activation levels could not be replicated in this study. By the used analysis method, it was possible to determine how many groups are empirically distinguishable with the questionnaire. The estimation suggested to separate patients into two instead of four different groups, e.g. low and high activation. Due to the general high scores in the study sample, the categorization into these four levels are less meaningful. This is clearly visible in the proportion of the sample and the cut-off scores of the different levels: only patients in the lower quantile are not in patient activation level 4. Probably, it would be sufficient to divide the sample into two groups.

4.1.3 Reliability

The psychometric properties of the PAM® survey are adequate. Our findings indicate moderate reliability (0.75). This is lower than the reliability of 0.84 in the study validating the German version of the PAM® survey [45] and also compared to other translations in Europe (Denmark: 0.89, Italy: 0.88) [42, 46]. Measurement of patient activation is conflicted with more error in the high ability level than in the medium and lower level. Consequently, survey results may vary in accuracy for different individuals. In the case of extremely high or extremely low person ability, for which only a few items with matching difficulty are available, questionnaires may no longer measure as precise as in the middle range of the trait distribution [99]. It may be due to our highly activated sample and the lack of difficult items that in our study reliability was lower [102]. In general, to investigate a specific psychological trait in a specific population, the appropriate questionnaire should be chosen for assessment. For example, if you want to measure depression in a general (non-clinical) sample you should use an instrument which 'only' detects if someone is depressive. However, if you want to measure depression in a clinical sample you should use a different questionnaire dividing the sample into different magnitudes of depression (e.g. low, middle, high).

Regarding reliability from a CTT perspective, items should measure the exact same construct to gain a good Cronbach's α value. [99] The moderate reliability concerning internal consistency could be explained that only 30% of the variance of all items was explained by one factor. Comparing these results to the Danish version of the PAM® survey, more variance was explained and reliability was higher.

4.1.4 Validity

In this study construct validity was investigated by factorial and divergent validity. About 30% of variance was explained by one underlying factor, presumably patient activation. This is a little bit lower than 35% of the variance explained by the factor patient activation in the data of German validation study [45] or 33% variance explained in the validation of the Hungarian sample [55]. However, the Danish version explained more variance with 43% in dysglycaemic patients [46] or 40 % in a neurological sample [75]. Although less variance was explained by one factor in this study, factorial validity is still acceptable.

To examine divergent validity, self-efficacy, quality of life and general mood were assessed. Patients showed generally high scores in these assessed traits, like their patient activation score. As found previously, patients who rate their health status better show higher PAM® scores. The association found in this study is comparable with previous studies [70, 114, 115]. A large association (correlation coefficients >0.50) was found between patient activation and self-efficacy, as before [54]. In the German validation study, this association was lower [45]. In a sample of patients with different chronic diseases it was small (correlation coefficients >0.10 and <0.30) [50]. However, higher patient activation is constantly associated with higher self-efficacy. Higher quality of life was associated with higher patient activation as well (compare [73, 74, 116]). In this study, this effect was found to be large, as in [73, 54], in previous studies it was moderate (correlation coefficients >0.30) [74] or even not significant [50]. However, overall there is clearly a positive association [78]. Divergent validity was further supported by the medium association between patient activation and better general mood. In previous studies a medium association between depressive symptoms and patient activation was found [74, 54, 117]. Furthermore, the association between social support and patient activation was investigated, which was medium and therefore higher than expected (compare [117, 118]) It was even found to be large in older people with long-term conditions and multimorbidity [149]. Moreover, as anticipated, there was a medium association between health information seeking and patient activation (compare: [52]).

Further validity indicators were given through the conversation with patients: in questionnaire development one important step is to let participants think aloud while answering items. In the case of this study, due to the interview situation, patients could state if they did not understand an item, and it could be explained so that patient activation was assessed. Furthermore, since many answered in sentences or stories, it could be verified if patients got the item meaning. For example, after asking *item 4*, some patients began to list all their medications and the reason why they take it. Regarding *item 6*, they talked about their relationship and sympathy for their medical doctor. Many retired patients stated at *item 13* that reduced stress levels allowed them to maintain their daily healthy routine and improve their health status. Some patients offered their response on the rating scale first and then elaborated on their answers, while others needed reminders to summarize their stories concisely in a single response option. It is worth mentioning that interviewers refrained from interfering even if they might have chosen a different response category in some cases.

4.2 Patient Activation Measure® disadvantages

While the PAM® survey is a useful tool, it is not without drawbacks. The PAM® survey is not freely available and a license is required to use it in any context (`https://www.insigniahealth.com/pam/`). This limits the accessibility of the PAM® survey and reduces practicability in an every day clinical setting.

Moreover, the company licensing the PAM® survey keeps the algorithm to calculate the PAM® score a secret. To obtain the score in a research project, an excel sheet is provided, submitting the entered data to Insignia Health and returning the score for the individual. There are no bulk operations available. Dealing with a lot of data this process is troublesome and error-prone. On top of that, the algorithm to calculate the sum score is altered every now and then. Furthermore, cut-off values for patient activation levels changed over time. These are kept a secret as well. In the literature, different classifications are found, with a trend of increasing cut-off values. In the beginning, cut-off scores were the following: Level 1 (score 0–47.0), level 2 (47.1–55.1), level 3 (55.2–67.0) and level 4 (67.1–100) [62]. A few years later, they were changed to level 1 (0.0–47.0), level 2 (47.1–55.1), level 3 (55.2–72.4), and level 4 (72.5–100) [150]. However, used cut-off values are rarely reported since they are intellectual property of Insignia Health. These last two points discussed makes it difficult to compare results of the PAM® survey across different studies.

Moreover, there are some psychometric aspects limiting the usefulness of the PAM® survey. The ceiling effects found in this study are not uncommon for the PAM® survey, (compare [42, 46, 148]). Especially, in one study, more than 50% of patients endorsed all the items [50]. From a psychometric perspective, the questionnaire may work well enough to differentiate patients at lower ability levels, but highly activated patients may not be well differentiated. This limits its usefulness in certain populations. While the PAM® survey was constructed, only five items were created to capture patient activation in level 3 and 4. If the questionnaire is to be used in a sample, in which a high level of patient activation can be assumed, additional difficult items are required that allow differentiation in the high ability range.

4.3 Influence of administration mode

A further interesting methodological point of this study is, that the questionnaires had to be read aloud to the patients, because they could not read well in-between examinations of the eye. Usually, the used questionnaires in this study are completed by individuals themselves. However, the mode of questionnaire administration can have serious effects on assessed data, for example through cognitive burden and response bias [151, 152]. These effects impact the reliability, consistency of responses, and the degree of agreement expressed. Basically, there are two different modes of completing questionnaires: those involving an interviewer, such as telephone interviews, and those that do not require an interviewer, like self-administration [153]. Essentially, in answering questionnaires, there are four main steps that require cognitive effort from respondents: understanding the question, recalling relevant information from memory, evaluating the connection between the retrieved information and the question, and transmitting the response [151].

Comparing differences between administration modes reveals three factors varying from mode to mode: interviewer effects, media related factors and factors influencing information transmission [153]. Information transmission includes delivery of information, means of communication and handling communication flow. Media related factors concerns the presentation of the questionnaire (oral vs. visual). These media related factors also influence cognitive processes [153] and affect the cognitive effort on individuals [151] while answering questionnaires. The most burdensome methods are likely visual and written self-administration, as they require respondents to be literate in the survey language, have no visual impairments, and possess the manual ability to complete the questions. For example, these methods involve ticking a box on a paper questionnaire to indicate responses. They also need to follow routing instructions [151]. The least burdensome method maybe is the face-to-face interview, as it only requires respondents to understand and speak the same language as the questions, without the need for reading skills. There are more advantages to this completion mode: an interviewer who establishes a friendly atmosphere can improve the likelihood of obtaining responses, sustain motivation when dealing with lengthy questionnaires, clarify any uncertainties, offer visual aids, employ memory techniques to aid recall, and control the sequence of questions. Moreover, interviewers can receive training to follow complex instructions and instructions for terminating the interview [151].

However, one disadvantage is the possible interviewer bias. The interviewer may consciously or unconsciously influence individuals responses to the questionnaire [152]. Another possible disadvantage is a response bias: through the presence of a interviewer, more social acceptable answers may be given.

Regarding this study, the interview mode partially developed to a conversation and the interviewer got more information than the items asked for. Most of the patients answer more than just “Agree” – they do like to tell their personal story. For example, one patient said she had cookbooks specifically for people with diabetes, but could not use the recipes due to expensive ingredients. She is already engaged in the management of her diabetes, but needed more information how to prepare a healthy meal. In a clinical setting, the additional information communicated by the patients could be beneficial to capture the health challenge patients are currently struggling with. Additional information given to a questionnaire by patients proved to be helpful before in a different area: a comprehensive diagnosis of depression necessitates the utilization of both self-reported survey data and interpersonal communication. A systematic review of the diagnostic process for depression has established that supplementary patient-provided information enhances the accuracy and precision of health problem identification. Therefore, the suggested diagnostic procedure involves administering a self-report questionnaire initially, followed by a subsequent personal interview [154].

4.4 Representativeness of the sample

After gaining insight how the PAM® survey works in an everyday clinical setting with chronic disease patients, a further open question remains, if the study findings are generalizable to a broader Austrian population suffering from a chronic disease. In comparison to Austrians in general, our study sample reveals some similarities and differences. The first difference is the age in this sample (69 years) compared to the Austrian population, who is on average 43 years old. The second difference is a lower proportion of women in the sample (42%) than in the Austrian population with 51 % women.

Comparing the self-rated health status between the study sample and the Austrian population, at least two factors have to be considered: age and chronic disease status. Our sample has a median age of 69 and suffers from at least one chronic disease, and self-rated their health status as "very good" in 11% , "good" in 43% and "moderate" in 40% of cases. Among the elderly population (75+ age category), 38% rate their health status as moderate. Less than half of individuals with chronic diseases rated their health as at least good[4]. Taking the influence of age and chronic disease into account, the self-rated health status of participants of this study may be comparable to the general Austrian population.

In terms of specific health conditions, hypertension which most of the study sample suffer from is the third most chronic disease in Austria and more importantly, in the age group corresponding to the study sample the most common. Among Austrians aged 75 and over, around 51% of men and 56% of women are affected by hypertension. Diabetes is a common comorbidity in this study sample, due to the population of interest. Compared to the Austrian population, the proportion of people suffering from diabetes is higher in the sample. Austrians aged 75 and over, 16.2% men self-reported being diabetic, with 15.5% being diabetic women [4].

Overall, the conclusion was drawn that the results of this study can be generalized to Austrians suffering from a chronic disease. Therefore, the PAM® survey should produce reliable and valid results in a variety of healthcare settings in Austria.

4.5 Limitations

4.5.1 Selection bias

In any survey, there is a concern regarding the possibility of selection bias due to its voluntary nature, as described in many studies (e.g. [155, 69]). In general, selection bias describes that some individuals have a greater chance to be included in a study than others. Patients who choose to participate may be already more activated, while patients who show low activation have no interest in talking about their management of health. Nonetheless, almost all eligible patients (94%) agreed to participate, and the impact of non-participation on the results is expected to be minimal. While selection bias is a common concern in many studies, the effect in this particular study is likely to be small, given the high inclusion rate of eligible patients. Another potential issue with this study is the possibility of sampling bias, a form of selection bias. This describes that some individuals of the target population are more likely than others to be included into the study [156]. In this study, the inclusion of patients was influenced by accessibility and availability. It is unclear whether the patients interviewed are truly representative for patients with macular edema. Nevertheless, it is noteworthy that almost all recurring patients were interviewed throughout the two-year study period.

4.5.2 Sample size

Another limitation of this study is the sample size. IRT models require a certain sample size for stable estimations, especially if DIF should be estimated [97]. Therefore, initially the aim was to include 700 patients. However, as the data collection progressed more slowly than initially thought and already lasted over one year, a new aim was set to reach a minimum of 500 patients. This study managed to achieve a sample size exceeding 500, which is generally adequate for stable estimations. However, due to the answer patterns observed in this study, the initial models including four answer categories could not be estimated stably. It is recommended that each answer category should be chosen more than five times to gain stable model estimation [121], which was not the case in this study. Estimating models with three categories, instead of the originally four, improved the results. Nevertheless, the sample size limitation was further highlighted during DIF analysis. Due to small amount of patients in certain subgroups, estimations could not be done as intended, and some groups had to be combined or reduced. For example, to investigate if the interviewer had an effect on the answer patterns, the original plan was to compare all four interviewers to each other at once. However, it was only possible to compare one interviewer against the other three together.

4.5.3 Social desirability

Another topic arises with the interview situation: in comparison with other questionnaire completing modes, a face-to-face interview faces a high risk of social desirability bias [151]. Social desirability bias describes the tendency of participants in studies to present themselves in a way that is supposedly more socially acceptable than reality [38]. Self-report questionnaires are answered in a way, that answers of participants are not answered in a way to represent themselves, but to fit to social norms and expectations. This adjustment is usually done unconsciously to present a positive image of oneself or to avoid negative judgments. In general, there are two different types of social desirability: underreporting and overreporting. Underreporting occurs if socially undesirable behavior is understated (e.g. smoking). Overreporting describes the overstating of socially desirable behavior (e.g. eating healthy) [38]. Furthermore, social desirability consists of two parts: self-deceptive enhancement and impression management [99]. The latter assumes that people strive to manage and control the impression they make on others. Impression management is not a behavior in exceptional situations, but a very essential element of our behavior in everyday social contexts. Self-deceptive enhancement represents the rather unconscious tendency to produce favorable self-evaluations, however oneself regards as honest [99].

In this study, patients may have over reported their engagement in their health management, stressed by the interview situation. This administration mode is more conflicted with social desirability than self-administration [153]. The effect of social desirability is generally stronger in face-to-face interviews than in written surveys, since in the latter the questionnaire administrator is not present and thus subjective anonymity is more likely to be guaranteed. The influence of social desirability is favored by a high face validity of the questionnaire, whereby the participant succeeds in recognizing the measurement principle [99] and therefore is able to provide more socially acceptable answers.

In this study, based on the behavior of the patients, an atmosphere was established prior to the interview where their thoughts and feelings felt welcomed. Most of them talked freely about their life, lifestyles and health behaviors, including their everyday struggle with their disease and self-care.

To gain more insight if social desirability had an influence in this study, other studies using the questionnaire in different modes are compared. The PAM® survey was administered in face-to-face interviews before. It was used to assess patient activation of individuals with vision loss in an interview setting [68]. Their patient activation scores were comparable to those found in other samples [141, 47, 148]. Moreover, during the development of the PAM® survey, telephone interviews were used to present the items [12]. Telephone interviews are more conflicted with social desirability than self-administration [153]. The developers decided to collect first data about their items by telephone interviews, thus a prominent influence of social desirability could not have been expected. Otherwise, the data would have had too low quality for questionnaire development. Overall, this indicates that the PAM® survey can be completed in a face-to-face interview mode without being overly influenced by socially desirable answers.

4.5.4 COVID

Due to COVID-19 the data collection of the study had to be paused for two months and lasted longer than initially planned. After the pandemic hit Austria, the daily routine in the outpatient clinics were changed, leading to fewer patient appointments at the clinic. In the end, a sample of N=554 could be reached despite the circumstances. The pandemic did not only influence the duration of the study, but the patients behavior as well. Compared to before the pandemic, patients seemed more stressed about being at the hospital and wanted to stay in the clinic as shortly as possible. Due to this reason, some declined study participation.

A comparison of study outcomes before and after COVID is not possible due to the small amount of interviews done before the pandemic hit Austria. However, during the interviews the impact of the pandemic was observed in the answers of participants. During the completion of the questionnaires, pandemic-related rules such as contact restrictions amplified patients' perception of certain assessed traits. Especially, during the assessment of self perceived social support patients either reported that this period showed them how many people they can rely on or that they were at home alone for the last 2 months. According to the data gained by the questionnaires, perceived-social support was predominantly high in our sample. In general, approximately half of the Austrian population reports experiencing strong social support. However, a small proportion of both men (7.9%) and women (7.3%) reported having little social support [4]. Higher social support is associated with a better health status [149, 117]. The study sample, divided into quintiles based on social support, showed the lowest mean score in the lowest group (N=110, M=71.2, SD: 14.1) and the highest score in the highest group (N=111, M=76.8, SD: 14.7). Furthermore, in this study, patient activation and social support share 16% common variance, highlighting the importance of providing assistance to those in need. Due to the aging population in Austria, this issue may gain importance over the next years.

4.6 Future implications

Overall, during the whole study period the importance of patient engagement was shown in a variety of settings: medical doctors were fond of the idea to gain a deeper knowledge of health behavior of their patients. Patients themselves were enthusiastic, because finally they were asked about their psychological well-being and were not only physically examined. During the interviews, some patients mentioned the importance of being seen as individuals, not just as patients with medical conditions. In the future, it would be advantageous to add a short conversation concerning patient engagement into everyday clinical practice. Dealing with the topic itself can help patients recognize their potential in their health management.

Moreover, this study showed that barriers due to patients' characteristics when administrating PROMs can be overcome. Even when the PAM® survey is not completed by patients themselves, it remains reliable and valid. As an interviewer administering a questionnaire to patients, it is important not to appear judgy and clarify that there exist no right and wrong answers on PROMs. The building of trust in the first few minutes is essential for gaining honest answers. Purposes and consequences of questionnaire administration should be explained. In this study, before start of the interview, it was clearly stated that shared personal stories, thoughts and feelings remain secret and that data are only collected and analyzed pseudonymized. Many participants seemed to have confidence in the described process, creating an encouraging atmosphere for them to openly discuss their health behaviors.

4.7 Conclusions

This study demonstrates the applicability of the PAM® survey within a routine clinical environment, allowing for prompt completion of all questionnaire items. Due to patients' limited reading abilities between examinations, the questionnaire was administered orally by interviewers. This resulted in additional information given beyond the specific item responses. Most patients tended to share more than just an answer on the rating scale, providing valuable insights into their personal experiences. This supplementary information could assist medical doctors in identifying the current challenges patients face in managing their health.

From a psychometric standpoint, the questionnaire proves to be a reliable and valid tool for assessing patient activation when administered orally. However, the results indicate that the PAM® survey performs well only within certain levels of patient activation, lacking sufficient items to capture the higher end of the activation spectrum. All items exhibited ceiling effects, with "Disagree strongly" and "Disagree" options rarely selected. This was observed for the more difficult items as well. This raises the question if differentiation of highly activated patients needs enhancement and therefore additional items targeting higher activation levels should be created to identify their current challenges in health management.

An unresolved inquiry pertains to the value of incorporating the PAM® within the organizational context of everyday clinical practice. This consideration arises due to the associated annual license fees and the requirement to submit data to Insignia Health. These two factors contribute to the perceived lack of attractiveness in utilizing the PAM®.

Overall, the PAM® serves as an effective initial screening tool, but to gain deeper insights into patient behavior of highly activated patients, the inclusion of more difficult items is necessary. Nonetheless, the PAM® survey successfully identifies patients with low levels of activation, providing an opportunity to encourage them to acknowledge their abilities, knowledge, and confidence in managing their chronic disease.

Bibliography

[1] World Health Organization. World health statistics 2022: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization, 2022. ISBN: 978-92-4-005114-0.

[2] Organization, W. H. The top 10 causes of death [Internet]. The top 10 causes of death. 2020. [cited 2023 Feb. 23]; Available from: `https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death`;

[3] GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet (London, England) 396(10258), 1204–1222. ISSN: 1474-547X. DOI: `10.1016/S0140-6736(20)30925-9`.

[4] Austria, S. Österreichische Gesundheitsbefragung 2019: Hauptergebnisse des Austrian Health Interview Survey (ATHIS) und methodische Dokumentation. Statistik Austria: Vienna, Austria.

[5] Barlow, J., Wright, C., Sheasby, J., Turner, A., and Hainsworth, J. Self-management approaches for people with chronic conditions: a review. Patient Education and Counseling 48(2), 177–187. ISSN: 07383991. DOI: `10.1016/S0738-3991(02)00032-0`.

[6] Hardman, R., Begg, S., and Spelten, E. Exploring the ability of self-report measures to identify risk of high treatment burden in chronic disease patients: a cross-sectional study. BMC Public Health 22(1), 163. ISSN: 1471-2458. DOI: `10.1186/s12889-022-12579-1`.

[7] Havelka, M., Luanin, J. D., and Luanin, D. Biopsychosocial Model – The Integrated Approach to Health and Disease. Coll. Antropol.

[8] The Lancet. What is health? The ability to adapt. The Lancet 373(9666), 781. ISSN: 01406736. DOI: 10.1016/S0140-6736(09)60456-6.

[9] Bircher, J. Towards a Dynamic Definition of Health and Disease. Medicine, Health Care and Philosophy 8(3), 335–341. ISSN: 1386-7423, 1572-8633. DOI: 10.1007/s11019-005-0538-y.

[10] Krahn, G. L., Robinson, A., Murray, A. J., Havercamp, S. M., Havercamp, S., Andridge, R., et al. It's time to reconsider how we define health: Perspective from disability and chronic condition. Disability and Health Journal 14(4), 101129. ISSN: 19366574. DOI: 10.1016/j.dhjo.2021.101129.

[11] Engel, G. L. The Need for a New Medical Model: A Challenge for Biomedicine. Science 196(4286), 129–136. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.847460.

[12] Hibbard, J. H., Stockard, J., Mahoney, E. R., and Tusler, M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health services research 39(4), 1005–1026.

[13] Uchino, B. N. Social Support and Health: A Review of Physiological Processes Potentially Underlying Links to Disease Outcomes. Journal of Behavioral Medicine 29(4), 377–387. ISSN: 0160-7715, 1573-3521. DOI: 10.1007/s10865-006-9056-5.

[14] Magnezi, R., Glasser, S., Shalev, H., Sheiber, A., and Reuveni, H. Patient activation, depression and quality of life. Patient Education and Counseling 94(3), 432–437. ISSN: 07383991. DOI: 10.1016/j.pec.2013.10.015.

[15] Lampert, T., Kroll, L. E., Lippe, E. von der, Müters, S., and Stolzenberg, H. Sozioökonomischer Status und Gesundheit. Bundes-gesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 56(5), 814–821.

[16] Wagner, E. H., Austin, B. T., and Von Korff, M. Organizing care for patients with chronic illness. The Milbank Quarterly, 511–544.

[17] Higgins, T., Larson, E., and Schnall, R. Unraveling the meaning of patient engagement: A concept analysis. Patient Education and Counseling 100(1), 30–36. ISSN: 07383991. DOI: 10.1016/j.pec.2016.09.002.

[18] McAllister, M., Dunn, G., Payne, K., Davies, L., and Todd, C. Patient empowerment: The need to consider it as a measurable patient-reported outcome for chronic conditions. BMC Health Services Research 12(1), 157. ISSN: 1472-6963. DOI: 10.1186/1472-6963-12-157.

[19] Vanstone, M., Canfield, C., Evans, C., Leslie, M., Levasseur, M. A., MacNeil, M., et al. Towards conceptualizing patients as partners in health systems: a systematic review and descriptive synthesis. Health Research Policy and Systems 21(1), 12. ISSN: 1478-4505. DOI: 10.1186/s12961-022-00954-8.

[20] Simmons, L. A., Wolever, R. Q., Bechard, E. M., and Snyderman, R. Patient engagement as a risk factor in personalized health care: a systematic review of the literature on chronic disease. Genome Medicine 6(2), 16. ISSN: 1756-994X. DOI: 10.1186/gm533.

[21] Franek, J. Self-management support interventions for persons with chronic disease: an evidence-based analysis. Ontario health technology assessment series 13(9), 1.

[22] D'Agostino, T. A., Atkinson, T. M., Latella, L. E., Rogers, M., Morrissey, D., DeRosa, A. P., et al. Promoting patient participation in healthcare interactions through communication skills training: A systematic review. Patient Education and Counseling 100(7), 1247–1257. ISSN: 07383991. DOI: 10.1016/j.pec.2017.02.016.

[23] Graffigna, G., Barello, S., Bonanomi, A., and Lozza, E. Measuring patient engagement: development and psychometric properties of the Patient Health Engagement (PHE) Scale. Frontiers in Psychology 6. ISSN: 1664-1078.

[24] Dean, S., Mathers, J. M., Calvert, M., Kyte, D. G., Conroy, D., Folkard, A., et al. "The patient is speaking": discovering the patient voice in ophthalmology. British Journal of Ophthalmology 101(6), 700–708. ISSN: 0007-1161, 1468-2079. DOI: 10.1136/bjophthalmol-2016-309955.

[25] Murphy, M., Hollinghurst, S., and Salisbury, C. Identification, description and appraisal of generic PROMs for primary care: a systematic review. BMC Family Practice 19(1), 41. ISSN: 1471-2296. DOI: 10.1186/s12875-018-0722-9.

[26] Pekonen, A., Eloranta, S., Stolt, M., Virolainen, P., and Leino-Kilpi, H. Measuring patient empowerment – A systematic review. Patient Education and Counseling 103(4), 777–787. ISSN: 0738-3991. DOI: 10.1016/j.pec.2019.10.019.

[27] Black, N., Burke, L., Forrest, C. B., Ravens Sieberer, U. H., Ahmed, S., Valderas, J. M., et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Quality of life Research 25, 1103–1112.

[28] Glasgow, R. E., Wagner, E. H., Schaefer, J., Mahoney, L. D., Reid, R. J., and Greene, S. M. Development and Validation of the Patient Assessment of Chronic Illness Care (PACIC). Medical Care 43(5), 436. ISSN: 0025-7079. DOI: 10.1097/01.mlr.0000160375.47920.8c.

[29] Kristjansson, E., Tugwell, P. S., Wilson, A. J., Brooks, P. M., Driedger, S. M., Gallois, C., et al. Development of the effective musculoskeletal consumer scale. The Journal of Rheumatology 34(6), 1392–1400. ISSN: 0315-162X, 1499-2752.

[30] Duke, C. C., Lynch, W. D., Smith, B., and Winstanley, J. Validity of a New Patient Engagement Measure: The Altarum Consumer Engagement (ACE) Measure™. The Patient 8(6), 559–568. ISSN: 1178-1653. DOI: 10.1007/s40271-015-0131-2.

[31] Johnson, M. O., Rose, C. D., Dilworth, S. E., and Neilands, T. B. Advances in the Conceptualization and Measurement of Health Care Empowerment: Development and Validation of the Health Care Empowerment Inventory. PLoS ONE 7(9). Ed. by Baradaran, H. R., e45692. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0045692.

[32] Xu, R. H., Cheung, A. W.-l., and Wong, E. L.-y. Development and validation of an instrument to measure patient engagement in Hong Kong Special Administrative Region, China. Patient Preference and Adherence Volume 12, 1667–1675. ISSN: 1177-889X. DOI: 10.2147/PPA.S171026.

[33] Osborne, R., Elsworth, G., and Whitfield, K. The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient education and counseling 66, 192–201. DOI: 10.1016/j.pec.2006.12.002.

[34] Small, N., Bower, P., Chew-Graham, C. A., Whalley, D., and Protheroe, J. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. BMC Health Services Research 13(1), 263. ISSN: 1472-6963. DOI: 10.1186/1472-6963-13-263.

[35] Bayliss, E. A., Ellis, J. L., and Steiner, J. F. Barriers to Self-Management and Quality-of-Life Outcomes in Seniors With Multimorbidities. The Annals of Family Medicine 5(5), 395–402. ISSN: 1544-1709, 1544-1717. DOI: 10.1370/afm.722.

[36] Anderson, R. M., Funnell, M. M., Fitzgerald, J. T., and Marrero, D. G. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. Diabetes care 23(6), 739–743.

[37] Barr, P. J., Scholl, I., Bravo, P., Faber, M. J., Elwyn, G., and McAllister, M. Assessment of Patient Empowerment - A Systematic Review of Measures. PLOS ONE 10(5). Ed. by Bond, K., e0126553. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0126553.

[38] Bühner, M. Einführung in die Test-und Fragebogenkonstruktion. Pearson Deutschland GmbH, 2011.

[39] Hibbard, J. and Gilburt, H. Supporting people to manage their health. An introduction to patient activation. London: The King's Fund.

[40] Graffigna, G., Barello, S., and Triberti, S., eds. Patient Engagement A Consumer-Centered Model to Innovate Healthcare. De Gruyter Open, Aug. 11, 2015. ISBN: 978-3-11-045243-3. DOI: 10.1515/9783110452440. [cited 2023 Feb. 26]; Available from: https://www.degruyter.com/document/doi/10.1515/9783110452440/html;

[41] Hibbard, J. H., Mahoney, E. R., Stockard, J., and Tusler, M. Development and Testing of a Short Form of the Patient Activation Measure. Health Services Research 40(6), 1918–1930. ISSN: 0017-9124, 1475-6773. DOI: 10.1111/j.1475-6773.2005.00438.x.

[42] Graffigna, G., Barello, S., Bonanomi, A., Lozza, E., and Hibbard, J. Measuring patient activation in Italy: Translation, adaptation and validation of the Italian version of the patient activation measure 13 (PAM13-I). BMC Medical Informatics and Decision Making 15(1), 109. ISSN: 1472-6947. DOI: 10.1186/s12911-015-0232-9.

[43] Rademakers, J., Maindal, H. T., Steinsbekk, A., Gensichen, J., Brenk-Franz, K., and Hendriks, M. Patient activation in Europe: an international comparison of psychometric properties and patients' scores on the short form Patient Activation Measure (PAM-13). BMC Health Services Research 16(1), 570. ISSN: 1472-6963. DOI: 10.1186/s12913-016-1828-1.

[44] Sacks, R. M., Greene, J., Hibbard, J., Overton, V., and Parrotta, C. D. Does patient activation predict the course of type 2 diabetes? A longitudinal study. Patient Education and Counseling 100(7), 1268–1275. ISSN: 0738-3991. DOI: 10.1016/j.pec.2017.01.014.

[45] Brenk-Franz, K., Hibbard, J. H., Herrmann, W. J., Freund, T., Szecsenyi, J., Djalali, S., et al. Validation of the German Version of the Patient Activation Measure 13 (PAM13-D) in an International Multicentre Study of Primary Care Patients. PLoS ONE 8(9). Ed. by Janda, M., e74786. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074786.

[46] Maindal, H. T., Sokolowski, I., and Vedsted, P. Translation, adaptation and validation of the American short form Patient Activation Measure (PAM13) in a Danish version. BMC Public Health 9(1), 209. ISSN: 1471-2458. DOI: 10.1186/1471-2458-9-209.

[47] Hellström, A., Kassaye Tessma, M., Flink, M., Dahlgren, A., Schildmeijer, K., and Ekstedt, M. Validation of the patient activation measure in patients at discharge from hospitals and at distance from hospital care in Sweden. BMC Public Health 19, 1–11.

[48] Steinsbekk, A. [Patient Activation Measure]. Tidsskrift for den Norske laegeforening 128(20), 2316–2318. ISSN: 0807-7096.

[49] Rademakers, J., Nijman, J., Hoek, L. van der, Heijmans, M., and Rijken, M. Measuring patient activation in the Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13). BMC Public Health 12(1), 577. ISSN: 1471-2458. DOI: 10.1186/1471-2458-12-577.

[50] Moreno-Chico, C., Paz, L. González-de, Monforte-Royo, C., Arrighi, E., Navarro-Rubio, M. D., and Gallart Fernández-Puebla, A. Adaptation to European Spanish and psychometric properties of the Patient Activation Measure 13 in patients with

chronic diseases. Family Practice 34(5), 627–634. ISSN: 0263-2136, 1460-2229. DOI: 10.1093/fampra/cmx022.

[51] Laranjo, L., Dias, V., Nunes, C., Paiva, D., and Mahoney, B. Translation and Validation of the Patient Activation Measure in Portuguese People with Type 2 Diabetes Mellitus. Acta Médica Portuguesa 31(7), 382. ISSN: 1646-0758, 0870-399X. DOI: 10.20344/amp.9072.

[52] Zrubka, Z., Vékás, P., Németh, P., Dobos, Hajdu, O., Kovács, L., et al. Validation of the PAM-13 instrument in the Hungarian general population 40 years old and above. The European Journal of Health Economics 23(8), 1341–1355. ISSN: 1618-7598, 1618-7601. DOI: 10.1007/s10198-022-01434-0.

[53] Ngooi, B. X., Packer, T. L., Kephart, G., Warner, G., Koh, K. W. L., Wong, R. C. C., et al. Validation of the Patient Activation Measure (PAM-13) among adults with cardiac conditions in Singapore. Quality of Life Research 26(4), 1071–1080. ISSN: 0962-9343, 1573-2649. DOI: 10.1007/s11136-016-1412-5.

[54] Magnezi, R. and Glasser, S. Psychometric Properties of the Hebrew Translation of the Patient Activation Measure (PAM-13). PLoS ONE 9(11). Ed. by Brucki, S., e113391. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0113391.

[55] Kosar, C. and Besen, D. B. Adaptation of a patient activatıon measure (PAM) into Turkish: reliability and validity test. African Health Sciences 19(1), 1811. ISSN: 1680-6905. DOI: 10.4314/ahs.v19i1.58.

[56] Ahn, Y.-H., Yi, C.-H., Ham, O.-K., and Kim, B.-J. Psychometric Properties of the Korean Version of the "Patient Activation Measure 13" (PAM13-K) in Patients With Osteoarthritis. Evaluation & the Health Professions 38(2), 255–264. ISSN: 0163-2787, 1552-3918. DOI: 10.1177/0163278714540915.

[57] Bahrom, N. H., Ramli, A. S., Isa, M. R., Baharudin, N., Sham, S. F. B., Yassin, M. S. M., et al. Validity and reliability of the Patient Activation Measure®(PAM®)-13 Malay version among patients with Metabolic Syndrome in primary care. Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia 15(3), 22.

[58] Zeng, H., Jiang, R., Zhou, M., Wu, L., Tian, B., Zhang, Y., et al. Measuring patient activation in Chinese patients with hypertension and/or diabetes: reliability and validity of the PAM13. Journal of International Medical Research 47(12), 5967–5976. ISSN: 0300-0605, 1473-2300. DOI: 10.1177/0300060519868327.

[59] Kinney, R. L., Lemon, S. C., Person, S. D., Pagoto, S. L., and Saczynski, J. S. The association between patient activation and medication adherence, hospitalization, and emergency room utilization in patients with chronic illnesses: a systematic review. Patient education and counseling 98(5), 545–552.

[60] Anderson, G., Rega, M. L., Casasanta, D., Graffigna, G., Damiani, G., and Barello, S. The association between patient activation and healthcare resources utilization: a systematic review and meta-analysis. Public Health 210, 134–141.

[61] Donald, M., Ware, R. S., Ozolins, I. Z., Begum, N., Crowther, R., and Bain, C. The role of patient activation in frequent attendance at primary care: a population-based study of people with chronic disease. Patient education and counseling 83(2), 217–221.

[62] Mosen, D. M., Schmittdiel, J., Hibbard, J., Sobel, D., Remmers, C., and Bellows, J. Is patient activation associated with outcomes of care for adults with chronic conditions? The Journal of ambulatory care management 30(1), 21–29.

[63] Dixon, A., Hibbard, J., and Tusler, M. How do People with Different Levels of Activation Self-Manage their Chronic Conditions?: The Patient: Patient-Centered Outcomes Research 2(4), 257–268. ISSN: 1178-1653. DOI: 10.2165/11313790-000000000-00000.

[64] Cuevas, H., Heitkemper, E., Ya-Ching, H., Dong, E. J., Garcia, A. A., and Zuniga, J. A. A systematic review and meta-analysis of patient activation in people living with chronic conditions. Patient Education and Counseling 104(9), 2200–2212. ISSN: 0738-3991. DOI: 10.1016/j.pec.2021.02.016.

[65] Innab, A. and Kerari, A. Impact of Behavioral Interventions on Patient Activation in Adults with Hypertension: A Systematic Review and Meta-Analysis. INQUIRY: The Journal of Health Care Organization, Provision, and Financing 59, 004695802210904. ISSN: 0046-9580, 1945-7243. DOI: 10.1177/00469580221090408.

[66] Hosseinzadeh, H., Verma, I., and Gopaldasani, V. Patient activation and Type 2 diabetes mellitus self-management: a systematic review and meta-analysis. Australian Journal of Primary Health 26(6), 431–442. ISSN: 1836-7399. DOI: 10.1071/PY19204.

[67] Mergelsberg, E. L., Olson, J. L., Moore, S., Jensen, J. F., Seivwright, H., Norman, R., et al. Providing evidence from practice: Evaluating 4.5 years of implementing diabetes education support program in carbohydrate management. Patient Education and Counseling 105(7), 2225–2233. ISSN: 07383991. DOI: 10.1016/j.pec.2022.02.001.

[68] Morse, A. R. and Seiple, W. Activation in individuals with vision loss. Journal of Health Psychology 26(13), 2603–2612. ISSN: 1359-1053, 1461-7277. DOI: 10.1177/1359105320922303.

[69] Skelly, A., Taylor, N., Fasser, C., Malkowski, J.-P., Goswamy, P., and Downey, L. Patient Preferences in the Management of Wet Age-Related Macular Degeneration: A Conjoint Analysis. Advances in Therapy. ISSN: 0741-238X, 1865-8652. DOI: 10.1007/s12325-022-02248-5.

[70] Skolasky, R. L., Green, A. F., Scharfstein, D., Boult, C., Reider, L., and Wegener, S. T. Psychometric Properties of the Patient Activation Measure among Multimorbid Older Adults: Psychometric Properties of the PAM. Health Services Research 46(2), 457–478. ISSN: 00179124. DOI: 10.1111/j.1475-6773.2010.01210.x.

[71] Shnaigat, M., Downie, S., and Hosseinzadeh, H. Effectiveness of patient activation interventions on chronic obstructive pulmonary disease self-management outcomes: A systematic review. Australian Journal of Rural Health 30(1), 8–21.

[72] Westman, B., Bergkvist, K., Karlsson Rosenblad, A., Sharp, L., and Bergenmar, M. Patients with low activation level report limited possibilities to participate in cancer care. Health Expectations 25(3), 914–924. ISSN: 1369-7625. DOI: 10.1111/hex.13438.

[73] Magadi, W., Lightfoot, C. J., Memory, K. E., Santhakumaran, S., Veer, S. N. van der, Thomas, N., et al. Patient activation and its association with symptom burden and quality of life across the spectrum of chronic kidney disease stages in

England. BMC Nephrology 23(1), 45. ISSN: 1471-2369. DOI: 10.1186/s12882-022-02679-w.

[74] Eyles, J. P., Ferreira, M., Mills, K., Lucas, B. R., Robbins, S. R., Williams, M., et al. Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis. Health and Quality of Life Outcomes 18(1), 121. ISSN: 1477-7525. DOI: 10.1186/s12955-020-01364-6.

[75] Packer, T. L., Kephart, G., Ghahari, S., Audulv, Versnel, J., and Warner, G. The Patient Activation Measure: a validation study in a neurological population. Quality of Life Research 24(7), 1587–1596. ISSN: 0962-9343, 1573-2649. DOI: 10.1007/s11136-014-0908-0.

[76] Moljord, I. E. O., Lara-Cabrera, M. L., Perestelo-Pérez, L., Rivero-Santana, A., Eriksen, L., and Linaker, O. M. Psychometric properties of the Patient Activation Measure-13 among out-patients waiting for mental health treatment: A validation study in Norway. Patient Education and Counseling 98(11), 1410–1417. ISSN: 07383991. DOI: 10.1016/j.pec.2015.06.009.

[77] Greene, J., Hibbard, J., and Tusler, M. How Much Do Health Literacy and Patient Activation Contribute to Older Adults' Ability to Manage Their Health?

[78] Golubinski, V., Oppel, E.-M., and Schreyögg, J. A systematic scoping review of psychosocial and psychological factors associated with patient activation. Patient Education and Counseling 103(10), 2061–2068. ISSN: 07383991. DOI: 10.1016/j.pec.2020.05.005.

[79] Graffigna, G., Barello, S., and Bonanomi, A. The role of Patient Health Engagement Model (PHE-model) in affecting patient activation and medication adherence: A structural equation model. PloS one 12(6), e0179865.

[80] Schönenberg, A., Teschner, U., Prell, T., and Mühlhammer, H. M. Validation and Psychometric Analysis of the German Translation of the Appraisal of Self-Care Agency Scale-Revised. Healthcare 10(9), 1785. ISSN: 2227-9032. DOI: 10.3390/healthcare10091785.

[81] Tranos, P. G., Wickremasinghe, S. S., Stangos, N. T., Topouzis, F., Tsinopoulos, I., and Pavesio, C. E. Macular edema. Survey of ophthalmology 49(5), 470–490.

[82] Daruich, A., Matet, A., Moulin, A., Kowalczuk, L., Nicolas, M., Sellam, A., et al. Mechanisms of macular edema: beyond the surface. Progress in retinal and eye research 63, 20–68.

[83] Fydrich, T., Geyer, M., Hessel, A., Sommer, G., and Brähler, E. Fragebogen zur Sozialen Unterstützung (F-SozU): Normierung an einer repräsentativen Stichprobe. Diagnostica 45(4), 212–216. ISSN: 0012-1924, 2190-622X. DOI: 10.1026//0012-1924.45.4.212.

[84] Dunkel, D., Antretter, E., Fröhlich-Walser, S., and Haring, C. Evaluation der Kurzform des Fragebogens zur Sozialen Unterstützung (SOZU-K-22) in klinischen und nichtklinischen Stichproben. PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie 55(5), 266–277. ISSN: 0937-2032, 1439-1058. DOI: 10.1055/s-2004-834746.

[85] Fydrich, T., Sommer, G., and Brähler, E. F-SozU, Fragebogen zur Sozialen Unterstützung. Hogrefe.

[86] Dalbert, C. Subjektives Wohlbefinden junger Erwachsener: Theoretische und empirische Analysen der Struktur und Stabilität. Zeitschrift für differentielle und diagnostische Psychologie.

[87] Chasiotis, A., Wedderhoff, O., Rosman, T., and Mayer, A.-K. GAINS-Goals Associated with Health Information Seeking. 2018.

[88] Chasiotis, A., Wedderhoff, O., Rosman, T., and Mayer, A.-K. Why do we want health information? The goals associated with health information seeking (GAINS) questionnaire. Psychology & Health 35(3), 255–274. ISSN: 0887-0446, 1476-8321. DOI: 10.1080/08870446.2019.1644336.

[89] Schwarzer, R. and Jerusalem, M. Skalen zur erfassung von Lehrer-und schülermerkmalen. 1999.

[90] Leibniz-Institut für Psychologie (ZPID) [Internet]. Open Test Archive. 2019. [cited 2023 Apr. 12]; Available from: https://www.testarchiv.eu/en;

[91] De Bruin, A. Health Interview Surveys: Towards International Harmonization of Methods and Instruments. WHO Regional Publications, European Series, No. 58. ERIC, 1996.

[92] Jährliche Personeneinkommen [Internet]. Statistik Austria. 2021. [cited 2023 Feb. 17]; Available from: https://www.statistik.at/statistiken/bevoelkerung-und-soziales/einkommen-und-soziale-lage/jaehrliche-personeneinkommen;

[93] Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria. [Internet]. 2023. [cited 2022 Sept. 15]; Available from: https://imi.medunigraz.at/en/services#c44617;

[94] LimeSurvey - einfache Online-Umfragen [Internet]. 2023. [cited 2022 Sept. 11]; Available from: https://www.limesurvey.org/de/;

[95] Reeve, B. B. and Fayers, P. Applying item response theory modeling for evaluating questionnaire item and scale properties. Assessing quality of life in clinical trials: methods of practice 2, 55–73.

[96] Cappelleri, J. C., Lundy, J. J., and Hays, R. D. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clinical therapeutics 36(5), 648–662.

[97] Kubinger, K. D., Rasch, D., and Yanagida, T. On designing data-sampling for Rasch model calibrating an achievement test. Psychological Test and Assessment Modeling 51(4), 370.

[98] McHorney, C. A. and Tarlov, A. R. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Quality of Life Research 4(4), 293–307. ISSN: 0962-9343, 1573-2649. DOI: 10.1007/BF01593882.

[99] Moosbrugger, H. and Kelava, A. Testtheorie und Fragebogenkonstruktion. 2., aktualisierte und u?berarbeitete Auflage. Springer-Lehrbuch. Berlin: Springer, 2012. ISBN: 978-3-642-20071-7.

[100] De Champlain, A. F. A primer on classical test theory and item response theory for assessments in medical education: Classical test theory and item response theory. Medical Education 44(1), 109–117. ISSN: 03080110, 13652923. DOI: 10.1111/j.1365-2923.2009.03425.x.

[101] Gliem, J. A. and Gliem, R. R. "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales". *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.* 2003.

[102] Embretson, S. E. and Reise, S. P. Item response theory. Psychology Press, 2013.

[103] Ostini, R. and Nering, M. Polytomous Item Response Theory Models. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc., 2006. ISBN: 978-0-7619-3068-6 978-1-4129-8541-3. DOI: 10.4135/9781412985413.

[104] Verhelst, N. D. and Glas, C. A. W. "The One Parameter Logistic Model". *Rasch Models: Foundations, Recent Developments, and Applications.* Ed. by Fischer, G. H. and Molenaar, I. W. New York, NY: Springer, 1995, 215–237. ISBN: 978-1-4612-4230-7. DOI: 10.1007/978-1-4612-4230-7_12. [cited 2023 Jan. 09]; Available from: https://doi.org/10.1007/978-1-4612-4230-7_12;

[105] Muraki, E. "A Generalized Partial Credit Model". *Handbook of Modern Item Response Theory.* Ed. by Linden, W. J. van der and Hambleton, R. K. New York, NY: Springer New York, 1997, 153–164. ISBN: 978-1-4757-2691-6. DOI: 10.1007/978-1-4757-2691-6_9. [cited];

[106] Andrich, D. Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. Applied Psychological Measurement 2(4), 581–594. ISSN: 0146-6216, 1552-3497. DOI: 10.1177/014662167800200413.

[107] Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement 34, 100–100.

[108] Hu, L.-t. and Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal 6(1), 1–55.

[109] Rouquette, A., Hardouin, J.-B., Vanhaesebrouck, A., Sébille, V., and Coste, J. Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. PloS one 14(4), e0215073.

[110] Linacre, J. M. What do infit and outfit, mean-square and standardized mean. Rasch measurement transactions 16(2), 878.

[111] Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 307–333.

[112] Wright, B. D. and Masters, G. N. Rating scale analysis. Chicago: Mesa Press, 1982. 206 pp. ISBN: 978-0-941938-01-3.

[113] Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Cook, K. F., Crane, P. K., Gibbons, L. E., et al. Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to measures of physical functioning ability and general distress. Quality of Life Research 16, 43–68. ISSN: 0962-9343, 1573-2649. DOI: 10.1007/s11136-007-9186-4.

[114] Haj, O., Lipkin, M., Kopylov, U., Sigalit, S., and Magnezi, R. Patient activation and its association with health indices among patients with inflammatory bowel disease. Therapeutic Advances in Gastroenterology 15, 175628482211287. ISSN: 1756-2848, 1756-2848. DOI: 10.1177/17562848221128757.

[115] Tusa, N., Kautiainen, H., Elfving, P., Sinikallio, S., and Mäntyselkä, P. Relationship between patient activation measurement and self-rated health in patients with chronic diseases. BMC Family Practice 21(1), 225. ISSN: 1471-2296. DOI: 10.1186/s12875-020-01301-y.

[116] Hendriks, S. H., Hartog, L. C., Groenier, K. H., Maas, A. H. E. M., Van Hateren, K. J. J., Kleefstra, N., et al. Patient Activation in Type 2 Diabetes: Does It Differ between Men and Women? Journal of Diabetes Research 2016, 1–8. ISSN: 2314-6745, 2314-6753. DOI: 10.1155/2016/7386532.

[117] Gleason, K. T., Tanner, E. K., Boyd, C. M., Saczynski, J. S., and Szanton, S. L. Factors associated with patient activation in an older adult population with functional difficulties. Patient education and counseling 99(8), 1421–1426. ISSN: 1873-5134; 0738-3991. DOI: 10.1016/j.pec.2016.03.011.

[118] Bos-Touwen, I., Schuurmans, M., Monninkhof, E. M., Korpershoek, Y., Spruit-Bentvelzen, L., Graaf, I. Ertugrul-van der, et al. Patient and Disease Characteristics Associated with Activation for Self-Management in Patients with Diabetes, Chronic Obstructive Pulmonary Disease, Chronic Heart Failure and Chronic Renal Disease: A Cross-Sectional Survey Study. Plos One 10(5), e0126400–e0126400. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0126400.

[119] Cohen, J. Statistical power analysis for the behavioral sciences. 2. ed., reprint. New York, NY: Psychology Press, 2009. 567 pp. ISBN: 978-0-8058-0283-2.

[120] Rubin, D. B. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, 2004. 324 pp. ISBN: 978-0-471-65574-9.

[121] Chalmers, R. P. mirt: A multidimensional item response theory package for the R environment. Journal of statistical Software 48, 1–29.

[122] Choi, S. W., Gibbons, L. E., and Crane, P. K. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. Journal of statistical software 39(8), 1–30. ISSN: 1548-7660; 1548-7660.

[123] R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing, 2021. [cited];

[124] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. Welcome to the Tidyverse. Journal of open source software 4(43), 1686.

[125] Revelle, W. Psych: procedures for psychological, psychometric, and personality research. 2020. R package version 2(7).

[126] Chaltiel, D. crosstable: Crosstables for Descriptive Analyses. 2022.

[127] Gohel, D., Fazilleau, Q., Nazarov, M., Robert, T., Barrowman, M., and Yasumoto, A. Flextable: Functions for tabular reporting. R package version 0.6 8.

[128] Wickham, H. ggplot2: elegant graphics for data analysis New York. NY: Springer.

[129] Zimbudzi, E., Lo, C., Ranasinha, S., Fulcher, G. R., Jan, S., Kerr, P. G., et al. Factors associated with patient activation in an Australian population with comorbid diabetes and chronic kidney disease: a cross-sectional study. BMJ open 7(10), e017695–2017–017695. ISSN: 2044-6055; 2044-6055. DOI: 10.1136/bmjopen-2017-017695.

[130] Greene, J. and Hibbard, J. H. Why Does Patient Activation Matter? An Examination of the Relationships Between Patient Activation and Health-Related Outcomes. Journal of General Internal Medicine 27(5), 520–526. ISSN: 0884-8734, 1525-1497. DOI: 10.1007/s11606-011-1931-2.

[131] Smith, S. G., Curtis, L. M., Wardle, J., Wagner, C. von, and Wolf, M. S. Skill Set or Mind Set? Associations between Health Literacy, Patient Activation and Health. PLoS ONE 8(9). Ed. by Baradaran, H. R., e74373. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074373.

[132] Breckner, A., Glassen, K., Schulze, J., Lühmann, D., Schaefer, I., Szecsenyi, J., et al. Experiences of patients with multimorbidity with primary care and the association with patient activation: a cross-sectional study in Germany. BMJ Open 12(8), e059100. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2021-059100.

[133] Lee, C. and Horowitz, C. R. Association Between APOL1 Risk Testing in Primary Care Settings and Patient Engagement Among Black Patients with Hypertension. preprint. In Review, Mar. 2, 2022. DOI: 10.21203/rs.3.rs-1180383/v1. [cited 2023 Mar. 06]; Available from: https://www.researchsquare.com/article/rs-1180383/v1;

[134] Smith, S. G., Pandit, A., Rush, S. R., Wolf, M. S., and Simon, C. J. The Role of Patient Activation in Preferences for Shared Decision Making: Results From a National Survey of U.S. Adults. Journal of Health Communication 21(1), 67–75. ISSN: 1081-0730, 1087-0415. DOI: 10.1080/10810730.2015.1033115.

[135] Kenney, M. Patient Activation Among Diverse Populations: A Systematic Review.

[136] Greiling, D. "Das österreichische Gesundheitswesen im internationalen Vergleich". *Gut versorgt?* Ed. by Fuchs, M., Greiling, D., and Rosenberger, M. Nomos Verlagsgesellschaft mbH & Co. KG, 2019, 21–44. ISBN: 978-3-8452-9487-2. DOI: 10.5771/9783845294872-21. [cited 2024 Mar. 27]; Available from: https://www.nomos-elibrary.de/index.php?doi=10.5771/9783845294872-21;

[137] Kringos, D. S., Boerma, W. G., Hutchinson, A., Saltman, R. B., and Organization, W. H. Building primary care in a changing Europe. World Health Organization. Regional Office for Europe, 2015.

[138] Graffigna, G., Barello, S., Bonanomi, A., and Menichetti, J. The motivating function of healthcare professional in eHealth and mHealth interventions for type 2 diabetes patients and the mediating role of patient engagement. Journal of diabetes research 2016.

[139] Mirror, Mirror on the Wall, 2014 Update: How the U.S. Health Care System Compares Internationally. DOI: 10.26099/9q4x-na97.

[140] Hibbard, J. H., Greene, J., and Overton, V. Patients with lower activation associated with higher costs; delivery systems should know their patients''scores'. Health affairs 32(2), 216–222.

[141] Van Vugt, H. A., Boels, A. M., De Weerdt, I., De Koning, E. J., and Rutten, G. E. Patient activation in individuals with type 2 diabetes mellitus: associated factors and the role of insulin. Patient Preference and Adherence Volume 13, 73–81. ISSN: 1177-889X. DOI: 10.2147/PPA.S188391.

[142] Hendriks, M. and Rademakers, J. Relationships between patient activation, disease-specific knowledge and health outcomes among people with diabetes; a survey study. BMC health services research 14(1), 393.

[143] Sun, V., Raz, D. J., Ruel, N., Chang, W., Erhunmwunsee, L., Reckamp, K., et al. A Multimedia Self-management Intervention to Prepare Cancer Patients and Family Caregivers for Lung Surgery and Postoperative Recovery. Clinical Lung Cancer 18(3), e151–e159. ISSN: 15257304. DOI: 10.1016/j.cllc.2017.01.010.

[144] Jerofke, T., Weiss, M., and Yakusheva, O. Patient perceptions of patient-empowering nurse behaviours, patient activation and functional health status in postsurgical patients with life-threatening long-term illnesses. Journal of Advanced Nursing 70(6), 1310–1322. ISSN: 03092402. DOI: 10.1111/jan.12286.

[145] Humphries, M. D., Welch, P., Hasegawa, J., and Mell, M. W. Correlation of Patient Activation Measure Level with Patient Characteristics and Type of Vascular Disease. Annals of Vascular Surgery 73, 55–61. ISSN: 08905096. DOI: 10.1016/j.avsg.2020.11.019.

[146] Masters, G. N. The analysis of partial credit scoring. Applied Measurement in Education 1(4), 279–297.

[147] Hung, M., Carter, M., Hayden, C., Dzierzon, R., Morales, J., Snow, L., et al. Psychometric assessment of the patient activation measure short form (PAM-13) in rural settings. Quality of Life Research 22(3), 521–529. ISSN: 0962-9343, 1573-2649. DOI: 10.1007/s11136-012-0168-9.

[148] Lightfoot, C. J., Wilkinson, T. J., Memory, K. E., Palmer, J., and Smith, A. C. Reliability and validity of the patient activation measure in kidney disease: results of rasch analysis. Clinical Journal of the American Society of Nephrology: CJASN 16(6), 880.

[149] Blakemore, A., Hann, M., Howells, K., Panagioti, M., Sidaway, M., Reeves, D., et al. Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. BMC Health Services Research 16(1), 582. ISSN: 1472-6963. DOI: 10.1186/s12913-016-1843-2.

[150] Greene, J., Hibbard, J., Sacks, R., Overton, V., and Parrotta, C. When Patient Activation Levels Change, Health Outcomes And Costs Change, Too. Health affairs (Project Hope) 34, 431–7. DOI: 10.1377/hlthaff.2014.0452.

[151] Bowling, A. Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health 27(3), 281–291. ISSN: 1741-3850, 1741-3842. DOI: 10.1093/pubmed/fdi031.

[152] Cook, C. Mode of administration bias. The Journal of Manual & Manipulative Therapy 18(2), 61–63. ISSN: 1066-9817. DOI: 10.1179/106698110X12640740712617.

[153] Leeuw, E. D. d., Hox, J. J., Dillman, D. A., and Methodology, E. A. of, eds. International handbook of survey methodology. EAM book series. New York ; London: Lawrence Erlbaum Associates, 2008. 549 pp. ISBN: 978-0-8058-5753-5 978-0-8058-5752-8.

[154] Costantini, L., Pasquarella, C., Odone, A., Colucci, M. E., Costanza, A., Serafini, G., et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review. Journal of Affective Disorders 279, 473–483. ISSN: 01650327. DOI: 10.1016/j.jad.2020.09.131.

[155] Newland, P., Lorenz, R., and Oliver, B. J. Patient activation in adults with chronic conditions: A systematic review. Journal of Health Psychology 26(1), 103–114. ISSN: 1359-1053. DOI: 10.1177/1359105320947790.

[156] Winship, C. and Mare, R. D. Models for sample selection bias. Annual review of sociology 18(1), 327–350.

Chapter V

Appendix

5.1 Appendix A

5.1.1 Study protocol

Activation in Patients with Chronic Diseases

Study Protocol

Dissertation

Magdalena Holter, MSc.[1]

Supervisors:

Univ.-Prof. Dipl.-Ing. Dr. Andrea Berghold[1]

Univ.-Prof. Dr. Andreas Wedrich[2]

Priv.-Doz. Univ.-Ass. Mag. Dr. Alexander Avian[1]

1: Institute for Medical Informatics, Statistics and Documentation

Medical University of Graz

Auenbruggerplatz 2, 8036 Graz

2: Department of Ophthalmology

Medical University of Graz

Auenbruggerplatz 4, 8036 Graz

Version 2.2 31.10.2019

Content

1 Theoretical Background

The number of patients suffering from a chronic disease is steadily rising worldwide. In Austria it is estimated that 36% of the population suffer from a chronic disease (Statistik Austria, 2014)[1]. With a chronic disease comes greater responsibility, thus it is important that patients can cope with their illness. There are several theoretical models stating different risk and protective behaviors and factors for one's health. In this study the biopsychosocial approach to health will be used to describe different influences (Engel, 1977)[2]. This model states that biological, psychological and social facets and their interaction determine physical health and illness. Known biological risk factors are e.g. age and high bodyweight. Depressiveness and addictive behaviour are examples for psychological risk factors, whereas self-efficacy and health behaviour are psychological resources for health. Social support is a resource regarding social processes. In the social facet, socioeconomic status (SES) is a known risk factor, a low SES predicts a poor health outcome (Lampert et al., 2013)[3]. These risk and resource factors are only limited examples of influences towards our health.

Patients suffering from a chronic disease have to deal with complex treatments and may have to change their lifestyle. They should also be aware when their health status deteriorates and they have to seek help. Therefore, it is important for the patients to have knowledge about their disease and the abilities to manage it. From the personal perspective they need confidence to live with the disease and incorporate it into everyday life. A validated tool is needed to measure how patients suffering from a chronic disease cope with this challenges described. In 2004 Hibbard et al.[4] developed a questionnaire named 'Patient Activation Measure' (PAM). It assesses the knowledge, skills and confidence of patients in managing their own health or chronic disease. The original questionnaire consisted of 22 items which were reduced to 13 items in a later version of the questionnaire called PAM-13 (Hibbard, 2005)[5]. The items are rated on a Likert type scale. A patient can be categorized into one of four levels of patient activation, with level 4 being the highest one, where a healthy lifestyle is maintained even during stressful periods. The reliability and validity of the questionnaire was endorsed by results of classical test theory (CTT) and item response theory (IRT).

The PAM-13 was applied to various patient populations (e.g. mental problems, neurological diseases, multimorbid patients) in English speaking countries. The PAM-13 is mainly used in 3 different settings: (i) to stratify the population and identify risks for diseases, (ii) to align the support to the patient activation and (iii) to measure an impact of an intervention. Regarding the patient population with chronic disease, Donald et al. (2011)[7] showed that if patients are less activated, they visit their main health care provider more often. A study conducted by Mosen et al. (2007)[8] showed that patients suffering from a chronic disease who are more activated tend to use more self-management services, show more self-management behavior and are more compliant in taking their medication. Furthermore they are more satisfied with their care and show more life satisfaction. In addition, they believe in being in control of the management of their own health (Dixon et al., 2009)[9]. Hence, it is important to have a closer look at different patient populations and their activation. Gaining insight about their activation is a way to help patients cope with their disease.

The PAM-13 was also translated into several languages including German (Brenk-Franz, 2013)[6]. However, it was only applied once to validate the German version in a very broad patient population in primary care. The psychometric properties of the German version were analyzed by using classical test theory. A positive association with self-efficacy has been found, which underlines the success of the translation. It may be due to the composition of the sample (or the translation), but the available data show more activated patients in German speaking countries than in the US. However, there also may be cultural differences between countries in Europe and the US or the differences can be due to the health care systems of the respective countries. Therefore, it is possible that patients from different countries suffering from a chronic disease have to face different challenges in maintaining their health status. In Austria and Germany, the concept of patient activation is barely studied yet, leaving questions about cultural differences with other countries to be answered. The German version of the PAM-13 has not been analyzed by item response theory (IRT) yet. By IRT construct validity can be confirmed. Item parameters are independent of the sample, so comparison with results of the PAM-13 in other languages are possible. Moreover, criteria-oriented interpretation of the measured characteristic is possible. Furthermore, interaction between patient activation and risk and resources from biological, psychological and social facets are of interest.

2 Objectives and Research Questions

The goal is to gain further insight how a specific questionnaire, the PAM-13, can be used in a clinical setting in patients with a chronic disease, regarding psychometric properties. Further, some health determinates of different facets will be investigated. Regarding chronic diseases, this study will feature patients with macular oedema. This illness often occurs in the course of diabetes. Affected patients are able to influence their health status via their behaviour. Therefore, these patients are regarded as an appropriate population for this research question.

2.1 Main objective

The psychometric properties (reliability, validity) of the PAM-13 will be investigated using item response theory (Rasch Rating Scale model).

2.2 Secondary objectives

We aim to investigate psychological, social and biological indicators for health behaviour to gain insight about this patient population. Different aspects of each of the three facets will be investigated with the goal to regard them and their interplay in one big picture. Known risk factors for macular oedemas included in this study will be age, smoking habits, body mass index (BMI), diabetes, HbA1c, previous cataract surgery, hypertonia, and cholesterol. Regarding the psychological facet, patient activation, general mood, general quality of life, motivations for information seeking and self-efficacy will be included. Social support, marital status and SES are aspects of social resources and risk factors to be assessed. Further, self-reported health status will be regarded and included.

Besides these already listed methodological aspects, we will describe the distribution of patient activation by the PAM-13 from a clinical point of view and deduce the needs of the patients in order to handle their disease in a high quality manner.

3 Methods

3.1 Study design

This is a questionnaire-based prospective cross-sectional study. It will consist of an ad-hoc sample from the population of outpatients with macular oedema from the Department of Ophthalmology of Graz. The aim is to include 700 patients. After receiving informed consent each patient included will be asked to complete the questionnaires once. This will take approximately half an hour for each participant. In general, the study is set to run for one year. Before the start of the study in autumn 2019, a pre-testing will take place to ensure feasibility.

3.1.1 Pre-Testing

To assess the duration of completing all questionnaires while they are read out loud, time will be measured in a test setting with people who are similar to the patient population regarding age. This procedure will be carried out with ten people. They will not be included in the main study.

3.2 Study population

Patients with macular edema will be included in this study. Since this is a heterogeneous disease with different etiologies, we will focus on two different types. These are diabetic macula edema and edema due to retinal vein occlusion. They both have similar risk factors and both are due to vascular diseases. However, diabetic macular edema occurs slowly, bilateral and subacute, whereas macular edema by retinal vein occlusion appears more quickly and monocular. Patients with retinal vein occlusion will be included shortly after the diagnosis. Diabetic patients are included either right after the diagnosis or 3 to 6 months after start of the treatment. Regarding age, the main population is expected to be over 60 years, since the illness usually occurs later in life.

An outpatient clinic for patients with a macula edema due to diabetes is held two times per week and one time per week for patients with a macula edema due to retinal vein occlusion at the Department of Ophthalmology. These patients will be approached about participating in the study.

Inclusion criteria:

- Willingness to participate in the study (signed informed consent)
- Macular edema due to diabetes or retinal vein occlusion
- Good German language skills (able to complete questionnaires)
- Adult patient (at least 18 years)

Exclusion criteria:

- Macular edema due to a different reason (not diabetes or retinal vein occlusion)
- Hearing difficulty
- Cognitive impairment
- Part of pre-testing

3.3 Sample size

The main objective of this study is to analyse the PAM-13 by IRT with different IRT models, including Rasch models e.g. the Rating Scale Model. A sample size calculation or power calculation for testing the model fit of a Rasch model is not feasible, because they use chi-squared distributed test-statistics, whose degrees of freedom are independent of the sample size (Kubinger, 2009)[10]. In general, sample size depends on the chosen IRT model: an increasing sample size is required with more complexity of the model. Further, the more response options the questionnaire provides, the more participants are required, as more item parameters must be estimated. A minimum sample size of 500 is recommended for a two-parameter model (Reeve & Fayers, 2005)[11]. The purpose of the analysis is influencing the required sample size as well: The aim of this study is to obtain precise and stable estimations on item characteristics and person scores. For this purpose a sample size of at least 500 patients is recommended (Cappelleri, 2014)[12]. Also, more items to be analysed require a bigger sample size (Cappeleri, 2014)[12].

There is one suggestion to calculate the power of a Rasch model via differential item functioning (DIF), assuming that there is a criterion dividing the participants into at least two groups (Kubinger, 2009)[10]. DIF describes how an item works in different subgroups and if it measures the same abilities for members of those subgroups. Simulations were done which showed that with decreasing number of items and increasing number of participant's, power rises and a DIF can be detected more easily. For example if the questionnaire consists of 20 items and there are two groups with the same mean but the same DIF for two items, with a sample size of 50 a power of 43% would be achieved, whereas with a sample size of 100 a power of 79% would be reached. No simulations with a higher number of participants were done, which would be more adequate for the underlying study, but it is clearly visible that with a larger sample size the power to detect the same DIF increases.

Regarding DIF and the PAM-13, DIF was found in one of the two German versions (Zill et al. 2013)[13], for sex and age (younger or older than 65), in 3 and 5 items, respectively. This DIF was detected in a big sample of 4300 participants.

Since the PAM-13 consists of 13 Items with 4 response options and item characteristics and person scores are to be estimated, at least 500 patients should be included in this study. Since DIF was previously found in a large sample, and we plan to analyse several subgroups (chronic disease of the patient, mode of completing the questionnaire, sex, age), we intend to include a high number

(N=700) of participants to achieve and adequate power to detect possible DIF in the official German version. In Graz there are approximately 1200 patients with a macula oedema due to diabetes or retinal vein occlusion each year. The aim to include 700 patients in one year seems possible.

3.4 Outcome measures

3.4.1 Patient Activation Measure (PAM -13)

The PAM-13 assesses the knowledge, skills and confidence of patients in managing their own health or chronic disease. The items are rated on a Likert type scale with four response categories from 'disagree' to 'agree strongly'. The answers to the 13 items are summed up and transformed on a scale between 0 – 100 (Hibbard et al., 2005)[5]. The German version of the PAM-13 showed sufficient reliability through internally consistency with a Cronbach's alpha of 0.84. Indications for validity was given by factorial structure and a trait-trait correlation of r=0.43 between the score of the PAM-13 and general self-efficacy (Brenk-Franz et al., 2013)[6].

The PAM-13 is licensed by Insignia Health, a company based in Portland, Oregon, USA, www.insigniahealth.com. One condition to use the questionnaire without a fee for academic purpose is to forward them study relevant data in anonymized form (for more details see 'Data handling'). Insignia Health states various other criteria as well, some of them are: the PAM-13 is not to be used in a commercial way, a sample size of at least 75 patients is required, and the plan to publish the results. By applying for a PAM-13 research license we grant Insignia Health to use our data for product improvement and validation efforts.

3.4.2 Fragebogen zur Sozialen Unterstützung (F-Sozu)

To measure self-perceived social support there are three different forms of the F-Sozu, which differ in the number of items. The 22 Item version was chosen for this study. They are rated on a Likert type style scale ranging in five steps from 'disagree' to 'agree strongly'. It is possible to build one general score ('social support') but also to create three subscales (emotional support, practical assistance and social integration). Sufficient reliability of the 22 item version was found in earlier Studies (Cronbach's α = 0.91 for the general score, Cronbach's α = 0.72 – 0.87 for the subscales). Validity for the F-Sozu is given through the factorial structure. In a factor analysis one factor 'social support' was found. Further validity is given through correlations with external criteria, such as self-confidence and life satisfaction. The subscales of the 22 item version differ in the size of the correlation with these external criteria, providing more information about social support than the general score (Sommer & Fydrich, 1999)[14].

3.4.3 Health Status

Current Health Status is assessed by the question 'How is your health in general?' with 5 answer options from 'very good' to 'very poor'. Further, one item out of the EQ-5D-5L (Herdman et al., 2011) is used, where the health status on this particular day is determined. A rating is made on a scale from 0 (worst health you can imagine) – 100 (best health you can imagine) and the intended number is written in a box. Reliability, measured by Cronbach's alpha, is α =0.86 for the scale. High correlations with other health measures provide convergent validity for the score.

3.4.4 Habituelle subjektive Wohlbefindensskala (HSWBS)

This scale measures general mood by six items and general quality of life by seven items. The resulting 13 items are judged on a Likert type scale with 6 response categories ranging from 'agree exactly' to 'agree not at all'. For each of the two scales a sum score is built (Dalbert, 1992)[15]. Internal consistency is $\alpha = .83$ for the scale general mood and $\alpha = .87$ for the scale general quality of life. One factor for the scale general mood was found, indicating factorial validity. Further factorial validity was given through a model, containing the factor 'habitual subjective well-being', which was built by the two subscales. Strong positive correlations with the life-satisfaction scale of the FPI-R show construct validity.

3.4.5 Goals Associated with Health Information Seeking (GAINS)

This is a recently developed questionnaire (Chasiotis et al., 2018)[16] about one's motivation to seek information about health topics. It consists of 16 items which are rated on a Likert type style scale ranging in five steps from 'totally disagree' to 'fully agree'. The items are summed up to form 4 scales: 'understanding', 'calming', 'action planning' and 'hope'. All items are summed up to gain a score for the general information seeking requirement. Since the questionnaire is newly developed, no data on psychometric criteria like reliability and validity are available.

3.4.6 Skala zur allgemeinen Selbstwirksamkeitserwartung (SWE)

This questionnaire assesses subjective belief to successfully cope with new demanding situations by own strength. General self-efficacy is measured by 10 items which are judged on a Likert type scale with four response categories ranging from 'disagree' to 'agree' (Jerusalem & Schwarzer, 1999)[17]. In several German samples a Cronbach's α between 80 and 90 was found. The scale is unidimensional. More validity is given by correlations of self-efficacy with various other constructs, such as negative correlations with depression, anxiety and burnout.

3.4.7 Demographic data

As stated before, this study will be questionnaire-based. However, to gain more information about the patients and to keep the duration of the questioning short, a few factors will be looked up in the 'EyeMed' system. This is an online system used by the Department of Ophthalmology of Graz for documenting the medical history of patients. Following information will be gathered:

- Date of diagnoses of the macular edema
- Diagnoses details?
- Onset of therapy for macular edema
- Blood parameters (same as questioned in demographic data, if they fit time wise, not older than 3 months)

Demographic data of interest are shown in detail in Table 1. Demographic Data. The answer options for some of demographic data are taken from Statistik Austria and are marked with an asterisk (*).

Table 1. Demographic Data.

Nr.	Notation	Details
1	Sex	Male, female, other
2	Age	In years
3	Marital status	Married, solid partnership, single, divorced, widowed
4	Education*	6 possibilities matching the International Standard Classification of Education: Kein Schulabschluss; Pflichtschulabschluss; (Lehre, BMS, Polytechnische Schule); Matura (AHS, BHS); (Kolleg, Akademie); (Fachhochschulabschluss Hochschulabschluss)
5	Job status	Employed, worker, civil servant, housewife/-man, self-employed, studying, retiree, jobless
6	Weight	In kg
7	Height	In cm
8	Diabetes	Yes, no, choose: Type I or Type II; Diagnosis since __ years; part of the Disease management program

9	Smoking	Yes, former-smoker, never, how many years do/did you smoke? Non-smoker since __ years
10	Hypertonia	Do you take medication regulating blood pressure? Yes, no
11	HbA1C	In mmol/mol
12	Cholesterol	In mg/dl
13	Previous cataract surgery	Yes, no
14	Other illness	Yes, no
15	Completion	Questionnaires were: Self-filled, read out loud
16	Income*	<800€, up to 1.125€, up to 1.500€, up to inclusive 1.950€, over 1.950€

3.4.8 Procedure

Table 2 gives an overview of currently planned measures to be used in the order to be presented, which facet they belong to regarding the biopsychosocial model and the estimated time to complete the questionnaire.

Table 2. Outcome measures.

Nr.	Name	Facet	Estimated duration (in min.)
1	Patient Activation Measure (PAM-13)	Psychological	4
2	F-Sozu K-22	Social	5
3	Health Status	Biological	1
4	Habituelle subjektive Wohlbefindensskala (HSWBS)	Psychological	5
5	Goals Associated with Health Information Seeking (GAINS)	Psychological	5
6	Skala zur allgemeinen Selbstwirksamkeitserwartung (SWE)	Psychological	3
7	Demographic data	Biological Social	3
			26

3.5 Data collection

Questionnaires are read out loud if patients can't read them on their own. Data collection will be made by trained personnel, trained students and the person mainly responsible for this study. Training will be guided by MH, it will include practicing the test procedure and handling of study materials. A standard procedure will be established to guarantee the reliability and the comparability of each testing. Data collection will take place in the ambulatory of the Department of Ophthalmology, while patients wait in-between examinations.

Medical students about to take their diploma have the possibility to apply for this project. They will use part of the data for their master thesis. However, they will investigate different research questions which can be answered with descriptive statistics.

3.6 Data handling

Data will be pseudonymized. Each participant will get a code. This will be done by the person who enters the patient into study. A web-based pseudonymization tool called 'iPSN' will be used for that. This tool was recently developed at the institute for Medical Informatics, Statistics and Documentation and ensures safe data handling. Only authorized persons will have access to the translation list.

A survey on LimeSurvey will be prepared before the start of the study, containing the items of the questionnaires and the demographic data. If the questions are read out loud, the interviewer will check the answers on a tablet on LimeSurvey. No names will be entered in the online system. If the patients want to fill out the questionnaires on their own, they can choose between paper and pencil or the tablet. If they choose the paper and pencil procedure, the data will be entered into LimeSurvey by students and the person mainly responsible for this study. Before data entry, a training will take place to ensure quality of data entries.

LimeSurvey will be hosted on a server of the Institute for Medical Informatics, Statistics and Documentation of the Medical University of Graz. Data will be stored there as well.

Participants will be informed that data concerning the PAM-13 and study relevant data will be forwarded anonymized to Insignia Health. This includes questionnaire based data collected during the study as well as demographics of the participants. Age will be forwarded as a number, no birth dates will be send. Each participant will be presented through a random number. As mentioned before, forwarding these data is one condition to use the questionnaire without a fee for academic purpose.

3.7 Data analysis

Plausibility check will be performed before data analysis. Categorical data will be presented as absolute and relative frequencies, and continuous data as means and standard deviations or medians and interquartile ranges, as appropriate. To analyze the research questions, different procedures will be used:

1. Item Response Theory (IRT) to evaluate the German version of the PAM-13.
2. Structural equation model to investigate the relationship between patient activation and indicators of the three facets for health behaviour of the biopsychosocial model.

The primary objective of this study is to analyse the PAM-13 by IRT. There are several IRT models for analysing questionnaires. The Rasch Rating scale model will be used to calculate item and person statistics, because the same model was used by the developers of the PAM-13 (Hibbard, 2005) and in further studies evaluating the questionnaire. One-dimensionality of the scale will be investigated by this model as well. Further, the Partial credit model will be applied on the data of the PAM-13, an extension of the original Rasch model for ordinal data. Before applying the models on the data, assumptions like local stochastic independence will be investigated. For both models, fits will be evaluated.

In addition, DIF will be assessed by estimating item parameters for groups of participants. Particularly, groups of patients with a different chronic disease and the mode of completing the questionnaires are of interest (read out loud, vs. paper and pencil vs. tablet). Moreover, subgroups of sex and age (younger vs. older participants) will be examined, since DIF in those groups were found in previous studies. DIF will be investigated from the model e.g. by using Likelihood Ratio Test and graphical analysis.

The secondary objective will be investigated by a structural equation model. It will be based on theoretical background from current literature about health resources, risks and behaviours. Indicators for each of the three facets (biological, psychological and social) will build one factor each. The psychological facet will be built by a positive influence on health behaviour by patient activation, self-efficacy, quality of life and motivation for health information seeking and a negative influence of a low mood. Regarding the social facet, a resource is well perceived social support, marital status and a risk factor low socioeconomic status. Health behaviour is negatively influenced by the biological facet through older age, being a smoker, overweight, suffering from a manifest disease like diabetes, previous cataract surgery, hypertonia, high blood sugar level (HbA1C) and high cholesterol. A positive influence towards health behaviour of the biological facet is good self-rated health. Individual influences of the indicators on the factor will be evaluated. Interactions between the facets will be investigated. Model fits will be investigated and alternative models will be regarded to reach a high model fit.

Statistical analysis will be performed by SAS software (version 9.4; SAS Institute, Inc., Cary, NC, USA) and R software (version 3.5.0 or higher) probably using the packages lordif[18], ltm[19], and mirt[20].

4 Ethical considerations

Patients have to give written informed consent to participate in the study. Participants can retract their consent at any time, as long as data have not been analyzed. Data will be handled carefully. Only authorized people will have access. Data will be entered pseudonymized into LimeSurvey. As mentioned before, only required data will be send to Insignia Health, however, fully anonymized. Ethical approval will be obtained from the ethic committee of the Medical University of Graz before start of the study.

5 Resources

Most of the questionnaires included in this study can be used without a fee. For the PAM-13 there is an administrational charge. The F-Sozu is distributed by Hogrefe, they provide a discount for researchers. The costs for the F-Sozu listed in the following table is estimated, but should be an upper limit. Using LimeSurvey will minimize copying costs. Estimated costs are displayed in Table 3. Estimated Costs.. The institute for Medical Informatics, Statistics and Documentation will fund costs for the questionnaires.

Table 3. Estimated Costs.

Questionnaire	Costs in euro
PAM-13	70
F-Sozu	350
Sum	420

6 Abbrevations

SES: Socioeconomic Status

PAM: Patient Activation Measure

BMI: Body Mass Index

DIF: Differential Item Functioning

F-Sozu: Fragebogen zur Sozialen Unterstützung

HSWBS: Habituelle subjektive Wohlbefindensskala

GAINS: Goals Associated with Health Information Seeking

SWE: Skala zur allgemeinen Selbstwirksamkeitserwartung

CTT: Classical Test Theory

IRT: Item Response Theory

7 References

1. Statistik austria, gesundheitsbefragung 2014. https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/gesundheitszustand/chronische_krankheiten/index.html. Updated 18.10.2016. Accessed 11/16, 2018.

2. Engel GL. The need for a new medical model: A challenge for biomedicine. *Science*. 1977;196(4286):129-136.

3. Lampert T, Kroll LE, von der Lippe E, Müters S, Stolzenberg H. Sozioökonomischer status und gesundheit. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*. 2013;56(5-6):814-821.

4. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Serv Res*. 2004;39(4p1):1005-1026.

5. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. *Health Serv Res*. 2005;40(6p1):1918-1930.

6. Brenk-Franz, K., Hibbard, J. H., Herrmann, W. J., Freund, T., Szecsenyi, J., Djalali, S., Schneider, N. . Validation of the german version of the patient activation measure 13 (PAM13). . 2013.

7. Donald M, Ware RS, Ozolins IZ, Begum N, Crowther R, Bain C. The role of patient activation in frequent attendance at primary care: A population-based study of people with chronic disease. *Patient Educ Couns*. 2011;83(2):217-221.

8. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is patient activation associated with outcomes of care for adults with chronic conditions? *J Ambulatory Care Manage*. 2007;30(1):21-29.

9. Dixon A, Hibbard J, Tusler M. How do people with different levels of activation self-manage their chronic conditions? *The Patient: Patient-Centered Outcomes Research*. 2009;2(4):257-268.

10. Kubinger KD, Rasch D, Yanagida T. On designing data-sampling for rasch model calibrating an achievement test. *Psychological Test and Assessment Modeling*. 2009;51(4):370.

11. Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. *Assessing quality of life in clinical trials: methods of practice*. 2005;2:55-73.

12. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther*. 2014;36(5):648-662.

13. Zill JM, Dwinger S, Kriston L, Rohenkohl A, Harter M, Dirmaier J. Psychometric evaluation of the german version of the patient activation measure (PAM13). *BMC Public Health*. 2013;13:1027-2458-13-1027. doi: 10.1186/1471-2458-13-1027 [doi].

14. Fydrich T, Geyer M, Hessel A, Sommer G, Brähler E. Fragebogen zur sozialen unterstützung (F-SozU): Normierung an einer repräsentativen stichprobe. *Diagnostica*. 1999;45(4):212-216.

15. Dalbert C. HSWBS-habituelle subjektive wohlbefindensskala. . 1992.

16. Chasiotis A, Wedderhoff O, Rosman T, Mayer A. GAINS-goals associated with health information seeking. . 2018.

17. Schwarzer R, Jerusalem M. Skalen zur erfassung von lehrer-und schülermerkmalen. *Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen.Berlin*. 1999.

18. Choi SW, Gibbons LE, Crane PK. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *J Stat Softw*. 2011;39(8):1-30.

19. Rizopoulos D. Ltm: An R package for latent variable modeling and item response theory analyses. *Journal of statistical software*. 2006;17(5):1-25.

20. Chalmers P. Multidimensional item response theory [R package mirt version 1.24]. .

8 Appendix

8.1 Questionnaires

Instrument zur Erfassung der aktiven Patientenbeteiligung (PAM)

Nachstehend finden Sie einige Aussagen, die Leute manchmal machen, wenn sie sich über ihre Gesundheit unterhalten. Bitte geben Sie an, wie stark diese Aussagen für Sie zutreffen.

		stimmt genau	**stimmt eher**	**stimmt kaum**	**stimmt nicht**
1	Letzten Endes bin ich derjenige, der dafür verantwortlich ist, für meine Gesundheit zu sorgen.	☐	☐	☐	☐
2	Das Wichtigste für meine Gesundheit ist eine aktive Rolle in meiner Gesundheitsversorgung zu übernehmen.	☐	☐	☐	☐
3	Ich bin überzeugt, dass ich selbst etwas unternehmen kann, um Krankheiten vorzubeugen.	☐	☐	☐	☐
4	Ich weiß bei jedem meiner Medikamente weshalb ich es nehme.	☐	☐	☐	☐
5	Ich bin überzeugt, zu wissen, wann ich zum Arzt gehen muss und wann ich ein Gesundheitsproblem selbst behandeln kann.	☐	☐	☐	☐
6	Ich bin überzeugt, dass ich meinem Hausarzt meine Sorgen mitteilen kann, auch wenn er mich nicht direkt darauf anspricht.	☐	☐	☐	☐
7	Ich bin überzeugt, dass ich die zu Hause notwendigen medizinischen Behandlungen selbst durchführen kann.	☐	☐	☐	☐
8	Ich kenne die Ursachen meiner Beschwerden.	☐	☐	☐	☐
9	Ich kenne verschiedene Behandlungsoptionen für meine Erkrankungen	☐	☐	☐	☐
10	Ich war bisher in der Lage, Veränderungen meiner Lebensgewohnheiten − wie gesunde Ernährung und Bewegung − aufrechtzuerhalten	☐	☐	☐	☐
11	Ich weiß, wie ich einer Verschlechterung meiner Gesundheit vorbeugen kann.	☐	☐	☐	☐
12	Ich bin überzeugt, Lösungen zu finden, wenn sich meine Gesundheit verschlechtert.	☐	☐	☐	☐
13	Ich bin überzeugt, dass ich Veränderungen meiner Lebensgewohnheiten - wie Diät und körperliche Bewegung - auch in stressigen Zeiten fortführen kann	☐	☐	☐	☐

Fragebogen zur Sozialen Unterstützung (F-Sozu)

Anleitung: In diesem Fragebogen geht es um Ihre Beziehungen zu wichtigen Menschen, also zum Partner, zu Familienmitgliedern, Freunden und Bekannten, Kollegen und Nachbarn. Wir möchten erfahren, wie Sie diese Beziehungen erleben und einschätzen.
Bitte kreuzen Sie an, inwieweit die jeweilige Feststellung auf Sie zutrifft. Falls Sie genau zutrifft, kreuzen Sie bitte die 5 an, falls sie überhaupt nicht zutrifft, kreuzen Sie bitte die 1 an; und falls Ihre Antwort dazwischenliegt, kreuzen Sie die 2, 3 oder 4 an, je nachdem, wie sehr Sie der Feststellung zustimmen.
Wenn in den folgenden Aussagen allgemein von „Menschen" oder von „Freunden / Angehörigen" die Rede ist, dann sind **die Menschen gemeint, die für Sie wichtig sind**.

		Trifft überhaupt nicht zu				**Trifft genau zu**
1	Ich habe Menschen, die sich um meine Wohnung (Blumen, Haustiere) kümmern können, wenn ich mal nicht da bin.	☐	☐	☐	☐	☐
2	Es gibt Menschen, die mich so nehmen, wie ich bin.	☐	☐	☐	☐	☐
3	Meinen Freunden / Angehörigen ist es wichtig, meine Meinung zu bestimmten Dingen zu erfahren.	☐	☐	☐	☐	☐
4	Ich wünsche mir von anderen mehr Verständnis und Zuwendung.	☐	☐	☐	☐	☐
5	Ich kenne einen sehr vertrauten Menschen, mit dessen Hilfe ich in jedem Fall rechnen kann.	☐	☐	☐	☐	☐
6	Bei Bedarf kann ich mir Werkzeug oder Lebensmittel ausleihen.	☐	☐	☐	☐	☐
7	Ich habe Freunde / Angehörige, die auch mal gut zuhören können, wenn ich mich aussprechen möchte.	☐	☐	☐	☐	☐
8	Ich kenne fast niemanden, mit dem ich gerne ausgehe.	☐	☐	☐	☐	☐
9	Ich habe Freunde / Angehörige, die mich auch einfach mal umarmen.	☐	☐	☐	☐	☐
10	Wenn ich krank bin, kann ich ohne Zögern Freunde / Angehörige bitten, wichtige Dinge (z.B. Einkaufen) für mich zu erledigen.	☐	☐	☐	☐	☐
11	Wenn ich mal tief bedrückt bin, weiß ich, zu wem ich gehen kann.	☐	☐	☐	☐	☐
12	Ich fühle mich oft als Außenseiter.	☐	☐	☐	☐	☐

		Trifft überhaupt nicht zu				**Trifft genau zu**
13	Es gibt Menschen, die Leid und Freude mit mir teilen.	☐	☐	☐	☐	☐
14	Bei manchen Freunden / Angehörigen kann ich auch mal ganz ausgelassen sein.	☐	☐	☐	☐	☐
15	Ich habe einen vertrauten Menschen, in dessen Nähe ich mich sehr wohl fühle.	☐	☐	☐	☐	☐
16	Ich habe genug Menschen, die mir wirklich helfen, wenn ich mal nicht weiter weiß.	☐	☐	☐	☐	☐
17	Es gibt Menschen, die zu mir halten, auch wenn ich Fehler mache.	☐	☐	☐	☐	☐
18	Ich wünsche mir mehr Geborgenheit und Nähe.	☐	☐	☐	☐	☐
19	Es gibt genug Menschen, zu denen ich ein wirklich gutes Verhältnis habe.	☐	☐	☐	☐	☐
20	Es gibt eine Gemeinschaft von Menschen (Freundeskreis, Clique), zu der ich mich zugehörig fühle.	☐	☐	☐	☐	☐
21	Durch meinen Freundes- und Bekanntenkreis erhalte ich oft gute Tipps (z.B. guter Arzt, wichtige Informationen).	☐	☐	☐	☐	☐
22	Es gibt Menschen, denen ich alle meine Gefühle zeigen kann, ohne dass es peinlich wird.	☐	☐	☐	☐	☐

		Sehr gut	**Gut**	**Mittel-mäßig**	**Schlecht**	**Sehr schlecht**
1	Wie ist Ihre Gesundheit im Allgemeinen?	☐	☐	☐	☐	☐

Wir wollen herausfinden, wie gut oder schlecht Ihre Gesundheit HEUTE ist.

Diese Skala ist mit Zahlen von 0 bis 100 versehen.

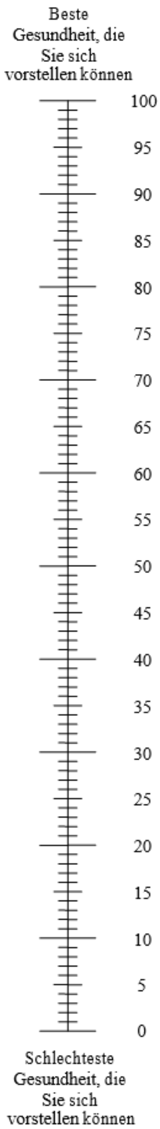
100 ist die beste Gesundheit, die Sie sich vorstellen können.

0 (Null) ist die schlechteste Gesundheit, die Sie sich vorstellen können.

Bitte kreuzen Sie den Punkt auf der Skala an, der Ihre Gesundheit HEUTE am besten beschreibt.

Jetzt tragen Sie bitte die Zahl, die Sie auf der Skala angekreuzt haben, in das Kästchen unten ein.

IHRE GESUNDHEIT HEUTE = ☐



ALLGEMEINES BEFINDEN

Bitte beurteilen Sie im Folgenden, wie es Ihnen **im Allgemeinen** geht und wie zufrieden Sie im Allgemeinen mit Ihrem Leben sind. Nehmen Sie bitte zu jeder der folgenden Aussagen Stellung und entscheiden Sie, inwieweit jede auf Sie ganz persönlich zutrifft. Kreuzen Sie dazu bitte jeweils eine Zahl zwischen 1 und 6 an, und lassen Sie bitte keine Aussage aus.

		stimmt genau	**stimmt weitgehend**	**stimmt ein wenig**	**Stimmt eher nicht**	**Stimmt weitgehend nicht**	**Stimmt überhaupt nicht**
1	Mein Leben könnte kaum glücklicher sein, als es ist.	☐	☐	☐	☐	☐	☐
2	Ich fühle mich meist ziemlich fröhlich.	☐	☐	☐	☐	☐	☐
3	Ich glaube, dass sich vieles erfüllen wird, was ich mir für mich erhoffe.	☐	☐	☐	☐	☐	☐
4	Wenn ich an mein bisheriges Leben zurückdenke, so habe ich viel von dem erreicht, was ich erstrebe.	☐	☐	☐	☐	☐	☐
5	Ich halte mich für eine glückliche Person.	☐	☐	☐	☐	☐	☐
6	Ich bin mit meinem Leben zufrieden.	☐	☐	☐	☐	☐	☐
7	Ich glaube, dass mir die Zeit noch einige interessante und erfreuliche Dinge bringen wird.	☐	☐	☐	☐	☐	☐
8	Ich bin nicht so fröhlich wie die meisten Menschen.	☐	☐	☐	☐	☐	☐
9	Ich bin mit meiner Lebenssituation zufrieden.	☐	☐	☐	☐	☐	☐
10	Ich bin selten in wirklicher Hochstimmung.	☐	☐	☐	☐	☐	☐
11	Ich sehe im allgemeinen mehr die Sonnenseiten des Lebens.	☐	☐	☐	☐	☐	☐
12	Wenn ich so auf mein bisheriges Leben zurückblicke, bin ich zufrieden.	☐	☐	☐	☐	☐	☐
13	Ich fühle mich meist so, als ob ich vor Freude übersprudeln würde.	☐	☐	☐	☐	☐	☐

GAINS
Goals Associated with Health Information Seeking

Chasiotis, A., Wedderhoff, O., Rosman, T. & Mayer, A.-K. (2018)

Im Alltag kann es dazu kommen, dass wir uns mit Situationen konfrontiert sehen, in denen wir uns durch das Auftreten und Erleben von Symptomen oder ärztlichen Diagnosen unsicher fühlen. Diese Situationen können es notwendig machen, sich über ein bestimmtes gesundheitliches Thema zu informieren und sich Informationen aus verschiedenen Informationsquellen (z.B. Bücher, Zeitschriften, Internet, Experten etc.) zu beschaffen, um eine Antwort auf eine konkrete gesundheitliche Frage zu finden.

Wir interessieren uns für die Ziele, die Menschen haben, wenn sie Informationen zu gesundheitlichen Problemen suchen. Bitte geben Sie an, welche Ziele Sie üblicherweise haben, wenn Sie sich wegen eines gesundheitlichen Problems informieren möchten.

Im Folgenden finden Sie dazu eine Reihe von Aussagen, die sich auf Ihre Ziele bei einer Informationssuche beziehen. Bitte kreuzen Sie diejenige Aussage an, mit der Sie am ehesten übereinstimmen. Hierzu können Sie unter fünf Antwortalternativen wählen, die von „trifft überhaupt nicht zu" bis hin zu „trifft voll und ganz zu" reichen.

Es gibt keine richtigen oder falschen Antworten; es ist ausschließlich Ihre Selbsteinschätzung von Interesse. Bitte bearbeiten Sie alle Aussagen in der vorgegebenen Reihenfolge und lassen Sie keine Aussage aus.

	Ich benötige Informationen, um …	**Stimmt überhaupt nicht**	**Stimmt eher nicht**	**Stimmt teilweise**	**Stimmt eher**	**Stimmt völlig**
1	… etwas über mögliche Ursachen für das gesundheitliche Problem herauszufinden.	☐	☐	☐	☐	☐
2	… den medizinischen Hintergrund des gesundheitlichen Problems besser zu verstehen.	☐	☐	☐	☐	☐
3	… das beschriebene gesundheitliche Problem in vollem Umfang zu erfassen.	☐	☐	☐	☐	☐
4	… etwas über mögliche Folgen des gesundheitlichen Problems zu erfahren.	☐	☐	☐	☐	☐
5	… mich in Bezug auf das gesundheitliche Problem zu beruhigen	☐	☐	☐	☐	☐
6	… in Bezug auf das gesundheitliche Problem gelassener zu werden.	☐	☐	☐	☐	☐

	Ich benötige Informationen, um …	Stimmt überhaupt nicht	Stimmt eher nicht	Stimmt teilweise	Stimmt eher	Stimmt völlig
7	… in Bezug auf das gesundheitliche Problem weniger Angst zu haben.	☐	☐	☐	☐	☐
8	… mich abzusichern, dass das beschriebene gesundheitliche Problem harmlos ist.	☐	☐	☐	☐	☐
9	… Behandlungsmöglichkeiten in Bezug auf das gesundheitliche Problem zu finden.	☐	☐	☐	☐	☐
10	… herauszufinden, was ich selbst zur Bewältigung dieses gesundheitlichen Problems unternehmen kann.	☐	☐	☐	☐	☐
11	… mein weiteres Vorgehen in Bezug auf die Bewältigung des gesundheitlichen Problems planen zu können.	☐	☐	☐	☐	☐
12	… Möglichkeiten zu finden, Einfluss auf das Problem zu nehmen.	☐	☐	☐	☐	☐
13	… mich trotz des Problems wieder gut zu fühlen.	☐	☐	☐	☐	☐
14	… mich trotz des Problems wieder freuen zu können.	☐	☐	☐	☐	☐
15	… trotz des Problems wieder Mut zu fassen.	☐	☐	☐	☐	☐
16	… trotz des Problems zuversichtlich in die Zukunft schauen zu können.	☐	☐	☐	☐	☐

Selbstwirksamkeit

Nachstehend finden Sie einige Aussagen. Bitte geben Sie an, wie stark diese Aussagen für Sie zutreffen.

		stimmt nicht	stimmt kaum	stimmt eher	stimmt genau
1	Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.	☐	☐	☐	☐
2	Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.	☐	☐	☐	☐
3	Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.	☐	☐	☐	☐
4	In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.	☐	☐	☐	☐
5	Auch bei überraschenden Ereignissen glaube ich, dass ich gut mit ihnen zurechtkommen kann.	☐	☐	☐	☐
6	Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.	☐	☐	☐	☐
7	Was auch immer passiert, ich werde schon klarkommen.	☐	☐	☐	☐
8	Für jedes Problem kann ich eine Lösung finden.	☐	☐	☐	☐
9	Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.	☐	☐	☐	☐
10	Wenn ein Problem auftaucht, kann ich es aus eigener Kraft meistern.	☐	☐	☐	☐

Demografische Daten

1	**Geschlecht**	☐ Männlich ☐ Weiblich ☐ Sonstiges	
2	**Alter**	__ Jahre	
3	**Familienstand**	☐ Verheiratet ☐ Feste Partnerschaft ☐ Ledig ☐ Geschieden ☐ Verwitwet	
4	**Ausbildung**	☐ Kein Schulabschluss ☐ Pflichtschulabschluss (Lehre, BMS, Polytechnische Schule) ☐ Matura (AHS, BHS) ☐ Kolleg/Akademie ☐ Fachhochschulabschluss ☐ Hochschulabschluss	
5	**Arbeitsstatus**	☐ Angestellte/r ☐ ArbeiterIn ☐ BeamterIn ☐ Selbstständige/r	☐ Hausfrau/mann ☐ Studierender ☐ PensionistIn ☐ Arbeitssuchende/r
6	**Gewicht**	__ kg	
7	**Größe**	__ cm	
8	**Diabetes**	☐ Ja ☐ Nein Wenn Ja: ☐ Typ I ☐ Typ 2 Diagnose seit __ Jahren. Sind Sie Teil des Disease Management Programmes? ☐ Ja ☐ Nein	
9	**Raucher**	☐ Ja ☐ ehemaliger Raucher ☐nie geraucht Wie viele Jahre rauchen Sie bzw. haben Sie geraucht? __ Jahre Nichtraucher seit __ Jahren	
10	Nehmen Sie Medikamente um den **Blutdruck** zu regulieren?		☐ Ja ☐ Nein
11	**HbA1c**	__ mmol/mol	
12	**Cholesterin**	__ mg/del	
13	Hatten Sie schon eine **Kataraktoperation**?		☐ Ja ☐ Nein
14	Leiden Sie noch unter einer anderen manifesten Krankheit?		☐ Ja ☐ Nein
15	**Einkommen**	☐ <800€ ☐ bis zu 1.125€ ☐ bis zu 1.500€ ☐ bis inklusive 1.950€ ☐ über 1.950€	
16	**Ausfüllmethode**	☐ Selbst ausgefüllt ☐ Vorgelesen ☐ Papier ☐ Tablet	

5.1.2 Informed consent

PatientInneninformation[1] und Einwilligungserklärung zur Teilnahme an der Studie

Patientenaktivierung bei chronisch kranken PatientInnen

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer!

Wir laden Sie ein, an der oben genannten Studie teilzunehmen. Die Aufklärung darüber erfolgt in einem ausführlichen Gespräch.

Ihre Teilnahme an dieser Studie erfolgt freiwillig. Sie können jederzeit ohne Angabe von Gründen aus der Studie ausscheiden. Die Ablehnung der Teilnahme oder ein vorzeitiges Ausscheiden aus dieser Studie hat keine nachteiligen Folgen für Ihre medizinische Betreuung.

Unverzichtbare Voraussetzung für die Durchführung einer Studie ist jedoch, dass Sie Ihr Einverständnis zur Teilnahme an dieser Studie schriftlich erklären. Bitte lesen Sie den folgenden Text sorgfältig durch und zögern Sie nicht Fragen zu stellen.

Bitte unterschreiben Sie die Einwilligungserklärung nur

- wenn Sie Art und Ablauf der Studie vollständig verstanden haben,
- wenn Sie bereit sind, der Teilnahme zuzustimmen und
- wenn Sie sich über Ihre Rechte als Teilnehmer an dieser Studie im Klaren sind.

Zu dieser Studie, sowie zur Patienteninformation und Einwilligungserklärung wurde von der zuständigen Ethikkommission eine befürwortende Stellungnahme abgegeben.

1. Was ist der Zweck dieser Studie?

Der Zweck dieser Studie ist es, den Fragebogen zur Patientenaktivierung zu evaluieren. Außerdem wird das Gesundheitsverhalten betrachtet.

Diese Studie wird im Auftrag der Medizinischen Universität Graz durchgeführt.

2. Wie läuft die Studie ab?

Diese Studie wird an der Universitäts-Augenklinik durchgeführt und es werden insgesamt 700 Personen daran teilnehmen. Ihre Teilnahme wird voraussichtlich 30 Minuten dauern.

<u>Folgende Maßnahmen werden ausschließlich aus Studiengründen durchgeführt:</u>

Sie werden gebeten, sechs Fragebögen sowie demografische Angaben auszufüllen. Es handelt sich dabei um den Patient-Activation-Measure (PAM)-13 Fragebogen zur Erfassung der aktiven

[1] Wegen der besseren Lesbarkeit wird im weiteren Text zum Teil auf die gleichzeitige Verwendung weiblicher und männlicher Personenbegriffe verzichtet. Gemeint und angesprochen sind – sofern zutreffend – immer beide Geschlechter.

Patientenbeteiligung, den F-Sozu K-22 zur Erfassung der wahrgenommenen sozialen Unterstützung, zwei Fragen zum Gesundheitszustand, der habituellen subjektiven Wohlbefindensskala, den Zielen die mit der Suche nach Gesundheitsinformation verbunden sind (GAINS) und der Skala zur allgemeinen Selbstwirksamkeitserwartung. Die demografischen Angaben beinhalten zum Beispiel Geschlecht, Alter, Berufsstand, Raucherstatus und HbA1c. Man benötigt für das Ausfüllen etwa 30 Minuten.

3. Worin liegt der Nutzen einer Teilnahme an dieser Studie?

Die Teilnahme an dieser Studie hat keinen direkten Nutzen für Sie. Die Ergebnisse dieser Studie können jedoch wertvolle Informationen liefern um die Betreuung von Patienten mit chronischen Erkrankungen zu verbessern.

4. Gibt es Risiken, Beschwerden und Begleiterscheinungen?

Nein. Außer dem Ausfüllen der Fragebögen werden keine weiteren Untersuchungen durchgeführt.

5. Datenschutz

Bei den Daten, die über Sie im Rahmen dieser klinischen Studie erhoben und verarbeitet werden, ist grundsätzlich zu unterscheiden zwischen

1) jenen personenbezogenen Daten, anhand derer Sie direkt identifizierbar sind (z.B. Name, Geburtsdatum, Adresse, Bildaufnahmen...),
2) pseudonymisierten (verschlüsselten) personenbezogenen Daten, bei denen alle Informationen, die direkte Rückschlüsse auf Ihre Identität zulassen, durch einen Code (z. B. eine Zahl) ersetzt bzw. (z.B. im Fall von Bildaufnahmen) unkenntlich gemacht werden. Dies bewirkt, dass die Daten ohne Hinzuziehung zusätzlicher Informationen und ohne unverhältnismäßig großen Aufwand nicht mehr Ihrer Person zugeordnet werden können und
3) anonymisierten Daten, bei denen eine Rückführung auf Ihre Person nicht mehr möglich ist.

Der Code zur Verschlüsselung wird von den verschlüsselten Datensätzen streng getrennt am Institut für medizinische Informatik, Statistik und Dokumentation aufbewahrt.

Zugang zu Ihren nicht verschlüsselten Daten haben der Prüfarzt und andere Mitarbeiter des Prüfzentrums, die an der klinischen Prüfung oder Ihrer medizinischen Versorgung mitwirken. Die Daten sind gegen unbefugten Zugriff geschützt. Zusätzlich können autorisierte und zur Verschwiegenheit verpflichtete Beauftragte des Sponsors Univ.-Prof. Dipl.-Ing. Dr. Andrea Berghold sowie Beauftragte von in- und/ oder ausländischen Gesundheitsbehörden und jeweils zuständige Ethikkommissionen in die nicht verschlüsselten Daten Einsicht nehmen, soweit dies für die Überprüfung der ordnungsgemäßen Durchführung der klinischen Prüfung notwendig ist.

Eine Weitergabe der Daten erfolgt nur in verschlüsselter oder anonymisierter Form. Auch für etwaige Publikationen werden nur die verschlüsselten oder anonymisierten Daten verwendet.

Sämtliche Personen, die Zugang zu Ihren verschlüsselten und nicht verschlüsselten Daten erhalten, unterliegen im Umgang mit den Daten der Datenschutz-Grundverordnung (DSGVO) sowie den österreichischen Anpassungsvorschriften in der jeweils gültigen Fassung.

Im Rahmen dieser klinischen Studie ist eine Weitergabe von Daten in Länder außerhalb der EU vorgesehen. Vom Sponsor werden die entsprechenden Maßnahmen ergriffen, die gemäß der DSGVO vorgesehen sind.

Sie können Ihre Einwilligung zur Erhebung und Verarbeitung Ihrer Daten jederzeit widerrufen. Nach Ihrem Widerruf werden keine weiteren Daten mehr über Sie erhoben. Die bis zum Widerruf erhobenen Daten können allerdings weiter im Rahmen dieser klinischen Studie verwendet werden.

Aufgrund der gesetzlichen Vorgaben haben Sie außerdem, sofern dies nicht die Durchführung der klinischen Studie voraussichtlich unmöglich macht oder ernsthaft beeinträchtigt, das Recht auf Einsicht in die Ihre Person betreffenden Daten und die Möglichkeit der Berichtigung, falls Sie Fehler feststellen.

Sie haben auch das Recht, bei der österreichischen Datenschutzbehörde eine Beschwerde über den Umgang mit Ihren Daten einzubringen (www.dsb.gv.at).
Die voraussichtliche Dauer der klinischen Studie ist 12 Monate. Die Dauer der Speicherung Ihrer Daten über das Ende der klinischen Studie hinaus ist durch Rechtsvorschriften geregelt.

Falls Sie Fragen zum Umgang mit Ihren Daten in dieser klinischen Studie haben, wenden Sie sich zunächst an Ihren Prüfarzt. Dieser kann Ihr Anliegen ggf. an die Personen, die am Studienzentrum für den Datenschutz verantwortlich sind, weiterleiten.

Datenschutzbeauftragte/r des Prüfzentrums: datenschutz@medunigraz.at, 0664 88 96 17 48

6. Möglichkeit zur Diskussion weiterer Fragen

Für weitere Fragen im Zusammenhang mit dieser Studie stehen Ihnen Ihr Studienarzt und seine Mitarbeiter gern zur Verfügung.

Name der Kontaktperson: Wedrich Andreas, Univ.-Prof. Dr.

Erreichbar unter: +43/316/385-12394

7. Einwilligungserklärung

Name des Patienten:

Geb.Datum:

Ich erkläre mich bereit, an der klinischen Studie ***Patientenaktivierung bei chronisch kranken PatientInnen*** teilzunehmen.

Ich bin von Herrn/Frau .... ausführlich und verständlich über die klinische Studie, mögliche Belastungen und Risiken, sowie über Wesen, Bedeutung und Tragweite der klinischen Studie, sich für mich daraus ergebenden Anforderungen aufgeklärt worden. Ich habe darüber hinaus den Text dieser Patientenaufklärung und Einwilligungserklärung, die insgesamt 4 Seiten umfasst gelesen. Aufgetretene Fragen wurden mir vom Studienarzt verständlich und genügend beantwortet. Ich hatte ausreichend Zeit, mich zu entscheiden. Ich habe zurzeit keine weiteren Fragen mehr.

Ich werde den ärztlichen Anordnungen, die für die Durchführung der klinischen Studie erforderlich sind, Folge leisten, behalte mir jedoch das Recht vor, meine freiwillige Mitwirkung jederzeit zu beenden, ohne dass mir daraus Nachteile für meine weitere medizinische Betreuung entstehen.

Ich stimme ausdrücklich zu, dass meine im Rahmen dieser klinischen Studie erhobenen Daten wie im Abschnitt „Datenschutz“ dieses Dokuments beschrieben verwendet werden.

Eine Kopie dieser Patienteninformation und Einwilligungserklärung habe ich erhalten. Das Original verbleibt beim Studienarzt.

....
(Datum und Unterschrift des Patienten)

....
(Datum, Name und Unterschrift des verantwortlichen Prüfarztes)

5.2 Appendix B

5.2.1 Data analysis

5.2.1.1 IRT

```
---
title: "FinalModel"
output:
word_document:
reference_docx: Vorlage1.docx
---
# Load packages
```{r, include=FALSE, echo=FALSE}
library(mirt)
library(ggmirt)
library(lordif)
library(lavaan)
library(psych)
library(psychTools)
library(tidyverse)
library(flextable)
library(crosstable)
library(WrightMap)
library(knitr)
```
Load data

```{r echo=TRUE, comment=NA}
load("C:/Users/lenah/Nextcloud/Doktorat/Auswertung/Activation/
   Daten/actidata_imputed_09082022.RData")

act <- act_imputed

#select sample for main analysis
filter <- act %>% select(ID, HS_01, FSOZU, SWE, HSWBS) %>% na.
   omit() %>% select(ID)
act <- merge(filter, act, by="ID")

data <- select(act, starts_with(c("PAM_")))
```
```

```
filename <- "actidata_v03062022.Rdata"

date_prepared_report <-  as.character(Sys.Date())

#Merge lower two response categories due to low cell occupation
data2 <- data
data2[data2 == 1] <- 2
‘‘‘

# 1. assumption: unidimensionality
## CFA with three categories

‘‘‘{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
invisible(mod1 <- (mirt(data2, 1)))
mod1s <-(summary(mod1))
mod1f <- cbind(mod1s$rotF, mod1s$h2)
kable(mod1f,  digits=2)
‘‘‘

# Final model: generalized partial credit model

## Model calculation

‘‘‘{r,  echo=TRUE, error=TRUE, comment=NA}
results.gpcm <- mirt(data=data2, model=1, itemtype="gpcm", SE=
    TRUE, verbose=FALSE)
show(results.gpcm)
‘‘‘

## Fit indizes

‘‘‘{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
fit1.gpcm <- M2(results.gpcm, type='C2', na.rm=TRUE)
AICc <- extract.mirt(results.gpcm, 'AICc')
SABIC <- extract.mirt(results.gpcm, 'SABIC')
HQ <- extract.mirt(results.gpcm, 'HQ')
fit2.gpcm <- cbind(AICc,SABIC, HQ)
```

```
fit.gpcm <- cbind(fit1.gpcm, fit2.gpcm)
rownames(fit.gpcm) <- c("GPCM")
kable(fit.gpcm, digits=c(rep(3,9),rep(1,3)))
```
```

## Empirical plots

```{r,  echo=TRUE, error=TRUE, comment=NA}
itemfit(results.gpcm, empirical.plot = 1, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 2, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 3, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 4, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 5, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 6, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 7, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 8, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 9, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 10, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 11, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 12, na.rm=TRUE)
itemfit(results.gpcm, empirical.plot = 13, na.rm=TRUE)
```

Item statistics

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
   height=12.07}
cat("Itemfit statistics")
kable(itemfit(results.gpcm, fit_stats = c("S_X2","infit"), na.rm
   = TRUE), digits=2)
cat("Item difficulty parameter")
kable(gen.difficulty(results.gpcm, type = "IRF"), digits=2)
cat("Discrimination parameter")
kable(MDISC(results.gpcm), digits=2)
```

Item characteristic curves

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
```
```

```
    height=12.07}
#Option characteristic curves
plot(results.gpcm, type = 'trace',
main = "", par.settings = simpleTheme(lty=1:4,lwd=2),
auto.key=list(points=FALSE,lines=TRUE, columns=4), theta_lim = c
    (-6, 6))

#Item information function
plot(results.gpcm, type = 'infotrace',
main = "", par.settings = simpleTheme(lwd=2, col="#007934", font=
    "serif") , theta_lim = c(-6, 6))

plot(results.gpcm, type = 'infotrace', facet=FALSE, main = "",
    par.settings = simpleTheme(lwd=2), theta_lim = c(-6, 6))

‘‘‘

## Test information

‘‘‘{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
plt <- plot(results.gpcm, type = 'infoSE') #store the object
pltdata <- data.frame(lapply(plt$panel.args, function(x) do.call(
    cbind, x))[[1]])
#Estimate standard error
pltdata$yse<- 1/(sqrt(pltdata$y))

#Plot
ggplot(pltdata, aes(x)) +
geom_line(aes(y = yse, colour = "#bfc0c2"), lwd=4) + geom_line(
    aes(y = y, colour = "#007934"), lwd=4) +
scale_y_continuous(expression(paste('Information '(theta)), size
    =44),
sec.axis = dup_axis(name=expression(paste('Standard error '(theta
    )), size=44))) +
xlab(expression(paste('Person ability '(theta)), size=44)) +
theme_light() +
theme(panel.grid.minor = element_blank(), panel.grid.major =
    element_blank(), panel.border=element_rect(color="black",
    fill=NA),
```

```
plot.title = element_text(hjust = 0.5, size=48), text=element_
    text(size=44,  family="serif"),
axis.text.x = element_text(color="black", size=40),  axis.text.y
    = element_text(color="black", size=40), legend.position = "
    bottom", legend.title=element_blank())+
scale_colour_manual(values =c('#007934'='#007934','#bfc0c2'='#
    bfc0c2'), labels = c('Information','Standard error'), name=
    NULL)

#How much test information in specific areas
areainfo(results.gpcm, c(-6,6))
areainfo(results.gpcm, c(-6,0))
areainfo(results.gpcm, c(-0,6))
#Area were most of the sample is
areainfo(results.gpcm, c(-2.5,2))
```
```

## Wright Map

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=16.09, fig.
    height=12.07}

Theta <- fscores(results.gpcm)
coefG <- coef(results.gpcm, IRTpars = TRUE, simplify = TRUE)
Thresh <- coefG[["items"]]

Thresh <- as.data.frame(Thresh[ ,-1])
#Rename items
rownames(Thresh) <- c("1", "2", "3", "4", "5", "6", "7"," 8", "9"
    , "10", "11", "12", "13")

par(family="serif",ps = 18)

wrightMap(thetas=Theta, thresholds = Thresh, vertLines = TRUE,
thr.sym.col.bg =rep(c("#007934", "#6b6d74"), each=13), thr.sym.
    cex = 1.4,
main.title = "Wright Map GPCM", axis.persons = "Patient
    activation ability distribution",  dim.names = " ", family =
    "A")
```
```

```
## Corr Plot

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
act2 <- data.frame(act, Theta)

par(family="serif",ps = 19)
act2 %>% select(PAM, F1) %>% pairs.panels(stars=FALSE, method="
    spearman", hist.col="#007934", rug = FALSE, cex.cor=1, digits
     = 2, breaks = 50, smooth=FALSE, ellipses = FALSE)
```

Reliability

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
theta_se <- fscores(results.gpcm, full.scores.SE = TRUE)
cat("Empirical")
print(empirical_rxx(theta_se))

plot(results.gpcm, type="rxx", lwd=2)

cat("Marginal")
print(marginal_rxx(results.gpcm))

```

Person reliability and person separation index

```{r, echo=TRUE, error=TRUE, comment=NA, fig.width=15.09, fig.
    height=12.07}
#To calculate the MSE, I need the SE of my model
## Get the SE from my fitted model
sedata <- as.data.frame(fscores(results.gpcm, full.scores.SE =
    TRUE))
MSEp <- sum(sedata$SE_F1^2) #141.9957

#-----#
#SSDp
```
```

```
meanp <- mean(act2$F1)
act2$diffpc <- meanp - act2$F1
SSDp <- sum(act2$diffpc^2) #411.9926

#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   ~~~~~~~~~~~~~~~~~~~~~~~~~~#
#Person reliability index
pri <- 1 - (MSEp/SSDp) #0.655
cat("Person Reliability Index")
print(pri)
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   ~~~~~~~~~~~~~~~~~~~~~~~~~~#
#Separation

G <- sqrt((pri/(1-pri)))
cat("Person Separation Index")
print(G)

cat("Anzahl Gruppen empirisch")
#Anzahl Gruppen H = (4G + 1)/3
nstrata <- (4*G+1)/3
print(nstrata)

#or:
G2 = sqrt(0.75/(1-0.75))
nstrata2 <- (4*G2+1)/3

'''

# Cronbach's alpha

'''{r, echo=TRUE, comment= NA}

pam13 <- list(PAM= c("PAM_01", "PAM_02", "PAM_03", "PAM_04", "PAM
   _05", "PAM_06", "PAM_07", "PAM_08", "PAM_09", "PAM_10", "PAM_
   11", "PAM_12","PAM_13"))
pam <- scoreItems(pam13, data2)

cat("Alpha")
kable(pam$alpha, digits=3)
```

```
cat("Scale intercorrelations corrected for attenuation raw
   correlations below the diagonal, alpha on the diagonal
   corrected correlations above the diagonal")
kable(pam$corrected, digits=3)
cat("Average item correlation")
kable(pam$av.r, digits=3)
cat("Median item correlation")
kable(pam$med.r, digits=3)
cat("Item by scale correlations: corrected for item overlap and
   scale reliability")
kable(pam$item.cor,digits=3)
```

```

# Trait-trait correlations

```{r, results="asis", echo=FALSE, fig.width=8, fig.height=6.4}
cat("With overall scores")
act %>% select(PAM, FSOZU, HS_01, HSWBS, SWE) %>% pairs.panels(
   stars=TRUE, method="spearman", hist.col="#007934", rug =
   FALSE, cex.cor=1, digits = 2, breaks = 50, smooth=FALSE,
   ellipses = FALSE)
cat("Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05")
act %>% select(PAM, FSOZU, HS_01, HSWBS, SWE) %>%corr.test(method
   ="spearman") %>% print(short=FALSE) %>% kable()

cat("With scores of subcales")
act %>% select(PAM, FSozu_Pu, FSozu_Eu, FSozu_SI, HSWBS_GS, HSWBS
   _LZ) %>% pairs.panels(stars=FALSE, method="spearman", hist.
   col="#007934", rug = FALSE, cex.cor=1, digits = 2, breaks =
   50, smooth=FALSE, ellipses = FALSE)
act %>% select(PAM, FSozu_Pu, FSozu_Eu, FSozu_SI, HSWBS_GS, HSWBS
   _LZ) %>%corr.test(method="spearman") %>% print(short=FALSE)
   %>% kable()

cat("With GAINS, different N")
act %>% select(PAM, FSOZU, HS_01, HSWBS, SWE, GAINS) %>% pairs.
   panels(stars=TRUE, method="spearman", hist.col="#007934", rug
    = FALSE, cex.cor=1, digits = 2, breaks = 50, smooth=FALSE,
   ellipses = FALSE)
act %>% select(PAM, FSOZU, HS_01, HSWBS, SWE, GAINS) %>%corr.test
```
```

```
    (method="spearman") %>% print(short=FALSE) %>% kable()
```
‘‘‘

5.2.2 Additional results

5.2.2.1 Model comparison

Table 5.1: Polytomous models estimated with initially four categories. Dichotomous models estimated with "Disagreement" vs. "Agreement" and "1-3" against "4".

Parameter	**Rating scale model**	**Partial credit model**	**Generalized partial credit model**	**Graded response model**	**Dichotomous model 1**	**Dichotomous model 2**	**GPCM final model**
RMSEA (90% CI)	0.069 (0.061 – 0.077)	0.056 (0.047 – 0.066)	0.052 (0.041 – 0.062)	0.062 (0.052 – 0.072)	0.044 (0.033 – 0.053)	0.055 (0.046 – 0.065)	0.062 (0.052 - 0.072)
SRMSR	0.08	0.07	0.06	0.079	0.067	0.064	
TLI	0.866	0.912	0.926	0.894	0.893	0.922	0.905
CFI	0.827	0.913	0.938	0.912	0.9	0.9	0.921
SABIC	12575.2	12514.8	12513.8	12430.1	4446.6	8254.7	11496.6
AICc	12557.9	12475.51	12465.36	12381.71	4431.4	8239.5	11458.1
Total test information	40.00	39.00	30.50	29.45	12.98	13.0	22.60
Standard Error: Mean ± SE	0.40 ± 0.07	0.38 ± 0.07	0.51 ± 0.10	0.50 ± 0.07	0.82 ± 0.13	0.59 ± 0.06	0.50 ± 0.08
Empirical reliability	0.72	0.72	0.73	0.75	0.56	0.74	0.75
Marginal reliability	0.80	0.80	0.73	0.75	0.45	0.69	0.74
Pearson correlation with PAM® score	0.99 (0.99 - 0.99)	>0.99 (0.99 - 0.99)	0.98 (0.98 - 0.98)	0.97 (0.96 - 0.97)	0.84 (0.81 – 0.86)	0.92 (0.90 – 0.93)	0.97 (0.97 – 0.98)

RMSEA=root mean square error of approximation; SRMSR= standardized root mean square residual;
TLI=Tucker-Lewis index; CFI=comparative fit index;
SABIC=sample adjusted Bayesian information criterion. AICc = Akaike information criterion corrected.

5.2.2.2 Differential item functioning

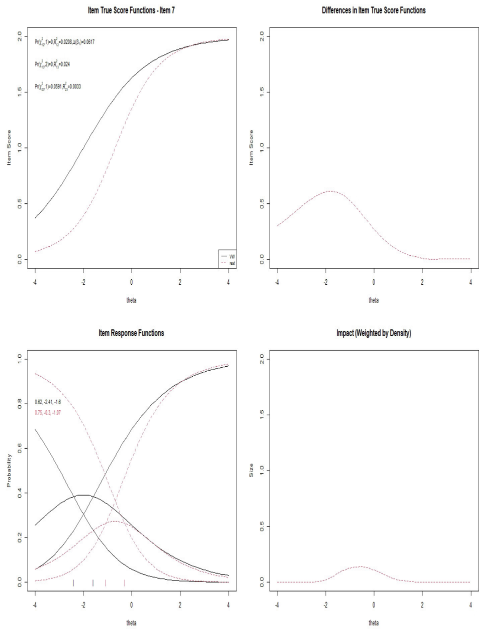


Figure 5.1: DIF interviewer for item 7: Uniform DIF for interviewer 2 compared to the other interviewers, presented as differences in item true score functions, displayed as the dotted line. Theta = person ability in logits.

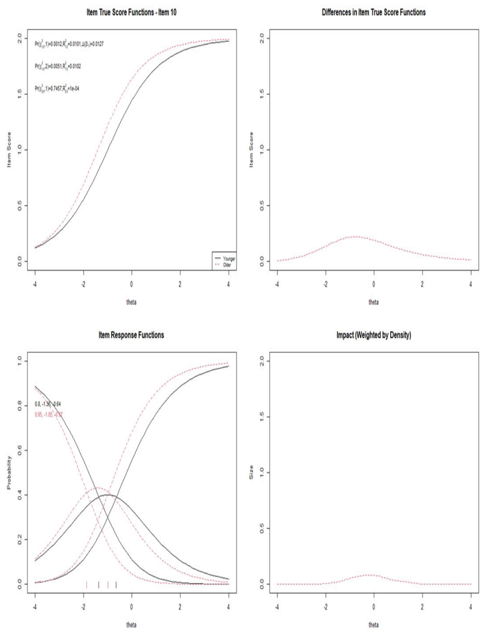


Figure 5.2: DIF age for item 10: Uniform DIF for age in groups (<70 and $\geq$ 70 years), presented as differences in item true score functions, the group $\geq$ 70 years is displayed as the dotted line. Theta = person ability in logits.

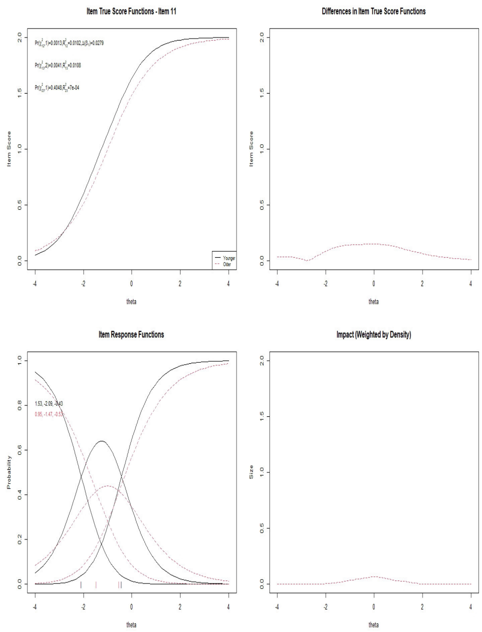


Figure 5.3: DIF age for item 11: Uniform DIF for age in groups (<70 and $\geq$ 70 years), presented as differences in item true score functions, the group $\geq$ 70 years is displayed as the dotted line. Theta = person ability in logits.

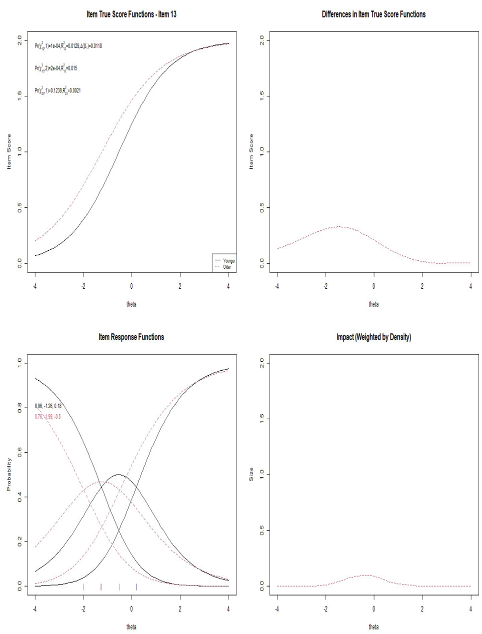


Figure 5.4: DIF age for item 13: Uniform DIF for age in groups (<70 and $\geq$ 70 years), presented as differences in item true score functions, the group $\geq$ 70 years is displayed as the dotted line. Theta = person ability in logits.

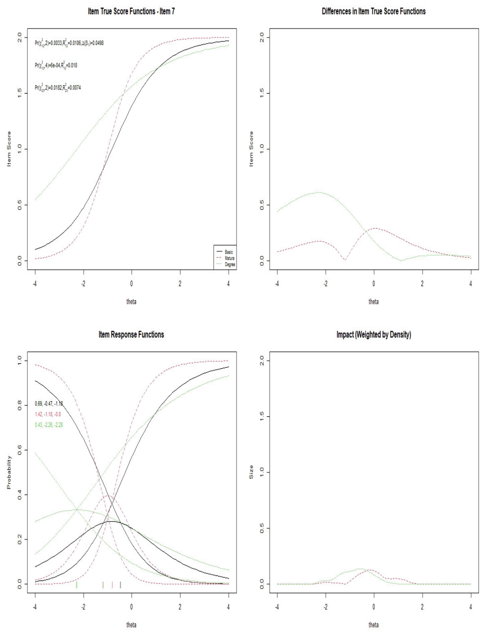


Figure 5.5: Non-uniform unbalanced DIF educational status for item 7: presented as differences in item true score functions. Reference group = basic education, dotted line = high school, solid line = higher education. Theta = person ability in logits.

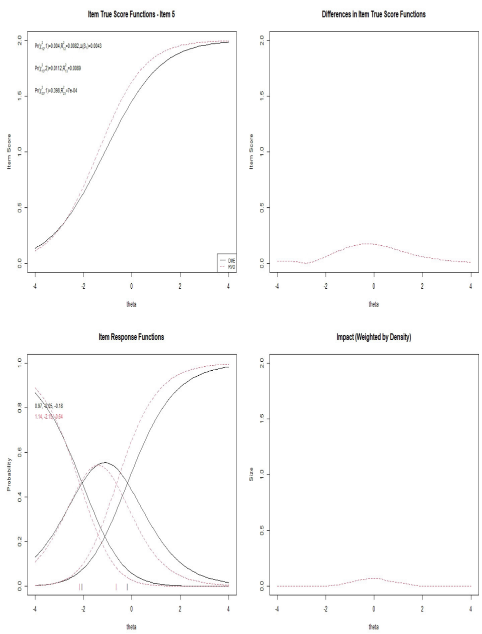


Figure 5.6: DIF type of macular edema for item 5: Uniform DIF for type of macular edema, presented as differences in item true score functions, the group RVO is displayed as the dotted line. Theta = person ability in logits.

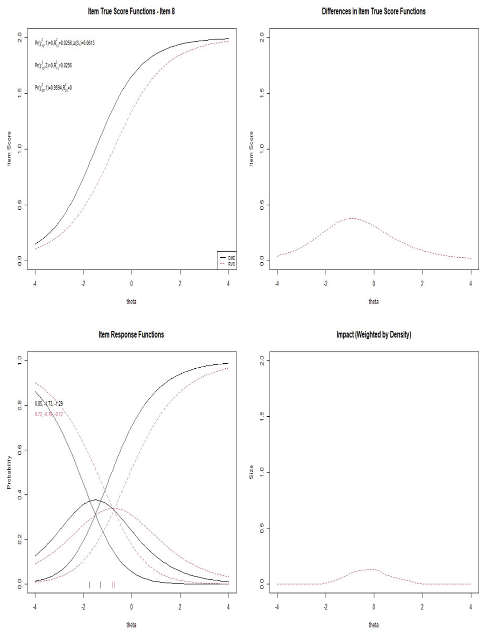


Figure 5.7: DIF type of macular edema for item 8: Uniform DIF for type of macular edema, presented as differences in item true score functions, the group RVO is displayed as the dotted line. Theta = person ability in logits.

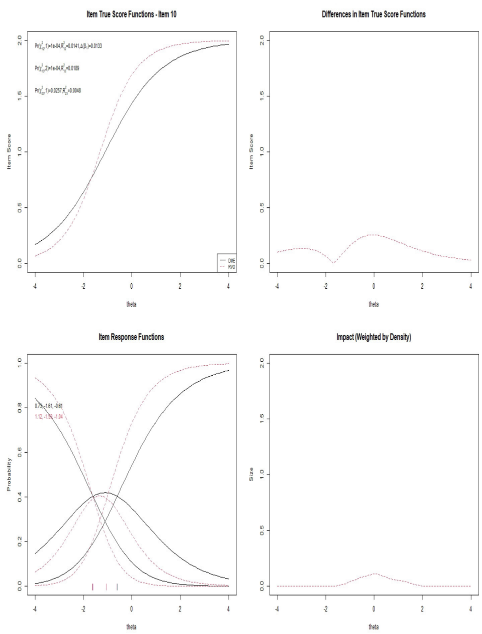


Figure 5.8: DIF type of macular edema for for item 9: non-Uniform DIF for type of macular edema, presented as differences in item true score functions, the group RVO is displayed as the dotted line. Theta = person ability in logits.

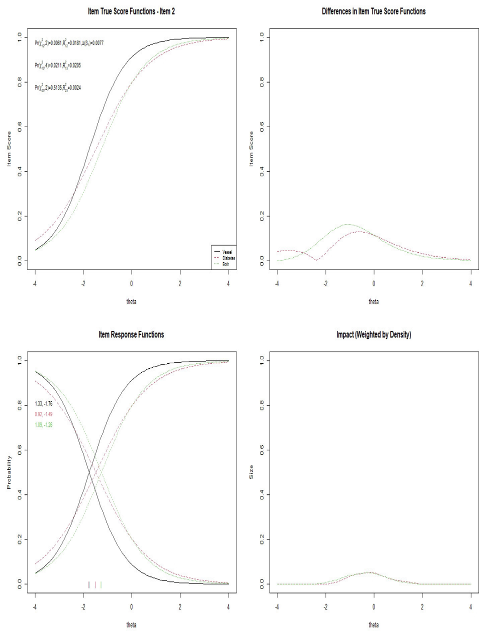


Figure 5.9: DIF comorbidity for item 2: Uniform DIF for comorbidity, presented as differences in item true score functions. Reference group = hypertension, dotted line= diabetes, solid line = hypertension & diabetes. Theta = person ability in logits.

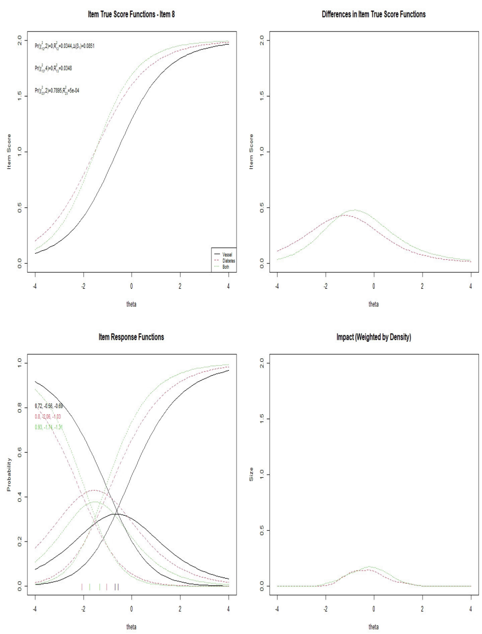


Figure 5.10: DIF comorbidity for item 8: Uniform DIF for comorbidity, presented as differences in item true score functions. Reference group = hypertension, dotted line= diabetes, solid line = hypertension & diabetes. Theta = person ability in logits.

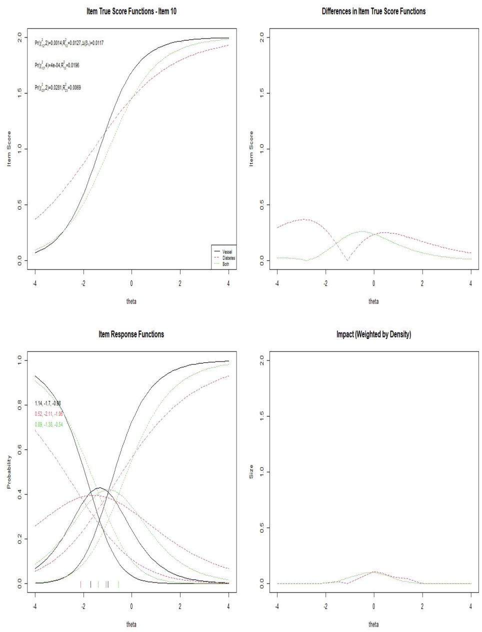


Figure 5.11: DIF comorbidity for item 10: Uniform DIF for comorbidity, presented as differences in item true score functions. Reference group = hypertension, dotted line= diabetes, solid line = hypertension & diabetes. Theta = person ability in logits.

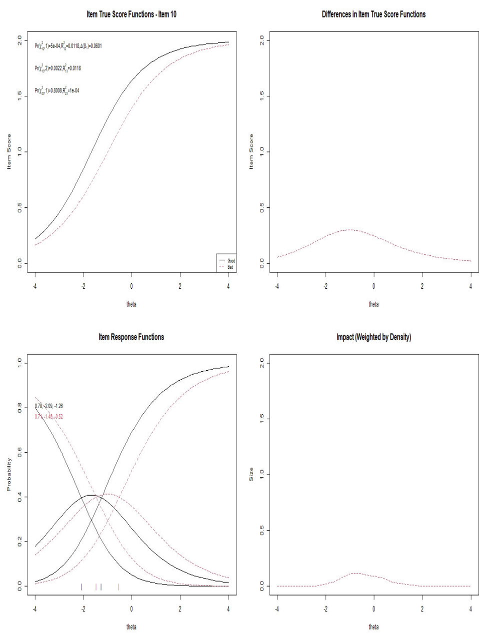


Figure 5.12: DIF health status for item 10. Uniform DIF for health status presented as differences in item true score functions. Reference group = very good & good, dotted line= moderate, bad & very bad. Theta = person ability in logits.