Psycho-oncology and assessment methods:

A critical reflection

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Lusine Rafaelovna Vaganian

geb. in Moskau

– Köln, Juni 2021 –

Psycho-oncology and assessment methods: A critical reflection - II

– Fach Psychologie –

Kumulative Dissertation zur Erlangung des Doktorgrades im Fachbereich

Klinische Psychologie und Psychotherapie

Erster Gutachter: Prof. Dr. Alexander L. Gerlach

Zweiter Gutachter: Jun.-Prof. Dr. Johannes C. Ehrenthal

Tag der mündlichen Prüfung: 20.10.2021

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Oktober 2021 angenommen.

Danksagung

– „Nichts ist unmöglich" –

…Hinter all den Theorien und Bemühungen, steckt eine Geschichte, ein Mensch, ein Lebensweg. Entscheidungen, die hierhergeführt haben und Menschen, die einen begleitet haben. Dieser Teil der Arbeit ist wohl der persönlichste und für mich ein sehr zentraler Teil. Denn Dankbarkeit ist eine schöne Fähigkeit und erst recht, wenn man sie auch nach außen tragen und sie teilen kann. Um solch ein Gesamtwerk zu vollbringen, verbringt man viel Zeit und Mühen damit, Monate, ja sogar Jahre. Wenn man schon so viel investiert hat, dann lohnt es sich kurz inne zu halten, den Lebensweg zu reflektieren und sich die Zeit zu nehmen sich gebührend zu bedanken. Daher nutze ich diesem Raum, um all dem und denjenigen zu danken, die sehr viel dazu beigetragen haben, dass ich nun an diesem Punkt in meinem Leben stehe. Grundsätzlich danke ich dem Leben, dass es mir trotz Berg- und Talfahrten stets Kraft gegeben hat weiter zu machen und durchzuhalten. Aber nichts wäre möglich ohne die Menschen, die einen geprägt, auf dem Lebensweg begleitet haben und mich unterstützt haben. Daher gilt mein ganz besonderer Dank Euch! Auch wenn hier nicht alle Menschen Erwähnung finden können, die mich bisher begleitet haben, bin ich euch nicht weniger dankbar…

…Um es meinem Lieblingsthriller-Autor Sebastian Fitzek gleichzutun, der, nicht nur, fesselnde Bücher schreibt, und dem ich hier an dieser Stelle auch danke, weil er mir mit seinen Büchern die nötige und spannende Ablenkung zwischen Prüfungs- und Lernphasen geschenkt hat, sondern weil er auch stets herrliche und inspirierende Danksagungen formuliert, hole ich an dieser Stelle etwas aus.

Das oben aufgeführte Zitat, ist nicht aus der Werbung einer Automarke, sondern ist an Alice im Wunderland angelehnt, denn das Unmögliche zu schaffen, gelingt einem nur, wenn man daran glaubt…Bei jedem Lebensweg geht es um Märchen und Träume und die Kraft daran zu glauben, diesen Weg zu meistern und seine Ziele zu verwirklichen. Eine Dissertation zu schreiben ist bei weitem nicht Unmöglich. Aber in manchen Phasen des Lebens kam es einem schon fast so vor. Eine Dissertation ist jedoch ein Ziel oder auch ein Traum, den man verwirklichen will. Um dies zu erreichen, erfordert es viel Kraft, und ganz besondere Menschen um sich herum…

Die Person, die mir das Träumen beigebracht und erlaubt hat, die mich darin gefördert und unterstützt hat, ist meine Mutter Irina. Sie hat mir nicht nur das Leben geschenkt, sondern mir alles ermöglicht und mich auf meinen Wegen unterstützt, gefördert und mich darin bestärkt meine Grenzen auszuweiten. Für all deine Liebe, Geborgenheit, Vertrauen in mich und meine Fähigkeiten sowie deine Zuversicht, liebevollen Worte, deine Bewunderung und noch so viel mehr, danke ich dir von ganzem Herzen!

Außerdem danke ich meinem Vater Rafael, der voller Stolz und Bewunderung, bei jedem kleinsten Schritt auf mich aufgepasst hat und der am liebsten bei jedem

kleineren sowie größeren Meilenstein ein Stadtfest zu meinen Ehren organisiert hätte. Papa auch dir danke ich für alle deine Mühen, deine Liebe und Wärme!

Meinem Bruder Armen gilt natürlich auch besonderer Dank. Der nicht nur meinen beruflichen Weg mitverfolgt und mitgefiebert hat, sondern mir auch persönlich sehr nahesteht. Ich danke dir für die stundenlangen Diskussionen über philosophische und gesellschaftliche Themen und insbesondere danke ich dir vor dem Hintergrund dieses Werkes, dass du mir immer den nötigen Ausgleich gegeben und mich von allem Stress und Unruhe abgelenkt hast sowie schöne spielerische Freuden bereitet hast.

Meinem Opa aus Armenien, der selbst noch bis ins hohe Alter an der Uni tätig war und der ein Mann voller Liebe, Wärme und hohen moralischen Werten war. Egal, wo du auch bist, ich weiß es wird dich freuen, dass ich diesen Weg gegangen bin und mein Werk vollbracht habe. Natürlich danke ich auch meinen Großeltern aus Russland, die beide stets an meinem Werdegang interessiert und so voller Stolz und Wärme zu mir sind, und trotz Entfernung meinem Weg hautnah beigelebt haben. Oma mit dir bin ich thematisch sämtliche Aufgaben und Probleme durchgegangen und wir haben sämtliche Lösungsmöglichkeiten durchdacht. Opa du hast mir immer die nötige Ruhe und Auszeit vorgeschrieben und mich stets geerdet.

Meinem Partner Kai gilt sehr großer Dank, der all die hochstressige Zeit mit mir gemeinsam durchgestanden hat, der mich immer wieder weggeführt hat von Theorien und Gedanken und mir die Freuden des Lebens hat spüren lassen, der mich unterstützt hat, wie er nur konnte und der mir mit kleinen und großen Überraschungen immer wieder das Leben versüßt hat. Ich danke dir mein Schatz für deine Bewunderung, deine Aufrichtigkeit, deine Neugierde sowie deine Aufgeregtheit, Anspannung und Jubel bei jedem kleinen und großen Meilenstein. Ich freue mich auf unseren weiteren gemeinsamen Lebensweg! Ich danke auch deiner Mama Gabi und ihrem Partner Georg dafür, dass sie mich stets wärmstens empfangen, und während der Zeit Ablenkung verschafft haben und mich stets zu den Wurzeln – insbesondere in der Gartenarbeit – zurückgeführt haben.

Meinen Tanten Natasha und Ruzana danke ich für ihre Leichtigkeit, ihre aufmunternden und motivierenden Worte und die Bewunderung und Stolz, den sie für meine Tätigkeiten aufbringen. Meinen Cousinen und Cousins, die überall in der Welt verstreut sind, und wir uns trotzdem so nah sind, danke ich, dass sie trotz Entfernung stets meinen Lebensweg mitverfolgen, mitfiebern und mitjubeln, sobald ein weiterer Meilenstein erreicht ist.

Freunde sind die Familie die man sich selbst aussucht! – Und das kann ich wahrlich behaupten. Ich bin für solch enge und liebe Menschen, die mir so nahe stehen unglaublich dankbar. Daher gilt an dieser Stelle ein besonderer Dank meinen engsten Freunden, die mich seit langer Zeit in vielen Phasen begleiten und mir Ausgleich, Wärme, Ruhe und Liebe schenken, da wo ich sie brauche…

…So wie Ina, die ich seit Beginn meiner Schulzeit kenne und ich mit ihr stets ausgiebig lachen, reden und die Seele baumeln lassen kann. Die Zeit mit dir verfliegt immer wie

im Flug, und du hast mir viel Kraft in allen schweren und leichteren Phasen gegeben! Danke, dass du so bist wie du bist! Marika, die seit meinem FSJ in mein Leben getreten ist und die im wahrsten Sinne des Wortes mich während meines Bachelorstudiums Tag und Nacht begleitet hat. Die seitdem ich sie kenne, bereits auf diesen Tag hin gefiebert hat und mir in vielen Jahren Halt, Kraft, Aufmunterung und Freude geschenkt hat. Sophie, die mit mir gemeinsam das konsequente Lernen, fernab der Heimat im Studium durchgestanden hat und auch jetzt trotz Entfernung, meinen Weg begleitet und mir immer wieder Aufmunterung schenkt. Dinh Vu, dir danke ich dafür, dass du mich stets motivierst und dass du so voller Stolz und Erstaunen von mir und meinen Leistungen sprichst und dass du immer ein wundervolles Wort für mich und jeden übrighast. Jeder von euch hat mir auf seine eigene Art und Weise vieles mitgegeben und mich stets unterstützt und meinen Lebensweg so sehr bereichert! Dafür danke ich euch von ganzem Herzen!

Auch meinen Freunden, die ich im Laufe meiner universitären Laufbahn kennenlernen durfte, danke ich für ihren Rückhalt. An dieser Stelle ganz besonders…Jenny du bist wahrlich eine Löwenmutter und strahlst unglaubliche Stärke aus! Susi, du bist ein großer Wirbelwind und begeisterst mich mit deiner Leichtigkeit. Arne, ich erinnere mich von Herzen gerne an die Zeit wo wir Stunden und Nächte verbracht haben, um Schlafanalyseauswertungen zu machen. Danke, dass du mich stets an deinem Wissen teilhaben hast lassen und mich motiviert hast meinen Weg zu gehen!

Sowie meinen Kollegen in der Pohligstr., danke, dass ihr ein herrliches Team seid und für wundervolle Weihnachtsfeiern und schöne Zeiten zwischendurch, hier an dieser Stelle insbesondere: Christina, Carolin, Anna, Timo und Helmut.

Wenn ich weit zurückblicke gibt es auch Lehrer in der Schulzeit, denen ich dafür danken möchte, dass sie mir vieles beigebracht haben und meinen gedanklichen Horizont erweitert haben, der mir auch heutzutage hilfreich ist. An dieser Stelle bedanke ich mich insbesondere bei Herrn Stricker, der ein großartiger Lehrer war und faszinierender Mensch ist, und bei Frau Scigala-Ostrek, die mich mit ihrer herzlichen und mütterlichen Art stets gefördert und unterstützt hat.

Während meiner studentischen und wissenschaftlichen Laufbahn, gab es natürlich weitere Vorbilder und Professoren, die mich in meinem Lebensweg gefördert, zu meiner Entwicklung beigetragen haben und meinen gedanklichen Horizont erweitert haben, denen ich hierfür danke! Obwohl ich seit Kindesbeinen meinen beruflichen Lebensweg schon genauestens kannte und plante, Psychotherapeutin zu werden, bekam ich während meines Bachelorstudiums Einsicht in die Forschungspraxis, die mich zu faszinieren begann. Hier an dieser Stelle ganz besonders geprägt durch Frau Prof. Dr. Elisabeth Kals, die ich für ihr Engagement für die Studenten, ihre unglaublich herzliche Art, ihre Zuversicht und Wissbegierde, ihre moralische und methodische Haltung schätze, und mich für ihre Unermüdlichkeit aufbauende und lobende Worte zu formulieren, das Vertrauen, was sie mir entgegengebracht, die vielen Perspektiven, die sie mir ermöglicht hat, sowie ihre Unterstützung in den Vorhaben ganz herzlich bedanken möchte.

Table of Content

Abstract

The application of appropriate methods in psychology and in science in general is an integral part of high quality research. However, recommendations and empirical practice often look quite different. The aim of this thesis is to critically reflect on which methodological principles are proposed and which are actually applied, using examples from the field of psycho-oncology. To this end, three studies were conducted to address essential methodological aspects in health research.

Individualized psycho-oncology is increasingly becoming the social and scientific norms. In order to be able to measure possible changes in treatment approaches depending on individual severity and the needs of the patients, an individualized methodological approach is also necessary, i.e., the assessment of clinically significant change. In the first study, two commonly used measures of clinical significance, i.e., the Reliable Change Index (RCI) and the Minimal Important Difference (MID), are compared and critically examined. Based on the analysis, it is recommended to use the RCI measure to avoid possible overestimation of treatment effects.

Several assessment instruments have emerged in psycho-oncology that aim to examine psychological processes related to the cancer disease and to evaluate treatment effects. Reliable and valid assessment instruments are important when planning clinical interventions. The psychometric properties of assessment instruments can be evaluated using classical test theory (CTT) or item response theory (IRT). The advantages of IRT approaches are that it is possible to obtain more detailed information about the scale or items and about the person's abilities. Due to these prevailing advantages, the use of IRT-based models is called for. The second and third study of the thesis, accordingly, examine the psychometric properties of two different assessment instruments, using item response theory, i.e., item analysis according to the Rasch model. The second study examines the World Health Organization's Disability Assessment Schedule 2.0 (WHODAS 2.0), a commonly used measure of disability. The instrument of the third study investigates the Positive Mental Health (PMH) scale. The WHODAS 2.0 proves to be well suited to assess disability in the psycho-oncological context, especially those who have an impairment will be adequately assessed with it, which are similar results to the CTT studies. The inclusion of positive psychology approaches has also been shown to be beneficial for cancer patients. The examined PMH scale in the third study is a unidimensional measure of

positive mental health and the scale can also be used well in the oncological context in its adapted version. However, the 8-item solution fits the model better, a contrast to findings of most CTT studies.

Overall, the present thesis critically reflects methodology taking examples from psycho-oncology. Rigorous scientific requirements of methods and how methods ultimately are implemented in research practice are sometimes two different things. Further studies on high quality methodology are needed and the application of rigorous requirements in practice should continue to be called for and implemented.

Abbreviation

ACT = Acceptance and Commitment Therapy

ADL = Activities of Daily Living

ANOVA = Analysis of Variance

APA = American Psychiatric Association

CTT = Classical Test Theory

DIF = Differential Item Functioning

DSM-5 = Diagnostic and Statistical Manual of mental disorder - fifth edition

DT = Distress-Thermometer

ECV = Explained Common Variance

GAD-7 = Patient health questionnaire – Generalized Anxiety Disorder Screener

GSE = General Self-Efficacy

HADS = Hospital Anxiety and Depression Scale

HADS-D - Hospital Anxiety and Depression Scale – German version

HADS-T = HADS total score

HSI = Hornheide Screening Instrument

ICC = Item Characteristic Curves

ICD-10 = International Classification of Diseases 10th Revision

ICF = International Classification of Functioning, Disability, and Health

IRT = Item Response Theory

isPO = Integrated, Cross-Sectoral Psycho-Oncological Care Program

MCID = Minimal Clinically Important Difference

MID = Minimum/Minimal Important Difference

LD = Local Dependence

PCA = Principal Component Analysis

PCM = Partial Credit Model

PHQ-9 = Patient health questionnaire – Depression

PMH = Positive Mental Health

PCOMs = Patient-Centered Outcome Measures

PROs = Patient-Reported Outcomes

PSI = Person Separation Index

QSC-R23 = Questionnaire on Stress in Cancer Patients

RCI = Reliable Change Index

RCT = Randomized Controlled Trial

RMT = Rasch Measurement Theory

SD = Standard Deviation

SE = Standard error

SEM = Standard Error of Measurement

WHODAS 2.0 = World Health Organization's Disability Assessment Schedule 2.0

WHO = World Health Organization

Figures

Tables

1. Introduction and Objectives

"Hence even in the domain of natural science, the aid of the experimental method becomes indispensable whenever the problem set is the analysis of transient and impermanent phenomena, and not merely the observation of persistent and relatively constant objects" (Wundt, 1910, p. 4).

As already emphasized by Wilhelm Wundt, the founder of experimental psychology, and thus the development of psychology as an independent science, the application of the scientific methodology is indispensable in the investigation of features that are not directly visible. However, this invisibility of measured constructs is reflected in the subject of psychology in the study of experience and behavior. Much time has passed since the beginnings of experimental research, and many methods and diverse theories have developed and emerged.

There is a plethora of methodological procedures that describe a standardized approach to the study of research objects. Meanwhile, there are many methodological theories that address HOW to test in the first place, and many methods that determine WHAT to test. The question that arises with abundance is which method to choose for one's scientific endeavor. Despite efforts at standardization and recommendations, scientific practice often looks quite different.

The example of psycho-oncology will be used to critically reflect on which methodological principles are proposed and actually applied, although the principles described here can of course be transferred and applied to other areas of psychology. In this thesis, psycho-oncology will be briefly described by first defining psycho-oncology and clarifying relevant historical aspects. Furthermore, the clinical picture of cancer and its associations with psychological components, will be described before a brief description of treatment options.

Subsequently, the methodological aspects of psycho-oncological research are presented and embedded in health services research. The assessment instruments for recording psychological components of the disease are introduced, and an overview of effectiveness research is given.

As mentioned at the beginning, the methodology is a central component in science. Whereby there are often differences in the required recommendations and empirical

practice. For example, approaches to effectiveness research in practice are often based on the group-level investigation, whereas for clinical implications, it is helpful in also examine individual-level significance, i.e., clinical significance. However, even within the investigation of clinical significance, there are several ways to do this. Therefore, the first study of the present thesis contrasted the most common methods of clinical significance and critically considered which methods should be used in the psycho-oncological context.

The remainder of this thesis will provide insight into the theories of how the psychometric quality of assessment instruments can be measured. The central theories are those of classical test theory (CTT) and item response theory (IRT), including the Rasch measurement theory (RMT). These are briefly introduced and compared. Again, a discrepancy between claims and implementation in scientific practice is apparent. Although IRT models provide more meaningful information about the psychometric properties of assessment instruments new, modern theories are called for in practice, it is less common in actual application. Following this call, the second study of the present thesis uses the Rasch model to evaluate the psychometric properties in the psycho-oncological context of a disability measurement instrument, i.e., the World Health Organization's Disability Assessment Schedule 2.0 (WHODAS 2.0) (Üstün et al., 2010), recommended in the Diagnostic and statistical manual of mental disorder (DSM-5) (American Psychiatric Association [APA], 2013) and is frequently used.

Another area that has long been neglected in psychology, in general, is the approach of positive psychology. This approach is about mental health and not just the absence of illness, but a state of well-being that positively impacts the whole range of life factors (Keyes, 2005; World Health Organization [WHO], 2004).

As aspects of positive psychology, predominantly positive mental health and factors such as optimism (Schiavon et al., 2017), are also shown to be relevant to the likelihood of survival in cancer patients, this long-neglected area represents an important starting point for clinical implications. In the wake of greater attention, numerous measures have been developed to examine people's positive abilities. For example, a brief inventory has also been developed. i.e., the positive mental health (PMH) scale (Lukat et al., 2016) and its psychometric properties have been tested using approaches from CTT. Due to the already explained demand for complementary

tests with newer theories, (e.g., item analysis according to the Rasch model), this demand was also followed here in the third study of the present work, and the psychometric properties of the PMH scale were examined with the aid of Rasch model. The present work ends with a summary reflection and an outlook on further research.

2. Background

2.1 Psycho-oncology

2.1.1 Definition and historical aspects

Psycho-oncology is a separate field of oncology, which includes different specialties such as medicine, psychology, psychiatry, psychosomatics, and sociology and deals with the experience and behavior as well as the social resources of cancer patients in connection with their cancer disease, treatment, and associated problems (Deutsche Krebsgesellschaft et al., 2014; Weis et al., 2007). Despite clinical care of patients and relatives, psycho-oncology contributes to collaborative research that ranges from the investigation of the significance of psychological and social factors for the development, early detection, diagnosis, treatment, rehabilitation, aftercare, and the entire course of cancer disease and their interactions over the life span to use and implement the corresponding findings in the prevention and during the continuum of the cancer illness, i.e., diagnosis, treatment, rehabilitation and palliative care (Deutsche Krebsgesellschaft et al., 2014; Holland, 2002; Weis et al., 2007).

Consider the magnitude and importance of the field, it is at first surprising that the beginning of psycho-oncology has emerged only since the 1970s (Breitbart & Alici, 2009; Holland, 2002). Due to the paucity of epidemiological knowledge and accompanying stigma of a cancer diagnosis, it was for an extended period, approximately until the 19th century, believed that informing a patient about the disease is immoral and inhumane, as the patient would get helpless and better be able to cope without the knowledge (Holland, 2002; Holland et al., 2015). At that time, a cancer diagnosis was automatically associated with death and led to seclusion, shame, and guilt for the affected person and the relatives (Bultz, 2016; Holland, 2002). With growing medical progress and opportunities in the early 20th century, such as the development of anesthesia and surgery improvement, it became possible to cure cancer if the tumor was detected early (Holland, 2002). However, it became essential to educate the public about the diagnosis, associated symptoms, treatment, and prevention, so 1913 public health campaigns began (Holland, 2002; Holland et al., 2015).

With growing public awareness and more options in research and medical treatment such as radiation in the 1900 - the 1920s and chemotherapy in the 1950s, there had

been an increasing interest in psychological responses to treatment and cancer disease (Breitbart & Alici, 2009; Holland et al., 2015). Having in mind the centuries of stigma also associated with mental illness and psychological treatments, it is not surprising that psycho-oncology, only developed until the 1970s (Breitbart & Alici, 2009; Holland et al., 2015; Kash et al., 2006).

In recent years, psycho-oncology has become increasingly established, and further research is rapidly moving forward with the growing understanding about the psychological aspects of cancer disease and the recognition of the importance of quality of life as well as the psychological impact of the cancer diagnosis on the experience and behavior of the affected persons and their relatives (Breitbart & Alici, 2009; Bultz, 2016; Kash et al., 2006). In this course, measurement instruments have been developed, such as the Distress Thermometer by Roth et al. (1998), to capture and examine the psychological distress associated with the cancer diagnosis (Kash et al., 2006). Overall, psycho-oncological care developed worldwide and became relevant among affected individuals and in research and politics. In order to continuously improve the care situation in Germany, action fields and goals were formulated in the National Cancer Plan (Bundesministerium für Gesundheit, 2017). One of the action fields is the further development of oncological care structures and quality assurance, in which the goal of cross-sectoral, integrated oncological care is described. This further development includes providing adequate psycho-oncological care for all affected individuals, improving identification of support needs, and ensuring inpatient and outpatient psycho-oncological care (Bundesministerium für Gesundheit, 2017). A brief overview of cancer and its psychological impact is provided in the following chapter before psycho-oncology treatment options are presented.

2.1.2 Cancer disease and its associations with mental health

Every year, almost 500,000 people are newly diagnosed with cancer in Germany (Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016). Cancer is a complex group of diseases, characterized by abnormal cells in any organ or tissue of the body, which grow uncontrollably and spread to other organs, i.e., metastasis (WHO, 2021). The uncontrollable growth is called a tumor. Classification of neoplasms into benign and malignant tumor is based on the growth pattern of the neoplasm (Zentrum für

Krebsregisterdaten im Robert Koch-Institut & Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., 2020). Overall, cancer refers to all malignant neoplasms, including lymphomas and leukemia. According to the WHO (2021), cancer is one of the leading causes of death worldwide, accounting for nearly 10 million deaths in 2020.

There are different forms of cancers, as they may differ in their locality and the severity of the disease. Table 1 shows a standard classification of cancer types based on the location according to the International Classification of Diseases 10th Revision (ICD-10) chapter II (WHO, 2019).

Table 1 Types of Cancer according to ICD-10

ICD-10 Code	Types of cancer
C00 - C14	Lip, oral cavity, and pharynx
C15 - C26	Digestive organs
C30 - C39	Respiratory and intrathoracic organs
C40 - C41	Bone and articular cartilage
C43 - C44	Skin
C45 - C49	Mesothelial and soft tissue
C50 - C50	Breast
C51 - C58	Female genital organs
C60 - C63	Male genital organs
C64 - C68	Urinary tract
C69 - C72	Eye, brain, and other parts of the central nervous system
C73 - C75	Thyroid and other endocrine glands
C76 - C80	Ill-defined, other secondary and unspecified sites
C7A - C7A	Malignant neuroendocrine tumors
C7B - C7B	Secondary neuroendocrine tumors
C81 - C96	Lymphoid, hematopoietic, and related tissue
D00 - D09	In situ neoplasms
D10 - D36	Benign neoplasms, except benign neuroendocrine tumors
D37 - D48	Neoplasms of uncertain behavior, polycythemia vera, and myelodysplastic syndromes

D3A - D3A	Benign neuroendocrine tumors
D49 - D49	Neoplasms of unspecified behavior

The most common types of cancer in Germany are prostate, lung, and colorectal in men, whereas breast, colorectal and lung cancer are most common among women and these are also associated with the most frequent tumor deaths (Zentrum für Krebsregisterdaten im Robert Koch-Institut & Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., 2020). It is currently assumed that the development of cancer is multicausal and results from the interaction between genetic and exogenous factors (Kusch et al., 2013; Weis et al., 2007). Several risk factors for cancer development have been identified, where health behavior is central and can be used for prevention. Changing lifestyle-related risk factors could prevent one-third to one-half of cancer diseases (Arem & Loftfield, 2017; Kusch et al., 2013). The modifiable risk factors are (Arem & Loftfield, 2017):

- Tobacco use
- Obesity
- Physical inactivity
- Alcohol use
- Virus and infections
- Exogenous Hormones

Numerous studies could not confirm the direct involvement of psychological factors (such as personality traits) in the onset or development of cancer, as it was initially assumed by research (Garssen, 2004; Kusch et al., 2013; Levenson & Bemis, 1991; Nakaya, 2014; Weis et al., 2007). Even if psychological factors do not play a significant role in the development of cancer, many psychological factors are associated with a cancer diagnosis, because cancer does not only refer to cell changes and the associated changes and limitations in one organ, but the disease and treatment affect the entire organism as well as the mind and the social environment of the affected person. Therefore, psycho-oncological care and research concepts involve the patient, the relatives, the social environment, and the treatment (Weis et al., 2007).

Coping with a cancer diagnosis, and mastering various disease and treatment associated tasks and changes can be a significant challenge for patients. Patients may

accordingly be burdened by multiple physical, psychological and social problems (Mehnert-Theuerkauf & Lehmann-Laue, 2019). Physical and mental distress are often the consequence of cancer and can deplete the quality of life, disease progression, and survival rates of patients (Chan et al., 2015; Grassi et al., 2017; Karakas & Okanli, 2014; Linden et al., 2012). The most common psychological consequences are anxiety and depression (Boyes et al., 2013; Bussmann et al., 2018; Mitchell et al., 2011; Mitchell et al., 2013; Niedzwiedz et al., 2019; Pitman et al., 2018). Therefore, these psychological factors are often a central focus for psycho-oncological treatment (Chan et al., 2015; Cook et al., 2018; Grassi et al., 2017).

2.1.3 Treatment options

In addition to the medical advances of the last century in cancer treatment, i.e., surgery, radiation treatment, chemotherapy, hormone treatment, psycho-oncological treatment options have also evolved. Psycho-oncological interventions are non-pharmacological interventions in which psychological and social work methods such as psychosocial counseling, psychoeducation, stress management training, psychotherapy, relaxation techniques are implemented to reduce psychological and social distress and improve the quality of life of cancer patients (Deutsche Krebsgesellschaft et al., 2014). Several meta-analysis and reviews demonstrated the effectiveness of psycho-oncological interventions in the reduction of emotional distress and improvement of quality of life (Akechi et al., 2008; Faller et al., 2013; Osborn et al., 2006; Piet et al., 2012; Rehse & Pukrop, 2003; Tatrow & Montgomery, 2006). However, such psycho-oncological care is currently not accessible to all patients with a cancer diagnosis in Germany. The S3 guidelines of psycho-oncological diagnosis, counseling and treatment of adult cancer patients (Deutsche Krebsgesellschaft et al., 2014) point out that psycho-oncological treatment should be oriented towards the needs of patients. In order to provide psycho-oncology care tailored to the needs of individual patients at the earliest possible stage of disease and to meet the requirements of the National Cancer Plan (Bundesministerium für Gesundheit, 2017), stepped psycho-oncology care programs are increasingly being established, e.g., stepped psycho-oncological care (Singer et al., 2017), integrated, cross-sectoral psycho-oncological care program (isPO) (Jenniches et al., 2020; Kusch et al., 2014).

2.2 Methodological aspects of research

Due to the heterogeneity of cancer and individual disease progression, the focus in recent years has increasingly been on patient-oriented medicine (Sinaiko et al., 2017). The requirement to care systems is also to provide the right patient at the right time with the proper care at the right place (Kusch et al., 2016; Kusch et al., 2013). Therefore, in order to ensure evidence-based patient-centered care, care providers were developing psycho-oncological care programs that can be used to realize quality-assured patient care; for example, continuous screening for mental stress, psychoeducation, and stepped psychosocial care are components of quality-assured patient care (Fann et al., 2012; Forsythe et al., 2013).

2.2.1 Health service research

Health services research is a multidisciplinary field of scientific investigations examining, i.e., whether interventions achieve the desired effect, even under conditions of reality of the health care system, and help identify how care can be improved (Lohr & Steinwachs, 2002). The objective of health services research is to link the knowledge generated by the fundamental and clinical research with the established clinical practice, to implement and to evaluate the entire process and its results and thus to provide new evaluated care concepts and includes the following approaches (Pfaff, 2003):

- Concept development, which is the development of innovative care concepts and structures
- Accompanying research, which is the implementation of these concepts under everyday conditions and accompanying evaluation
- Outcome research, which is the evaluation of the effectiveness of care models and programs by using, for example, patient-centered outcome measures (PCOMs) or patient-reported outcomes (PROs)

It is required that the assessment instruments of the outcome research have to be reliable, valid, easily interpretable, and change-sensitive (Glasgow et al., 2012;

Wyrwich et al., 2013). However, what assessment instruments are used in psycho-oncology?

2.2.2 Assessment of psychological distress in cancer patients

According to the National Cancer Plan (Bundesministerium für Gesundheit, 2017) and the S3 guidelines (Deutsche Krebsgesellschaft et al., 2014), psycho-oncological care should be oriented towards the needs of individual patients. Therefore, it is recommended to screen patients with cancer for psychological distress and psycho-oncological support needs (Jacobsen, 2007). Since patients suffer from psychological consequences of the cancer disease, but no clinical disorder has developed, the criteria for an ICD-10 diagnosis are not given. As a consequence, many patients would not be considered who would need psycho-oncological care. Appropriate screening instruments are needed to carry out needs-based diagnostics. There are several screening instruments to assess emotional distress in cancer patients, for example (Deutsche Krebsgesellschaft et al., 2014):

- Hospital Anxiety and Depression Scale – German version (HADS-D) (Herrmann-Lingen et al., 2011)
- Hornheide Screening Instrument (HSI) (Rumpold et al., 2001; Strittmatter et al., 2000)
- Distress-Thermometer (DT) (Roth et al., 1998)
- Questionnaire on Stress in Cancer Patients (QSC-R23) (Herschbach et al., 2003)
- Patient health questionnaire – Depression (PHQ-9) (Kroenke et al., 2001)
- Patient health questionnaire – Generalized Anxiety Disorder Screener (GAD-7) (Löwe et al., 2008; Spitzer et al., 2006)

Many assessment instruments showed satisfied high quality in terms of their psychometric properties and generalizability in screening for emotional distress in cancer patients (Vodermaier et al., 2009). According to the S3 guidelines, the HADS-D has the best evidence nationally and internationally and is recommended as the best screening method in cancer patients (Deutsche Krebsgesellschaft et al., 2014). The HADS assesses self-report distress and is an established tool for the assessing anxiety

and depression in cancer patients (Herrmann-Lingen et al., 2011; Mitchell et al., 2010; Vodermaier & Millman, 2011). The scale consists of 14 items with a total score (HADS-T) ranging from 0-42. Subscale scores for depression and anxiety may additionally be calculated. Higher scores on the HADS indicate more severe depression and anxiety. To identify patients with an increased need for psycho-oncological care and especially for depression symptoms in cancer patients, a sum score of HADS-T ≥ 15 can be used as the cut-off value (Jenniches et al., 2020; Mitchell et al., 2010; Vodermaier & Millman, 2011).

In addition to the benefit of early identification of patient support needs and individual clinical decision-making, the use of PROs, including screening instruments, is essential for outcome research and the accompanying investigation of the effectiveness of care programs or interventions, thus for clinical and research practice.

2.2.3 Effectiveness research

In outcome research, effectiveness is often considered a criterion for determining whether and to what extent the evaluated intervention can cause the desired effect under 'real-world' conditions (Meyer et al., 2014). Because much psychological research deals with mental components and treatment outcomes that are not directly visible, it is difficult to capture (Evans et al., 1998). For this purpose, it is crucial to select appropriate outcome measures. With increasing patient-centeredness in health services research, the use of PROs measures has risen over the past 40 years in clinical and research practice (Wyrwich et al., 2013). PROs comprise information about a patient's health status reported directly from the patient and provided a standardized method of capturing the patient's perspectives and experiences (Ahmed et al., 2012).

Another important criterion for evaluating of interventions in outcome research should be briefly mentioned at this point, namely efficacy. The purpose is the same as for effectiveness, only that the investigation takes place under different conditions. Efficacy determines whether the evaluated intervention achieves the expected results under controlled, ideal circumstances (Meyer et al., 2014). The gold-standard research design is the randomized controlled trial (RCT) for examining clinical efficacy. RCTs evaluate particular interventions compared to a placebo or other control group condition by testing for statistically significant differences at a group level (Ferguson et

al., 2002). As mentioned above, health services research, including psycho-oncology, examines interventions under conditions of reality of the health care system. However, typically statements of effectiveness are also based on the analysis of the statistical significance of interventions.

Statistical significance provides information about differences found concerning a lower probability level than expected if occurring by chance at a group level (Page, 2014). Put another way, the statistical methods by comparing means between groups, using distributions of scores of assessment measure before and after an intervention to identify whether the observed potential change is likely, not due to chance (Evans et al., 1998; Kendall et al., 2013). Thus, significance tests in controlled and well-conducted group studies provide meaningful evidence about the effects of specific interventions on a specific population (Bothe & Richardson, 2011). However, it is a misconception that statistical significance provides information about the strength of the relationship between intervention and outcome variable (effect size) or about clinically meaningful effects of an intervention (clinical significance) (Ferguson et al., 2002; Kraemer et al., 2003). Effect size is another group of statistics that provide information about the magnitude of average change and allow to sum and average the size of treatment effects (Lambert & Ogles, 2009). Considering the heterogeneity of individuals and their variability in the perception of and response to the benefits of an intervention, it becomes clearer that it is crucial to identify individual variation in clinical significance in health care. Statistically significant differences at the group level, even by considering the effect size measures, are insufficient for purposes of clinical outcome research because they yield no information about individual changes and, therefore, whether the treated individuals have returned to normal functioning or have made clinically meaningful changes (Bothe & Richardson, 2011; Jacobson et al., 1999; Jensen & Corralejo, 2017; Lambert & Ogles, 2009). In addition, the mere use of group-level significance in identifying responders to a particular treatment may lead to misclassification of patients as responders when they show no change at the individual level (Hays et al., 2018). This variation in interpretation of individual and group differences is due to the fact that each individual does not have the same experience in a change outcome (Guyatt et al., 2002). The most common methods are briefly mentioned below before presenting the study that critically examines the measurement methods of clinical significance.

Due to the gap created by the mere use of group statistical significance or effect sizes in an intervention or care program's effectiveness, there is an increasing call for the additional investigation of clinical significance in outcome research and evaluation (Lenz, 2021). In recent years, many different concepts have been developed to assess the clinical significance and are increasingly used to evaluate and improve psycho-oncological care programs (Bedard et al., 2013; Guyatt et al., 2002; Ogles et al., 2001). One of the most common and well-established concepts is the Reliable Change Index (RCI) by Jacobson and Truax (1991) (de Beurs et al., 2019). In RCI, it is possible to calculate the number of participants moving from a dysfunctional to a functional state using the combination of statistically significant change and clinically significant change (Kendall et al., 2013). According to this approach, it is possible to classify individual patients as 'recovered', improved', 'unimproved' or 'deteriorated' (Jacobson & Truax, 1991). Another commonly used group of methods is the Minimum/Minimal Important Difference (MID). The MID is defined as the "smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke et al., 1989, p. 408). Based on this, several methods have been developed, which often have different aims, making it difficult to decide which method should be used (King, 2011). A recent review compared the different MID methods and concluded that the different methods result in different estimates (Mouelhi et al., 2020).

Due to the different concepts, the question now arises which method should be used to investigate clinical significance. The importance of examining the significant change at the individual level is emphasized, especially in outcome research and evaluation (Hays, Brodsky, et al., 2005; Lenz, 2021). Nonetheless, in many studies, significance on an individual level is often not ensured (Breitbart et al., 2015). Following this scientific demand, the following study aims to critically compare two clinical significance methods and thus contribute to an appropriate choice of measurement method in psycho-oncology and clinical decision making.

2.3 Study 1 - Critical consideration of assessment methods for clinically significant changes of mental distress after psycho-oncological interventions

Critical consideration of assessment methods for clinically significant changes of mental distress after psycho-oncological interventions

Subtitle: Assessment of distress reductions

The study can be found under the following reference:

Vaganian, L., Bussmann, S., Gerlach, A. L., Kusch, M., Labouvie, H., & Cwik, J.C. (2020). Critical consideration of assessment methods for clinically significant changes of mental distress after psycho-oncological interventions. *International Journal of Methods in Psychiatric. 29*:e1821. https://doi.org/10.1002/mpr.1821

The following article is the manuscript version.

Abstract

Objectives: Considering the heterogeneity of cancer entities and the associated disease progression, personalized care of patients is increasingly emphasized in psycho-oncology. This individualization makes the use of measurements of individual clinically significant change important when studying the efficacy and effectiveness of psycho-oncological care. Two conceptualizations for the measurement of clinical significance are critically contrasted in this study: the Reliable Change Index (RCI) and the Minimal Important Difference (MID) method.

Methods: In total, 2121 cancer patients participated in the study and a subsample of 708 patients was reassessed about four months later. Psychological distress was measured using the Hospital Anxiety and Depression Scale (HADS-T). We evaluated two measures of clinical significance (RCI, MID) by comparing the respective numbers of improved, unimproved and deteriorated patients.

Results: Individually significant changes were observed with both methods, however determined rates of improvement differed substantially: MID (66.67%) and RCI (48.23%). Most importantly, according to MID 17.93% of patients were identified as being improved, although their respective improvements were not statistically significant and thus unreliable.

Conclusions: The benefits of RCI outweigh MID, and therefore, the RCI is recommended as a measure to assess change.

Key words: clinical significance, Hospital Anxiety and Depression Scale, minimal important difference, psycho-oncology, reliable change index

Introduction

Coping with the diagnosis of cancer, and mastering the associated tasks and changes, can be a significant challenge for patients. Physical and mental distress are often associated with cancer, and can deplete patients' quality of life, disease progression, and survival rates (Chan et al., 2015; Karakas & Okanli, 2014; Linden et al., 2012). The most common psychological consequences are anxiety and depression (Bussmann et al., 2018; Linden et al., 2012). Due to the heterogeneity of cancer entities and the individual disease progression, the focus in recent years has increasingly been on patient-oriented medicine (Sinaiko et al., 2017). Consequently, it is required of the care system to provide the right patient at the right time with the right care at the right place (Kusch et al., 2016; Kusch et al., 2013). In order to ensure evidence-based patient-centered care, care providers are also developing psycho-oncological programs that can be used to provide individualized quality-assured patient care, for example, continuous screenings for stress, psychoeducation, and stepped psycho-oncological treatments (Fann et al., 2012; Forsythe et al., 2013).

Usually, statements of effectiveness are based on the analysis of its statistical significance. On a group level, statistical significance gives information about differences found in terms of a probability level lower than would be expected if occurring by chance (Page, 2014). In controlled and well-conducted group studies, significance testing provides meaningful and necessary evidence about the impact of specific interventions on a given population (Bothe & Richardson, 2011). However, the disadvantages of analyses on a group level are that very marginal differences can be statistically significant if the sample size is large enough (Hays, Brodsky, et al., 2005; Hays et al., 2018). Furthermore, the results yield no information about individual change and, thus, cannot be used as an indicator of clinical significance (Bothe & Richardson, 2011; Lambert & Ogles, 2009). Similarly, when it comes to identifying responders to a particular treatment, the mere use of group-level significance can lead to misclassification of patients as a responder if they show no change on an individual level (Hays et al., 2018).

Therefore, in addition to statistical significance on a group level, the relevance of clinical significance and related concepts, are increasingly being used to improve change measurement and clinical decision-making. These approaches are also increasingly used for the assessment and improvement of psycho-oncological care

programs (Bedard et al., 2013; Guyatt et al., 2002; Ogles et al., 2001). The concept of clinical significance represents the assessment of significant change on an individual level. The methods used to accomplish this are either distribution-based or anchor-based approaches (Ogles et al., 2001; Page, 2014; Wyrwich et al., 2013).

Many different concepts have been developed to assess clinical significance. One of the leading concepts is the Reliable Change Index (RCI) (Jacobson & Truax, 1991). The definition of clinical significance is that a patient has returned from a so called dysfunctional (clinical) to a so called functional (healthy) state (Jacobson & Truax, 1991). To observe this, it is necessary to combine two criteria, a statistically significant change and the clinically significant change. Only based on both criteria, the individual change can be classified within defined categories. A patient is classified as "recovered", if the difference between the pre- and the post-test value is greater than the RCI (i.e., is statistically reliable), and if the post-test score has passed a predetermined cut-off point. Put another way, this classification may only take place if there is a statistically and clinically significant change. Accordingly, a patient is classified as "improved", if there is a statistically change, but the values did not pass the predetermined cut-off point. Thus, the patient's dysfunctional symptoms are still present subsequent to treatment. Furthermore, there is a category of patients who are classified as "unimproved" or "deteriorated". Unimproved means that patients revealed no statistical change, regardless of whether the cut-off point was crossed. Furthermore, patients who report a statistically significant worsening of symptoms are classified as "deteriorated".

However, beside the RCI, another concept of clinical significance, the minimal differences between two measurement points has been suggested, also known as Minimum/Minimal Important Difference (MID). Within this approach, the minimum significant difference or change for the patient should be represented as a score (Guyatt et al., 2002; Jaeschke et al., 1989; Revicki et al., 2006; Wyrwich et al., 2013). The statistical significance is not a requirement for the calculation (in contrast to the RCI) (Page, 2014). Firstly, the MID was defined as the "smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke et al., 1989, p. 408). The MID is calculated by the smallest significant difference between pre- and post-test value, which represents a "significant" change (Copay et al., 2007; Revicki et al., 2008; Revicki et al., 2006). In

order to define the significance of this change, the anchor-based approach compares the results with other measures using an external anchor or criterion, whereas distribution-based methods calculate the MID by using a measure of variability (Copay et al., 2007; Crosby et al., 2003). The advantage of the anchor-based methods is the comparison of results with an external anchor, while the advantage of distribution-based methods is that changes are presented free of random variations (Crosby et al., 2003). Each MID value for a given instrument may vary with regard to the studied population and the given context (Revicki et al., 2008). The aim of the MID is to provide feedback to the patient, as well as to the clinician, about the benefits and implications for further treatment. To balance the advantages and disadvantages of anchor-based and distributed-based approaches within the concept of MID, it is commonly recommended to calculate the MID using a combination of both methods (Bedard et al., 2013; Guyatt et al., 2002; Revicki et al., 2008). However, there is still no agreement which method or combination is the best (Guyatt et al., 2002). Because of the potential relevance of this decision for each individual patient, it is essential to make the right clinical decision. One study has already provided an overview about three different methods of clinical significance including standard error of measurement (SEM), standard error of prediction, and the RCI (Hays, Brodsky, et al., 2005). Note that the SEM is often used to calculate the MID (Ousmen et al., 2018). Whereas in the study of Hays, Brodsky, et al. (2005) the importance of examining the significant change at the individual level for improvement, consistency, or deterioration is emphasized, in many studies significance on an individual level is often not ensured (e.g. Breitbart et al., 2015). Furthermore, no recommendation was made with regard to which of the methods should be used. Because of the high relevance of clinical decision-making, this study aims to critically contrast the two measures of clinical significance (RCI, MID) based on the change in the symptoms of anxiety and depression by cancer patients due to psycho-oncological treatment.

Methods

Participants and procedure

Data collection took part as part of a standardized program of the Clinic I of Internal Medicine (Clinical Psychology) in cancer patients of the Centre of Integrated Oncology Cologne and from the region. All participants provided written informed consent. The data was collected at two measurement time points. The first measurement time point

(t1) was at the time of inpatient admission of the cancer patients and the second measurement time point (t2) was four months later. The questionnaire was handed out at t1 as part of the standardized care program (Kusch et al., 2014). At t2, the questionnaire was handed out again if patients stayed in the hospital or sent to patients by mail, if they were discharged from the inpatient unit before t2. In total, 2121 cancer patients (1643 women (77.5%)) with mean age of 53.02 (SD = 13.50) and mean Hospital Anxiety and Depression Scale total score (HADS-T) of 16.91 (SD = 8.56) participated in the study at t1 and 708 patients (582 women (82.2%); mean age of 53.23 (SD = 13.00) and mean HADS-T of 13.67 (SD = 7.88) filled out the questionnaire a second time at t2. The cancer diagnoses among the participants are presented in Table 2. All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The work was approved by the Ethics Commission of Cologne University's Faculty of Medicine (reference number 15-048).

Table 2 Percentage of cancer diagnoses among participants at t1 – first study

Cancer of	Percentage (%)
Breast	40.9
Lymphoid, haematopoietic and related tissue	18.2
Female genital organs	9.1
Digestive organs	7.0
Respiratory and intrathoracic organs	3.9
Eye, brain and other parts of central nervous system	3.3
Male genital organs	1.7
Thyroid and other endocrine glands	1.6
Skin	1.5
Urinary tract	1.4
Mesothelial and soft tissue	1.3
Lip, oral cavity and pharynx	1.1
Ill-defined, secondary and unspecified sites	0.8
Bone and articular cartilage	0.4
Residual category (including different forms of cancer)	7.8

Measures

Self-reported distress was measured by using the German version of the HADS (Herrmann-Lingen et al., 2011). The scale is an established tool for the assessment of anxiety and depression in cancer patients (Mitchell et al., 2010; Vodermaier & Millman, 2011). Furthermore, it is recommended as a screening tool for the measurement of psychological strain according to the S3-guidelines for the psycho-oncological management of adult cancer patients (Deutsche Krebsgesellschaft et al., 2014). The HADS is a self-rating questionnaire and consists of 14 items with a total score ranging from 0-42. It is also possible to calculate subscales scores, but in psycho-oncological contexts, patients often show combined and fluctuating manifestations of anxious and depressive symptoms, so that a global measure of the HADS-T can best represent the clinical situation (Herrmann-Lingen et al., 2011). Based on the HADS-T, cancer patients suffering from significant distress can be reliably distinguished from cancer patients without distress (Mitchell et al., 2010). Accordingly, the analyses of the present study will focus on the HADS-T score. There are different cut-off scores recommended for HADS-T. Specifically, in cancer patients a sum score of HADS-T $\geq$ 15 can be used as the cut-off value to identify patients with an increased need for psycho-oncological care and especially for depression symptoms (Mitchell et al., 2010; Vodermaier & Millman, 2011). The psychometric properties of the HADS indicate a reliable and valid instrument (Bjelland et al., 2002; Herrmann-Lingen et al., 2011). In the present study, Cronbach's alpha was excellent (HADS-T: α = .91).

Statistical analyses

Patients' characteristics are described by using means and standard deviations. Kolmogorov-Smirnov-test results indicated that the data was not normally distributed. Nonetheless, the statistical measure of change on group level was examined by multivariate analysis of variance for repeated measures, since the *F*-test is robust in a large sample if the assumption of normal distribution is not met. Tests were calculated two-tailed with an assumed significance level of $p < .05$.

For the analysis of clinical significance, only cancer patients who had higher values at the first measurement time point than a critical threshold of HADS-T $\geq$ 15 were included. The rationale for this decision was that the present study is focussed on the methodological comparison of two concepts of clinical change and whether these two concepts would influence clinical decision making differentially. We presumed that

individuals with relevant self-reported distress would more likely improve through psycho-oncological care allowing for the best possible comparison of the two concepts. Note, however, that we do not want to imply or indicate that patients with values under the cut-off threshold of ≥ 15 could not potentially benefit from psycho-oncological care. The calculation formula for RCI used here is: $RCI = \frac{Y-X}{Std\,(Y-X)} = \frac{Y-X}{Std(X)\sqrt{2(1-Rel(X))}}$

According to this formula, the RCI is calculated as the difference between the pre- (Y) and post-test (X) values divided by the standard measurement error of the difference (Jacobson & Truax, 1991). As a measure of reliability, α is recommended and will be used for the test instrument and the examined sample (Lambert & Ogles, 2009).

A RCI > 1.96 corresponds to a 95%-confidence interval and indicates that the individual change is statistically significant. In the present case, the change between the pre- and post-test values had to be greater than or equal to 5 to indicate a statistically significant change for each individual patient. To assess whether the change is also clinically significant, the post-value must pass a predetermined cut-off point, which separates the dysfunctional from the functional population (Jacobson & Truax, 1991). This cut-off point can be defined in three different ways (Jacobson & Truax, 1991; Lambert & Ogles, 2009). In the present study the most conservative cut-off point was used. This cut-off point is based on information from functional and dysfunctional population and is recommended to determine the significant clinical change if functional standards are known. We used the calculation formula for this cut-off point: $c = \frac{(SDpatient\;Mnonpatient)\,+(SDnonpatient\;Mpatient)}{(SDpatient+SDnonpatient)}$ (Lambert & Ogles, 2009). Thus, data from a healthy population was used to define the cut-off point (HADS-T: M = 9.45; SD = 6.80) (Hinz & Brähler, 2011). In consequence according to the calculation formula, the cut-off point of HADS-T = 16.52 was used in addition to the RCI to determine significant clinical change in the present study.

In order to define clinically significant change employing MID, studies were searched in which a MID was established for the HADS-T. We found one study that had calculated a MID of 1.5 for the HADS-T in a sample of patients with chronic obstructive pulmonary disease (Puhan et al., 2008). The MID of 1.5 was calculated using an anchor-based approach employing a linear regression analysis. Two self-report instruments that assessed the burden of disease (symptomatology and affect) were used as anchors. In the same study, an effect size of 0.5 SD unites of change score

was additionally used to estimate a MID of 1.17 for the HADS-T. In this case, the MID was estimated based on a distribution-based method as an alternative to the anchor based approach (Puhan et al., 2008). Investigators regularly consider an effect size of 0.5 *SD* units as an adequate estimate of clinically significant change (Walters & Brazier, 2003). Given that the HADS-T results only in positive integers, no actual difference results in using the anchor- or distribution-based method. More specifically, the MID between pre- and post-measurement always had to be ≥ 2-point change on the HADS-T.

Results

At t1 1251 of all 2121 patients (59.0%) exceeded a HADS-T score ≥ 15, indicating relevant distress; the mean HADS-T score of these patients was 22.61 (*SD* = 5.85). From all 708 patients who took part at both measurement time points, 396 patients (55.9%) who exceeded a HADS-T ≥ 15, reported a mean HADS-T at t1 of 22.08 (*SD* = 5.59) and at t2 of 17.40 (*SD* = 7.16). Significant differences between the group who took part at only t1 (group 1; 1413 patients) and the group who took part at both t1 and t2 (group 2; 708 patients) was found with respect to gender (a higher proportion of women in group 2; $\chi^2(1) = 13.68$, $p < .001$) and their respective HADS-T scores (higher scores in group 1; $t(2119) = 2.75$, $p = .006$) at measurement time point t1. There were no differences between the groups with regard to age ($t(2118) = -.488$, $p = .626$). Overall, there was a statistically significant improvement on the group level (HADS-T: $F(1,707) = 95.35$, $p < .001$; $\eta_p^2 = .119$) from t1 to t2 in all patients of group 2. Furthermore, patients who exceeded the critical threshold of the HADS-T score of ≥ 15 at the time of pre-examination showed also a significant change at t2 ($F(1,395) = 176.75$, $p < .001$; $\eta_p^2 = .309$), illustrating an even stronger effect of treatment in this subgroup of patients of group 2.

For the analysis of statistical and clinical significance on an individual level, firstly the RCI was calculated. Additionally, an external MID value was used as alternative reference. In order to better illustrate the effects of using the two different types of clinically significant change, in this analysis, only patients who reached or exceeded the critical threshold for psychological strain as measured by HADS-T scores ≥ 15 at the time of the pre-examination were included (i.e., the group of patients for whom psycho-oncological treatment is indicated).

According to RCI concept, 193 of 396 (48.74%) cancer patients exhibited a statistically reliable change of self-reported distress (improved; see Table 3). Statistically significant and clinically significant change in symptoms was seen in 191 cancer patients (48.23%) (recovered). In turn, 29 (7.32%) patients showed a reliable worsening of symptoms (deteriorated) and 174 (43.94%) patients showed neither statistical nor clinically significant changes (unimproved). The estimated change of ≥ 2 points on the HADS-T was used as a cut-off point to assess the clinical change using the MID concept. For HADS-T, 264 of the 396 analysed patients (66.67%) achieved a change of at least ≥ 2 points in HADS-T between t1 and t2. However, among the 264 patients, 71 patients (17.93%) did not meet the criteria of a reliable statistical change and 73 patients (18.43%) of significant clinical change according to the concept of the RCI. Table 3 illustrates this pattern of results using a crosstab to give an overview on patients' changes with regard to depressive and anxiety symptoms as determined by using the RCI and MID concept.

Table 3 Frequency and percentage of different change based on RCI and MID for cancer patients for HADS-T from pre- to post-examination – first study

	MID - Deteriorated (pre-post difference ≤ -2)	MID - Unimproved (2 > pre-post difference > -2)	MID - Improved (pre-post difference ≥ 2)	Total
RCI - Deteriorated (pre-post difference ≤ -5)	29 (7.32%)	0	0	29 (7.32%)
RCI - Unimproved (5 > pre-post difference > -5)	39 (9.85%)	64 (16.16%)	71 (17.93%)	174 (43.94%)
RCI - Improved (pre-post difference ≥ 5)	0	0	193 (48.74%)	193 (48.74%)
Total	68 (17.17%)	64 (16.16%)	264 (66.67%)	396 (100%)

Discussion

Before interpreting the data, it should be called to mind that the aim of this study is not the evaluation of a specific psycho-oncological treatment package, but rather a comparison of two commonly used methods of determining clinical significance. The efficacy of treatment is not emphasized in the present study, so the relatively high dropout rate and the missing control-group are not exceptional given the naturalistic sample. Thus, this study aimed to critically contrast two concepts for the measurement of clinical significance (RCI and MID) within a sample of cancer patients with respect to the identifiable ratio of individual improvement and deterioration in symptoms between the methods.

Overall, the patients with signs of substantial psychological stress before treatment showed significant improvement in symptoms of anxiety and depression. These results are in line with other studies, which previously highlighted the potentially highly positive effects of integrative psycho-oncological treatment on anxiety and depression in cancer patients (Grassi et al., 2017; Kost et al., 2009). To transfer these group-level results to measurements of clinical relevance for each individual patient, RCI and MID were calculated and critically compared. Both measurements of clinical change supported the claim of patient-oriented psycho-oncological care as being efficacious. Based on the RCI concept, 48.74% of patients exhibited a reliable statistical change (improvement) with regard to symptoms of anxiety and depression. The results of the analysis are in line with other studies focussing on individual case analyses (Grassi et al., 2017; Kost et al., 2009). In addition, 48.23% of patients reported a statistically and clinically significant change (recovered) based on their scores on the HADS-T. On the downside, 29 (7.32%) patients reported statistically increased levels of depression and anxiety.

The estimation of change and clinical significance was also performed employing the MID. According to the MID concept, 66.67% of patients showed a clinically meaningful improvement. Thus, compared with the RCI, the estimated number of patients with clinically significant improvements was clearly higher when using MID. However, 71 patients (17.93%) that supposedly had improved, did not meet the requirements for a statistically reliable change. Moreover, 73 patients (18.43%) did not meet criteria for significant clinical change according to the RCI concept (i.e., rather belong to a healthy as compared to a psychologically stressed population). In addition, 39 patients (9.85%)

were classified as deteriorated according to MID. Although again, this numerical change was not statistically reliable. Thus, on the one hand, the MID value as a cut-off point appears less stringent than the RCI's cut-off points, and on the other hand it is not ensured that identified changes are indeed reliable changes and consequently there is a substantial risk of overestimating the results when studying the effects of an intervention. Interestingly, using MID did not result in a larger group of individuals being detected whose psychological stress had deteriorated. Thus, the lower threshold of the MID only resulted in a more liberal detection of improvement and not of deterioration.

An important advantage of RCIs compared with the MID is that statistical significance of each individual change is a requirement for determining clinical significance. Or with other words, clinical significance is always statistically safeguarded (Lambert & Ogles, 2009). Also, the calculation of an individuals' clinical significance using the RCI concept is quick and efficient (Lambert & Ogles, 2009). A criticism is that the RCI is a rather conservative method and the cut-off points used are relatively strict criteria. Individuals with low initial symptom severity have little chance of undergoing a clinically relevant change as they have little room for improvement (Lambert & Ogles, 2009). However, we believe that being conservative when determining the benefit of an intervention is a merit rather than being problematic. Before ending treatment, it should be safely ensured that a person indeed has improved rather than discharging a patient that only ostensibly has improved. Another potential problem with the MID was not much of a problem in our sample due to the instrument we used. There are a number of different definitions for the calculation of the minimal important difference, which often have different aims (King, 2011). For the approach of the anchor-based methods as well as for the distribution-based methods, different authors suggest different types of calculations (King, 2011; Revicki et al., 2008). In a recent structured review, Ousmen et al. (2018) reported that the most commonly used distribution-based method was the 0.5 *SD*, followed by the SEM. Note that as the best approximation, a combination of both approaches is recommended (Bedard et al., 2013; Guyatt et al., 2002). However, this approach rarely used in practice (Ousmen et al., 2018). Furthermore, Hays, Farivar, et al. (2005) highlighted that despite the fact that several measures (e.g. SEM) are related to the MID, these measures do not provide direct information about the MID (e.g., new information about the size of change) and that their use should therefore be discouraged. In our sample, given that only integer numbers are calculated as individual scores on the HADS-T, these different MID scores resulted always in the

same HADS-T change score of 2. Thus, in the present study it seemed rather inconsequential, which method to use.

Study limitations

Beside the strength of large sample size, examination on patients' reports and the benefits for clinical and research practice, the study also has limitations. One limitation is that the difference from pre- to post-examination does not necessarily have to be related solely to psycho-oncological treatment. Frequently, patients respond to cancer diagnosis with signs of anxiety and depression, which often remit spontaneously without the need for psycho-oncological support (Cook et al., 2018). To examine this aspect in a study with an additional randomized control group would be interesting and important. Additionally, the data showed a very high drop-out rate of approximately 66.6% in the period between pre- (2121 cancer patients) and post-examination (708 cancer patients). This can primarily be explained by the fact that the data collection did not take place as part of a research project, but within routine care. It is possible, that patients switched to outpatient care, were no longer in treatment or even died. This was possibly especially likely because the data was collected in a highly specialized treatment facility, recruiting patients from a relative wide catchment area. Nonetheless, there are differences between the patients who dropped out and who took part at both measurement time points with regard to gender and the initial HADS-T score. To control for these differences would be interesting and important in further studies. Note, however, that the differences on the HADS-T are significant but that the mean HADS-T values of the drop-out group (M = 17.27, SD = 8.69) and the non-drop-out group (M = 16.18, SD = 8.24) nonetheless had higher values than the critical threshold of HADS-T ≥ 15. Arguably, the self-reported psychological distress was clearly relevant in both groups. Moreover, this group difference had no direct influence on the analysis of possible differences between the individual change measurements.

Clinical implications

Concepts to determine the clinical significance of treatments are increasingly being used for adequate change measurement and clinical decision-making. These approaches are also increasingly used in the assessment and improvement of psycho-oncological care. Therefore, it is important to critically assess which method can be used for a clinical decision. Although the MID ostensibly shows a higher percentage of improvements, it is statistically unreliable and as a consequence of its use an

overestimation of the effects of this form of intervention is possible. In addition, errors in clinical decision-making may result if patients' treatments are ended prematurely. Due to these weaknesses of the MID and the substantive advantages of the RCI, the RCI is recommended as a measure of change in the care research of cancer patients.

2.4   Psychometric properties of assessment instruments

Outcome research in health service research investigates the effectiveness of care models and programs by using PROs. The assessed PROs should be examined with reliable, valid, easily interpretable, and change-sensitive assessment instruments. The previous chapters have dealt with the question of which measurement instruments are used in the clinical practice of psycho-oncology and how the possible effectiveness of interventions on PROs can be investigated (i.e., what should be tested, statistical and clinical significance). The following chapters deal with the methodological question of how the psychometric quality of an instrument is evaluated or which other adequate assessment instruments can be used in psycho-oncological clinical and research practice.

Psychological assessment instruments aim to measure a latent psychological trait or construct, e.g., individuals' ability or personality traits, and draw inferences about the latent traits from the empirically obtained test results (Hambleton & Jones, 1993; Ziegler & Hagemann, 2015). For this purpose, it is essential that the quality or the psychometric properties of the respective outcome instruments are scientifically robust, i.e., met the standards and criteria of measurement science (Tennant & Conaghan, 2007). The most relevant criteria for evaluating health assessment and in clinical practice are (Souza et al., 2017):

- Reliability, which is a measure of formal accuracy and includes the aspects of stability, internal consistency, and equivalence
- Validity, which describes the property of an assessment instrument that it measures what it purports to measure and includes aspects of content, criterion, and construct validity

Testing of assessments psychometric properties is based on several theories and different methods, including classical test theory (CTT), item response theory (IRT), and sometimes as a separately named theory, Rasch measurement theory (RMT), which are briefly introduced below.

2.4.1 Classical test theory (CTT)

CTT assumes that the test result directly corresponds to the true degree of expression of the examined characteristic by adding the scores of different items of a measurement instrument (Petrillo et al., 2015). In addition, the measurement or test result is overlaid by measurement errors (e.g., lack of concentration, inappropriate items), which according to CTT, means that the observed test score (X) is based on the true score (T) plus measurement error (E) (Hambleton & Jones, 1993; Hays et al., 2000):

$$X = T + E$$

To determine the two unknown unobservable or latent variables T and E, assumptions are made in the CTT (Hambleton & Jones, 1993). First, it is assumed that the measurement error of an item is independent of the true score, and second, that the sum of the measurement errors of all items approaches zero (Hambleton & Jones, 1993; Streiner, 2010). This practically leads to the larger the number of items in the instrument or scale, the smaller the measurement error of the scale is (Streiner, 2010). Since the test scores represent the true value, there is a monotonically increasing relationship in the sense that the responses to the items reflecting the same latent trait increase/decrease, implying a higher/lower expression of the trait, depending on the coding of the items (Cappelleri et al., 2014). The test scores are treated as interval scale level, so it is possible to calculate sum scores and mean values, which does not correspond to reality (Streiner, 2010).

However, as CTT has evolved, different applicable models for test development and evaluation have emerged. The advantages of CTT models are that the assumptions are easy to satisfy in real test data and that the methodology is simple to implement. But there are also drawbacks to the theory and its methodology. First of all, one disadvantage is, that both person parameters (i.e., true scores) and item parameters (i.e., item difficulty and discrimination) are dependent on the test or subject sample (Hambleton & Jones, 1993). Consequently, if the scale is to be used with a different population, it must be renormed and the psychometric properties determined for this new population; likewise, any change in the instrument requires a reevaluation of the psychometric properties (Streiner, 2010). Another limitation of CTT is that the obtained

scores are completely test dependent. Therefore, it is difficult to compare the results of two different tests because they are not on the same scale, and it assumes equal measurement errors at all ability levels (Hambleton & Jones, 1993). Another disadvantage is the equivalence assumption, because this implies that all items contribute equally to the total score, or that a total score can thus be formed by summation. However, it is assumed here that the items are measured on an interval scale, which is rarely, if ever, the case (Streiner, 2010). These limitations and potential problems arising from the simple assumptions may lead to an inaccurate representation of the examined outcome results and to results not being comparable. An extensive description of the advanced methodology is given elsewhere (e.g., Crocker & Algina, 2006; Rost, 2004). Overall, CTT can be described as a long dominating, quantitative approach to testing the reliability and validity of a scale (Cappelleri et al., 2014).

2.4.2 Item response theory (IRT) or/and Rasch measurement theory (RMT)

To address the disadvantages of CTT, a different approach, IRT, was developed. An essential component of the theory is the use of stochastic models to derive statistical estimates of parameters, i.e., the location of persons and items on a latent continuum (Petrillo et al., 2015). Overall, IRT is a group of measurement models that explain the connection between the response items and the person location of an underlying latent trait (Hays et al., 2000). They describe the probability with which a person solves the item depending on the person parameter, i.e., the person's true ability, and the item parameter, i.e., the item's difficulty (Strobl, 2015). Item difficulty refers to the probability of endorsing an item or scoring high on an item in terms of the trait being measured (Hays et al., 2000). The item difficulty is reflected in item characteristic curves (ICC). The curves provide information about the discrimination and difficulty of an item (Cappelleri et al., 2014). The main assumptions of IRT are the unidimensionality of a scale, which means the scale exclusively assesses one underlying construct, and local independence of the items, which means there is no correlation between items when extracting the trait factor (Tennant & Conaghan, 2007). In IRT, no distribution assumptions (e.g., normal distribution) are necessary as in CTT since the estimation

procedure is not based on mean values and variances, but on frequency ratios of the response categories to each other (Strobl, 2015).

Several IRT models differ in terms of the number of parameters (1-, 2-, or 3-parameters) and in terms of the given item response format of the scale, i.e., dichotomous or polytomous items (Cappelleri et al., 2014). There is often heterogeneous terminology in the literature about whether Rasch measurement theory (RSM) is part of item response theory or should be treated separately. The main difference is that IRT typically describes a data set like CTT does, and the model must fit the data, whereas RMT has the primacy that the data should fit the model (Petrillo et al., 2015). The further theoretical discussion of this is not discussed in detail here.

However, the item analysis according to Rasch or often called Rasch model was conceived by the Danish mathematician Georg Rasch and is the simplest model within IRT (Gustafsson, 1980). Initially designed for dichotomous items, the Rasch model is usually classified as a 1-parameter model within IRT and has been further developed for polytomous items (Cappelleri et al., 2014). This approach is used to investigate psychometric properties in study 2 and study 3 of the present work, and further information about the relevant assumptions and used methodology is described in the manuscript version below.

There are several advantages by using IRT and the resulting information, which are summarized in the following. An advantage lies in a more comprehensive and accurate evaluation of PROs and the item characteristics (Hays et al., 2000). With the investigation of the fit statistic of each item, it is possible to make statements whether the items represent the latent trait and form a unidimensional scale and where the items on the scale are located (Streiner, 2010). Furthermore, it is possible to evaluate the response categories with the analysis of ordered or disordered thresholds and examine whether they are equally spaced. Item thresholds are the transition points between two adjacent respond categories. IRT assumes that measurement accuracy depends on where items are located and whether they provide enough information to discriminate between individuals, which contrasts with CTT, which assumes that measurement accuracy is constant across scales (Petrillo et al., 2015). Another advantage is the investigation of differential item functioning (DIF). If DIF is given, it means that, in different groups, the corresponding item indicates the latent

characteristic in different ways (Tennant & Conaghan, 2007). An extensive description of the advanced methodology and the various measurement models is given elsewhere (e.g., Andrich, 2011; Rost, 2004; Strobl, 2015). Overall, the IRT models are statistically more sophisticated and the assumptions more stringent than those of the CTT, and therefore the statements about the psychometric properties are more informative (Hambleton & Jones, 1993; Streiner, 2010).

2.4.3 Relevance for (psycho-oncological) assessment methods

In general, test theory is concerned with the relationship between the latent trait and the observed test behavior and the requirements that a test must meet in order to draw conclusions about the expression of the tested trait based on a test result (Hambleton & Jones, 1993). The theories addressed in the previous subsections address these issues and are applied to both construction and evaluation of assessment methods. A direct comparison between the three approaches and additional background can be found elsewhere (e.g., Petrillo et al., 2015). Simply put, CTT models are weaker in their assumptions and more effortless to satisfy by test data than IRT, including the Rasch model (Hambleton & Jones, 1993).

If IRT models offer many advantages over the CTT approach, it begs the question of why they are not more widely used and why there are still not many test developments and evaluations of psychometric properties using this approach. Streiner (2010) formulates hypotheses in this regard: First, human inertia, which is related to the fact that most learn CTT in college, and IRT has been slow to establish itself in college over the past 20 years. Second, CTT is much easier to understand. Also, most statistical software packages have all the modules to compute CTT models, whereas computation of IRT models must be implemented in much more complex and expensive programs. Nonetheless, Hays et al. (2000) highlight that the use of IRT methods in measuring health outcomes is growing rapidly in the 21st century and will lead to expansions in the utility and improvements of the methodology as experience increases.

Overall, it should be understood that it is not a matter of using one method or the other, as it seems much more instructive to see the approaches as complementary to each other in order to avoid possible biases or disadvantages of one method or the other.

However, recently, there has been an increasing call and use of IRT models in the development or evaluation of clinical assessment instruments (Thomas, 2019). This change also resonates with the few direct comparisons of CTT and IRT approaches available to date (e.g., Bjorner, 2019; Blanchin et al., 2011; Jabrayilov et al., 2016; Petrillo et al., 2015; Prieto et al., 2003) and for the oncology literature. For example, in a systematic review of PRO measures in oncologic breast surgery, Chen et al. (2010) emphasized the importance of new psychometric methods, i.e., IRT or Rasch models, in developing and validating scales for use in individual patient care as well as for group-level comparisons. However, following this trend and call, the following study examines the psychometric properties of an assessment instrument that is also relevant in psycho-oncology care using the methods of the Rasch model.

2.5 Study 2 - An Item Analysis according to the Rasch Model of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0)

An Item Analysis according to the Rasch Model of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0)

The study can be found under the following reference:

Vaganian, L., Bussmann, S., Boecker, M., Kusch, M. Labouvie, H., Gerlach, A. L., & Cwik, J. C. (2021). An Item Analysis according to the Rasch Model of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0). *Quality of Life Research*. http://doi.org/10.1007/s11136-021-02872-8

The following article is the manuscript version.

Abstract

Purpose: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) assesses disability in individuals irrespective of their health condition. Previous studies validated the usefulness of the WHODAS 2.0 using classical test theory. This study is the first investigating the psychometric properties of the 12-items WHODAS 2.0 in patients with cancer using item analysis according to the Rasch model.

Methods: In total, 350 cancer patients participated in the study. Rasch analysis of the 12-items version of the WHODAS 2.0 was conducted included testing unidimensionality, local independence, and testing for differential item functioning (DIF) with regard to age, gender, type of cancer, presence of metastases, psycho-oncological support, and duration of disease.

Results: After accounting for local dependence, which was mainly found across items of the same WHODAS-domain, satisfactory overall fit to the Rasch model was established (χ^2 = 36.14, p = 0.07) with good reliability (PSI = 0.82) and unidimensionality of the scale. DIF was found for gender (testlet 'Life activities') and age (testlet 'Getting around/Self-care'), but the size of DIF was not substantial.

Conclusion: Overall, the analysis results according to the Rasch model support the use of the WHODAS 2.0 12-item version as a measure of disability in cancer patients.

Introduction

About 15% of the world's population live with some form of disability (WHO & The World Bank, 2011). According to the WHO, a person's functioning and disability are best described by a dynamic interaction between contextual factors and health conditions (WHO, 2001). In addition to establishing a patient's diagnosis, it is necessary to assess the overall condition in particular areas of life (i.e., the disability of a patient with regard to home tasks, work or other social areas) in order to ensure sound clinical decision-making and selection of appropriate interventions for patients (Üstün et al., 2010). Since disability can affect many life areas, it is difficult to ensure a suitable, reliable and valid measurement of its impact on the live of a person.

In 2001, the WHO developed the International Classification of Functioning, Disability, and Health (ICF) and defined disability as "an umbrella term for impairments, activity limitations or participation restrictions" (WHO, 2001, p. 3). Based on the ICF, the WHODAS 2.0 was developed to provide a standardized method for measuring health and disability (Üstün et al., 2010). The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (APA, 2013), recommend the WHODAS 2.0 as "the best current measure of disability for routine clinical use" (Gold, 2014, p. 174).

The scale is an established tool for the assessment of functioning difficulties in six domains (cognition, mobility, self-care, getting along, life activities, and participation). It has been developed for individuals with any kind of disease and is available in three different length regarding the number of items (12, 12 + 24, and 36 items) and as interview-, self- or proxy-administered versions (Üstün et al., 2010). Usage of the WHODAS 2.0 is continuously increasing, and it is available in 47 languages and dialects (Federici et al., 2017). It has been validated for different health conditions, for example, depression (Luciano, Ayuso-Mateos, Fernández, et al., 2010), multiple sclerosis (Magistrale et al., 2015), or myocardial infarction (Kirchberger et al., 2014).

Cancer patients have to cope with their diagnosis and master the disease-associated tasks and changes. In addition, they may also suffer from disability. The disabilities experienced by cancer patients can differ substantially, due to the heterogeneity of cancer entities and individual disease progression. Thus, it is pertinent to consider the application of the WHODAS 2.0 in the oncological context as well.

Research studying the psychometric properties of the WHODAS 2.0 in an oncological context is rare. Only few studies exist based on classical test theory (CTT), which showed good to excellent reliability, good convergent and discriminant validity, and supported the 6-domain structure (Norouzi et al., 2020; Pösl et al., 2007). However, within Chinese breast cancer patients a 7-domain structure was identified (Zhao et al., 2013). An advantageous alternative to CCT is item analysis according to the Rasch Model, which can be used to assess the unidimensionality of the items, sampling invariance, and local dependence problems (Fischer, 1987; Gustafsson, 1980) According to Rasch (1965), this must be re-examined for each new population the measure is applied to (as cited in Gustafsson, 1980). Studies employing Rasch analysis on the WHODAS 2.0 have looked at different health conditions like myocardial infarction, stroke, osteoarthritis, depression, and brain injuries (Kirchberger et al., 2014; Küçükdeveci et al., 2013; Kutlay et al., 2011; Luciano, Ayuso-Mateos, Aguado, et al., 2010; Snell et al., 2020). These studies confirmed the assumption of unidimensionality for the 36 item version as well as 12 items short version of the WHODAS 2.0.

However, to the best of the authors' knowledge, Rasch analysis has not yet been applied to the WHODAS 2.0 in an oncological context. That is why this study aims to examine the applicability of the 12-items version of the WHODAS 2.0 among patients afflicted by various types of cancer with the aid of Rasch-analysis, especially to investigate the assumptions of unidimensionality, invariance across different exogenous variables, local independence of items, and the targeting.

Method

Participants and procedure

Participants were invited to participate in the study using SoSciSurvey (Leiner, 2014) as an online survey consisting of various questionnaires. The link was posted on social media platforms and online cancer support groups as part of a validation study (Cwik et al., 2021). All participants gave their informed consent online. Inclusion criteria were: age ≥ 18 years and current or in the past cancer diagnosis. Exclusion criteria were not defined. In total, $N = 350$ cancer patients (283 women (80.9%), 66 men (18.9%), 1 gender diverse (0.3%)) completed the 12-items version of the WHODAS 2.0.

We received the permission of WHO for utilization of the WHODAS 2.0 (License: CC BY-NC-SA 3.0 IGO). All procedures contributing to this work comply with the relevant national and institutional committees' ethical standards on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The work was approved by the Ethics Commission of the University's Faculty of Medicine (reference number 18-098).

Assessment instruments

WHODAS 2.0. Global health status was assessed using the German version of the 12-item self-administered version of the WHODAS 2.0 (Üstün et al., 2010). The scale is an established and validated tool for the assessment of functioning difficulties in six domains (understanding and communicating, mobility, self-care, getting along, life activities, and participation). The participants estimate how many difficulties they have had in performing various activities in the last 30 days on a 5-point Likert-type scale (none = 0, mild, moderate, severe, extreme/ cannot do = 4). Higher scores reflect a more significant disability (Üstün et al., 2010).

Statistical analyses

Data were analyzed using SPSS version 26.0 (IBM Corporation, Released 2019) and RUMM2030 software (Andrich et al., 2009). Patients' characteristics are described by means and standard deviations. One item is missing from one patient of the 12-item WHODAS version, which was replaced by using the mean of the other items as recommended by Üstün et al. (2010).

Item analysis methodology according to the Rasch model was used to assess the psychometric properties of the WHODAS 2.0 in an oncological context. This model allows a nuanced analysis of an instrument's psychometric properties because it focusses on the items and how persons respond to them. Person parameters are estimated, which express the individual extent of a latent trait, which in the case of WHODAS 2.0 is disability (da Rocha et al., 2013). Likewise, on the same latent trait, the item difficulty parameters are estimated. 'Easy' WHODAS-items would be items that are already scored high in the direction of disability by patients with only minor disabilities, whereas 'difficult' WHODAS-items would be items that are only scored high by patients with major disabilities. During the process of the item analysis according to the Rasch model, it is tested whether patients respond as expected to each item. For

example, a patient with major disabilities should also score high on an 'easy' WHODAS-item. In order to properly test the fit of the WHODAS-data to the Rasch model, this paper follows the current state-of-the-art Rasch analysis requirements (Tennant & Conaghan, 2007) and the CREATE guidelines for reporting valuation studies (Xie et al., 2015).

Given the polytomous WHODAS-items, the Partial Credit Model (PCM) (Masters, 1982) was used. According to the Rasch model, performing analysis comprises the investigation of how well the data meet the expectations of the measurement model, i.e., unidimensionality, local independence, and absence of differential item functioning (DIF). In this sense, the analysis according to the Rasch model can be understood as an iterative process in which potential deviations from the model's expectations are investigated and – if possible – resolved.

One fundamental requirement of the Rasch model is unidimensionality, i.e., the items of a scale should assess only one underlying construct. Unidimensionality was tested with principal component analysis (PCA) of the residuals (Pallant & Tennant, 2007). The idea is to use the items with the highest negative/positive loadings on the first component to create two subsets of items. The separate person estimates of these two subsets are used to identify significant differences using independent t-tests. The proportion of significant t-tests should not exceed 5% to confirm unidimensionality (Smith, 2002).

Another assumption is that of local independence. This implies that there should be no residual correlations between items when extracting the trait factor (Tennant & Conaghan, 2007). Locally dependent items respectively items which are linked in some way, can lead to overestimation of reliability, parameter estimation bias, and problems with construct validity (Christensen et al., 2017). Following the recommendations of Christensen et al. (2017) and Marais (2013), a cut-off value of 0.2 above the average residual correlation was used to assess local dependence (LD). One strategy to deal with LD if one does not want to delete scale items is to combine the locally dependent items into testlets by adding them together. Using the testlet-strategy results in a bi-factor equivalent solution. The proportion of explained common variance (ECV) (Andrich, 2016; Pomeroy et al., 2020; Rodriguez et al., 2016) of the general factor should be >0.9 to consider the scale as unidimensional. The ECV is indicated in

RUMM2030 as A-factor (Pomeroy et al., 2020). One more assumption is that there is no item bias with regard to exogenous variables (no DIF). If DIF is given, the difficulty of an item is different for different groups (e.g., males and females). In other words, in different groups, the corresponding item indicates the latent characteristic in different ways (Tennant & Conaghan, 2007). DIF analyses were examined using analysis of variance (ANOVA). We tested the items for DIF by looking at gender (woman, man), age (median split of the sample: below and above 54), type of cancer (breast, other forms of cancer, multiple cancers), presence of metastases (yes, no, unknown), psycho-oncological support (yes, no) and duration of disease (median split of the sample: below and above 3.9 years). In case of DIF, we evaluated the impact of DIF by computing equated scores (Christensen et al., 2019). Due to too small group sizes, we had to exclude the one gender diverse person for the DIF analysis of gender and combine the residual cancer types into one category, 'other forms of cancer' for the DIF analysis of cancer type.

Additionally, item fit as indicated by standardized residuals within a range of ± 2.5 and overall model fit indicated by a non-significant Chi-Square probability $p > 0.01$, were investigated (Pallant & Tennant, 2007; Siegert et al., 2010). Moreover, the ordering of item thresholds was analyzed. Item thresholds are the transition points between two adjacent respond categories. Disordered thresholds may affect scale scores' interpretation and validity (Andrich, 2013). There can be different causes for threshold disordering, such as that the respondents might have difficulties consistently differentiate between the different response options, or LD might cause the disordering. If the disordering is due to category differentiation problems, one way to handle this is by collapsing the disordered response categories.

The scale's internal consistency was estimated using Person Separation Index (PSI). The PSI is equivalent to Cronbach's alpha and can be interpreted similarly with a requirement of a minimum value of .7 for group and .85 for individual use (Tennant & Conaghan, 2007). Targeting was assessed graphically based on the person-item threshold distribution graph. Person-item maps demonstrate how person parameters and item thresholds are distributed along the trait dimension.

Results

Mean age of the $N = 350$ participants was 52.34 years ($SD = 14.07$) and all participants completed the WHODAS 2.0 questionnaire. A selection of descriptive statistics and an overview of cancer diagnoses among the participants are presented in Table 4.

Table 4 Characteristics of cancer patients ($N = 350$) – second study

Gender	
Male	66 (18.9)
Female	283 (80.9)
Divers	1 (0.3)
Age (in years)	52.34±14.07 (20-83)
Job situation	
Active	146 (41.7)
Certified sick	56 (16.0)
Different form	148 (42.3)
Types of cancer	
Breast	182 (52.0)
Urological	37 (10.6)
Prostate, testicular	33 (9.4)
Gynecological	29 (8.3)
Hematological	26 (7.4)
Intestinal, rectal	20 (5.7)
Skin	13 (3.7)
Lungs, Bronchia	10 (2.9)
Ear, Nose, Throat	7 (2.0)
Gastric, esophageal, pancreatic	7 (2.0)
Parts of central nervous system	5 (1.4)
Soft tissue	3 (0.9)
Residual category (including other forms of cancer)	29 (8.3)
Metastases	

No	260 (74.3)
Yes	78 (22.3)
Unknown	12 (3.4)
Current psycho-oncological, psychological, psychotherapeutic support	
No	251 (71.7)
Yes	99 (28.3)
HADS-T Score - Distress (HADS T ≥15)	154 (44.0%)

Values are presented in frequency (%) or mean±standard deviation (range). HADS-T = Hospital Anxiety Depression Scale (Herrmann-Lingen et al., 2011) (To identify patients with an increased need for psycho-oncological care and especially for depression symptoms in cancer patients, a sum score of HADS-T ≥ 15 can be used as the cut-off value) (Jenniches et al., 2020)

The initial analysis of all 12 items of the WHODAS 2.0 showed a satisfactory overall model fit (χ^2 = 88.21, p = 0.01). However, several items displayed LD, two items showed item-misfit, DIF was found for items 1 and 12 in relation to age, for items 7 and 11 in relation to gender, and for item 12 related to disease duration and disordered thresholds in six items. In the initial analysis, LD was found for item 1/2/7, item 3/6, item 7/8, item 8/9, and 10/11.

As LD seemed to be the major problem, we focused at first on accounting for it. We stepwise combined the locally dependent items with the highest residual correlation, starting with the item pair 8 and 9 (r = 0.554; critical-LDvalue = 0.1). These successively combined locally dependent items were consistent with the six WHODAS 2.0 domains (in order of testlet formation: 'Self-care', 'Getting along with people', 'Getting around', 'Understanding and Communicating'', Life activities' and 'Participating in society'). After combining the two items of each domain into one testlet, LD was still present between the domain-testlets 'Getting around' and 'Self-care' (r = 0.102; critical-LDvalue = 0.1), which were subsequently combined to one common testlet. The fit statistics of the testlets of the WHODAS 2.0 (Üstün et al., 2010) can be found in table 5.

Table 5 Final analysis fit statistic of the WHODAS 2.0 12-items version (testlets ordered by location) – second study

Testlet	Item	Location	SE	Residual
Participating in society	4) How much of a problem did you have joining in community activities (for example, festivities, religious or other activities) in the same way as anyone else can? 5) How much have you been emotionally affected by your health problems?	-0.64	0.04	-0.72
Life activities	2) Taking care of your household responsibilities? 12) Your day to day work?	-0.25	0.04	-2.09
Getting around/ Self-care	1) Standing for long periods such as 30 minutes? 7) Walking a long distance such as a kilometer (or equivalent)? 8) Washing your whole body? 9) Getting dressed?	0.23	0.03	0.18
Understanding and communication	3) Learning a new task, for example, learning how to get to a new place? 6) Concentrating on doing something for ten minutes?	0.28	0.04	0.17
Getting along with people	10) Dealing with people you don't know? 11) Maintaining a friendship?	0.39	0.04	1.31

SE = Standard error

After applying these strategies, there was no further evidence of LD nor item misfit. The assumption of unidimensionality could be confirmed. The *t*-test showed satisfactory results with 11 significant tests (3.30%). The A-factor was 0.94, indicating a high explained common variance across the five testlets and confirming the scale's unidimensionality as well.

However, in the final analysis DIF related to age was found for the testlet 'Getting around/ Self-care' and related to gender for testlet 'Life activities'. Elderly persons seemed to have more difficulties in the domain 'Getting around/ Self-care' than younger persons with the same level of disability, and women seemed to have more difficulties in 'Life activities' than men with the same level. We investigated the impact of found DIF with the before mentioned methods. After splitting the testlet 'Life activities' for gender-DIF and computing equated scores, only a minor difference was found, with the biggest difference being 1.5 score-points. As the gender-DIF was considered as being not substantial, we decided not to split this testlet for gender in the final solution. The situation was similar regarding the age-DIF, although the difference in equated scores between the younger and older patients was slightly higher, with a maximum score difference of about 2 points in the middle range of the person location (between 0 and 1). However, in the other parts of the disability dimension, the difference was negligible. Additionally, we conducted an analysis to examine the impact of the found age-DIF in the present sample: Mean WHODAS 2.0-person parameters between the younger and older patients once with and once without adjusting for DIF were compared. The effect size (Cohen, 1988) for the comparison of younger and older patients without DIF adjustment was $d = 0.44$, whereas it was $d = 0.52$ with DIF-adjustment. Based on only minor differences in both lines of analyses, we decided not to split for age in the final solution.

After adjusting for LD, two testlets displayed disordered thresholds: the testlet 'Understanding and communicating' showed negligible disordering in the first two thresholds: Threshold 1 = -0.61; Threshold 2 = -0.63. The other thresholds in this testlet were ordered. The testlet 'Getting along with people' (item 10 and 11) showed more disordering. Several lines of additional analyses were performed, e.g., collapsing for the initial item 11, which had displayed disordering in the initial analysis as well, or rescoring items 10 and 11. However, disordering for the testlet 'Getting along with people' still remained, and model fit did not improve. For this reason, and the reason that the final solution with five testlets (without rescoring) met the expectations of the measurement model, we did not make any further optimization regarding threshold ordering.

The overall model fit of the final solution was satisfactory ($\chi2 = 88.21$, $p = 0.07$) with good reliability PSI = 0.82. Table 6 shows the summary fit statistic of the initial analysis, as well as of the analysis with the six domains and of the final analysis.

Table 6 Summary fit statistic – second study

Analysis	Overall item-trait interaction			Uni-dimensionality t-test (%)	Reliability	Targeting				Item misfit	Differential Item Functioning (DIF)	Disordered thresholds
						Item Residual		Person Residual				
	Chi-square	df	p-value	test (%)	PSI	Mean	SD	Mean	SD	Item number	Item number (source of DIF)	Item number
Initial	88.21	60	0.01	5.41	0.87	-0.37	1.63	-0.29	0.95	2, 9	Age: 1, 12 Gender: 7, 11 Disease duration: 12	1, 7, 8, 9, 11, 12
Testlets (6 WHODAS 2.0 Domains)	38.17	30	0.15	4.80	0.82	-0.51	1.40	-0.29	0.80	None	Age: Testlet "Getting around", "Life-activities" Gender: Testlet "Life activities"	Testlet "Getting around", "Understanding and Communicating", "Self-Care", "Getting along with people"
Final	36.14	25	0.07	3.30	0.82	-0.23	1.26	-0.31	0.96	None	Age#: Testlet "Getting around/ Self-Care", "Life-activities" Gender#: "Life-activities	Testlet "Understanding and communicating", "Getting along with people"

DF= degrees of freedom; PSI = person separation index; SD= standard deviation; **#** = It was not adjusted for the found DIF (see text for more details).

Figure 1 shows the targeting of the scale with a mean person location value of $M = -0.78$ ($SD = 1.03$). This result means that the patients had a lower mean level of disability than the average difficulty of the scale (which is 0). The person distribution was slightly off-centered, with more people showing lower levels of disability and only a relatively small number of persons with high levels of disability. The item threshold distribution shows that the scale measures a wide range of disability, except for very low levels and very high levels of disability.
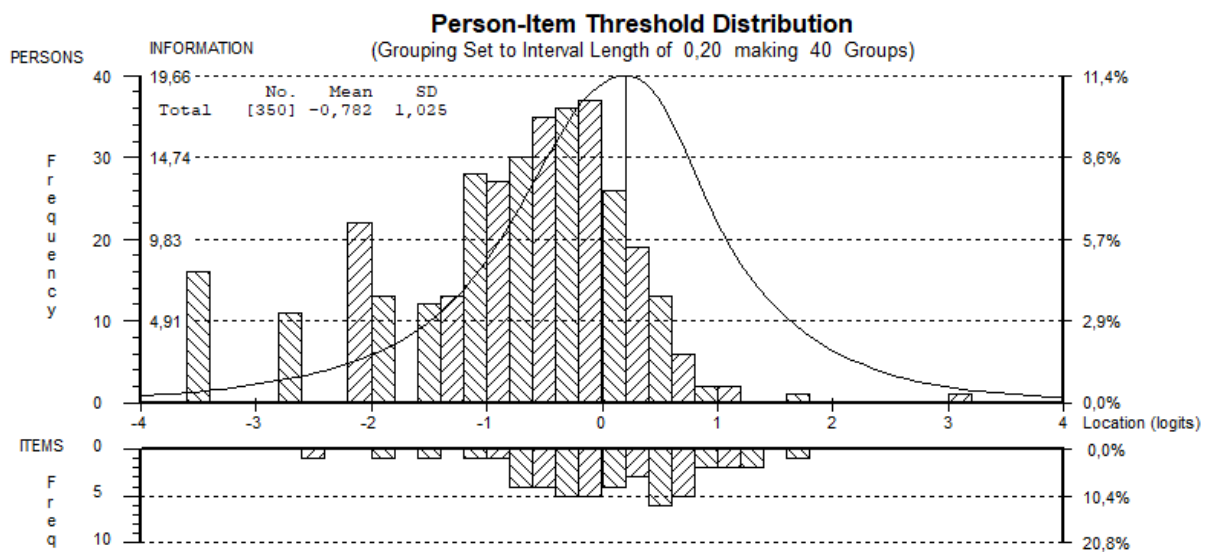
Figure 1 Person-Item threshold distribution (final analysis) – second study

Note. On the top half of the graph, the distributions of persons and at the bottom half the item thresholds are shown for the final analysis of the WHODAS 2.0 12-item version with higher values indicating higher level of disability (top of the half) and higher difficulty (bottom half). At the left side the frequency and at the right side the percentage of persons respectively items are shown.

A transformation table of the WHODAS 2.0-scores to interval-level person parameters is provided in table 7.

Table 7 Conversion table of Rasch logits – second study

WHODAS 2.0 Score	Interval-scaled person estimate
0	-3.49
1	-2.68
2	-2.15
3	-1.80
4	-1.55

5	-1.35
6	-1.19
7	-1.06
8	-0.94
9	-0.84
10	-0.75
11	-0.67
12	-0.59
13	-0.52
14	-0.45
15	-0.39
16	-0.33
17	-0.27
18	-0.22
19	-0.16
20	-0.11
21	-0.06
22	-0.01
23	0.04
24	0.09
25	0.14
26	0.19
27	0.24
28	0.29
29	0.34
30	0.39
31	0.44
32	0.49
33	0.54
34	0.60
35	0.65
36	0.71
37	0.78
38	0.85
39	0.93
40	1.01
41	1.12
42	1.23
43	1.37
44	1.53
45	1.74
46	2.01
47	2.43
48	3.08

Discussion

This study aimed at assessing and is the first to provide information about the psychometric properties of the 12-item version of the WHODAS 2.0 within a sample of cancer patients using modern psychometric analysis, i.e., Rasch analysis. The use of Rasch analysis has numerous potential advantages over CTT when assessing self-reported health outcomes. For example, it allows a nuanced analysis of the psychometric properties because of its focus on single items and how persons respond to them, it permits testing bias or DIF in different subgroups, and facilitates a transformation of ordinal into interval level scores. The use of the interesting and cancer-specific DIF variables should be highlighted. Overall, the Rasch measurement model's application on the WHODAS 2.0 showed a good model fit with good reliability after making some modifications related to LD.

The scale showed several pairs of locally dependent items corresponding to the domains of the WHODAS 2.0 (Üstün et al., 2010). After combining the locally dependent item pairs successively into domain-specific testlets, one last LD could be observed between the testlets 'Getting around' and 'Self-care', which had to be combined to one common testlet. In terms of content, this makes sense since both assess facets of activities of daily living (ADL). The findings of LD within the scale are comparable with other studies. For example, Luciano, Ayuso-Mateos, Fernández, et al. (2010) reported correlated pairs of items within the domains 'Getting around', 'Self-care' and 'Getting along with people' or Snell et al. (2020) within the domain 'Self-care'. We found LD in all domains of the WHODAS 2.0 like Kutlay et al. (2011) or Küçükdeveci et al. (2013) and additionally one between the two domains assessing ADL.

DIF was tested by gender, age, type of cancer, the presence of metastases, psycho-oncological support, and duration of disease. For most of these external variables, no DIF was found. However, in contrast to other studies (e.g. Kirchberger et al., 2014), uniform DIF occurred related to age for testlet 'Getting around/ Self-care' and related to gender for testlet 'Life activities'. After investigating the impact of the found DIF with splitting for gender and computing equated scores, we only found a relatively small inconsiderably difference in the equated scores, so we decided not to split for gender. However, there was a bigger difference with a maximum score difference of about 2 points in the middle range of the person's location regarding age. This result denotes

that patients with the same level of disability responded differently to the items of the ADL-testlet dependent on their age. Specifically, elderly individuals seem to have more difficulties in this domain than younger persons with the same level of disability. However, this difference becomes visible only in the middle range. In contrast, patients with either a high or low level of disabilities responded comparable in the areas of high or low level of disability, irrespective of their age. Another consideration about the found minor DIF might be that this is not a measurement bias, but the difference could be expected. People develop indeed more difficulties with higher age in areas of 'Getting around' and 'Self-care', so a split for age would not be necessary. Given that the DIF was found only in a tiny part of the assessed dimension and given the only minor differences (in term of effect sizes) when comparing younger and older patients with and without the DIF adjustment as well as the contentual reflection, about expected differences, we decided not to adjust for DIF. However, our sample is relatively young, with a mean age of 52.34 years. In a sample with more elderly patients, a more relevant age-DIF might be found.

The confirmation of unidimensionality of the scale is consistent with other Rasch analyses on the WHODAS 2.0 (Kirchberger et al., 2014; Luciano, Ayuso-Mateos, Aguado, et al., 2010). Additionally, targeting (Fig. 1) was satisfactory for the present sample with a mean person location value of $M = -0.78$ ($SD = 1.03$). However, for low and high levels of disability, the targeting is not as good as item thresholds are missing in these areas of the dimension. The WHODAS 2.0 was initially developed to provide a standardized method for measuring health and disability in the general population (Üstün et al., 2010). Our results indicate that even in a sample of patients with cancer, the differentiation in the lower segment of disability is not optimal – an area where probably most of the people of a healthy population would be located. However, the differentiation within a healthy population or persons with no respectively very low levels of disability may not be so relevant for assessment of oncology patients and the improvement of clinical decision making in psycho-oncology. However, more difficult items are also missing, making it hard to precisely assess disability in patients with a high level of disability using the 12-items version of the WHODAS 2.0. A good example is the *Getting around*-domain. In the 12-item version, the items "Standing for long periods such as 30 minutes?" and "Walking a long distance such as a kilometer (or equivalent)?" are indicators for this domain – activities that might be far too difficult to perform for severely ill patients. Here it might be sensible to either include some more

items of the WHODAS 2.0 36 items version or develop a better targeted short scale for patients with a higher level of disability (e.g., with WHODAS items like: "Moving around inside your home.").

In the initial analysis of our study, disordered thresholds were found for six items. In the testlet solution, less disordering was found, indicating that at least part of the threshold disordering in the initial analysis was due to LD. However, the testlet "Getting along with people" displayed disordered thresholds, a phenomenon often observed for testlets. Therefore, the ordering of thresholds should be further investigated in future WHODAS-studies.

Besides some strengths, the present study also has limitations. There is a relatively high percentage of breast cancer patients in the sample of this study. Accordingly, the results may only be generalized to cancer patients with caution. Due to small group sizes, we had to combine the residual cancer types into one category, 'other forms of cancer', for DIF analysis. To examine the influence of various cancer forms decidedly, especially cancer types with more severe disease progress, additional research would be interesting and important. Nevertheless, in our study, we could use the presence of metastases or the duration of disease as an indicator for the severity. Both of these indicators showed no DIF. Also, the sample's psychological distress, measured by the HADS-T, is roughly equally distributed across the cancer forms. We therefore can assume that the type of cancer does not unduly influence the response behavior. Furthermore, the sample was recruited from social media platforms and within online cancer support groups. As a result of this, the sample is relatively young, with a mean age of 52.34 years. The scale's targeting was good for the present sample but already shows an off-centered person distribution with a relatively small frequency of persons with a high level of disability. This result indicates a bias by low disability levels in this sample. Also, a high percentage (41.7%) of the cancer patients have an active job situation, indicating a relative fit sample. The item threshold distribution shows that the scale measures a wide range of disability but not across the entire range. With respect to this and the small age-DIF we found in our study, future research should examine a sample with a higher level of disability and perhaps include some additional items suited for the assessment of higher levels of disability.

Conclusion

The present study provides essential information about the psychometric properties of the 12-items version of the WHODAS 2.0 in the oncological context. The Rasch analysis of the 12-items version of the WHODAS 2.0 showed that this measurement may be used well in the oncological context, especially those who have an impairment are adequately assessed with it. The instrument is non-biased with respect to gender, type of cancer, the presence of metastases, psycho-oncological support, and duration of disease. There might be only a need for critical consideration with respect to age, especially in the elderly.

2.6   Positive Mental Health

2.6.1 Definition and historical aspects

Another current requirement is the consideration of positive aspects in mental health assessment. Even, and perhaps even more so in the era of the global Corona pandemic, psychological aspects of illness are gaining importance, and there is a need to reflect the challenges on mental health (Moreno et al., 2020).

However, for a long time, the deficit-oriented approach was followed in psychology, so that the focus was only on psychopathology, symptoms, illness and its treatment (Seligman & Csikszentmihalyi, 2000). In recent years, this focus began to change. This has taken into account the research findings of positive psychology and the realization that mental health is not just the absence of illness, but a state of well-being that has a positive impact on a whole range of life factors (Keyes, 2005; WHO, 2004). In a narrower sense, the term mental health has been established earlier. The history of the development of the concept of mental health can be read elsewhere (e.g., Bertolote, 2008; Keyes, 2007).

The term positive psychology was first used by Abraham Maslow in the mid-20th century and picked up about 40 years later by Martin Seligman, who called for more attention to be paid to the good in people and in the world (Snyder et al., 2021). Positive psychology is concerned with positive subjective experiences such as well-being, satisfaction, hope, and optimism, as well as positive individual characteristics such as interpersonal skills, perseverance, and future orientation, and with positive institutions and communities (Seligman & Csikszentmihalyi, 2000). With the advent of positive psychology, a new distinction was made from earlier historical and methodological ideas and research focused on psychopathology to the interconnecting facets of mental health and its positive impact on human life (Snyder et al., 2021). Accordingly, facets of well-being, respectively positive mental health, and mental health problems, may also be present at the same time (Lukat et al., 2016). Attempts to conceptualize mental health assume that there are several facets, which can be divided into eudaemonic well-being, i.e., positive psychological and social functioning in life, and hedonic well-being, i.e., positive emotions toward one's life (Keyes, 2007).

The increased interest in positive mental health motivated the development of several assessment instruments (Miret et al., 2015). To this end, a scale combining the hedonic and eudaemonic aspects of mental health was developed, aiming to capture positive mental health with a short, person-centered, and unidimensional questionnaire, namely the PMH scale by (Lukat et al., 2016). The evaluation of the psychometric properties of this scale according to the aspects already listed in the previous chapters of the Rasch model will be conducted in the third study. Prior to this, a brief outline of the importance of positive mental health in clinical research and practice, particularly in the psycho-oncology context, will be provided.

2.6.2 Relevance for psycho-oncological context

The increased call for studying positive mental health is not unique to research, but also has clinical implications. For example, Keyes et al. (2010) highlight that increases in mental health predict decreases in mental illness and losses in mental health predict increases in mental illness. In addition, research shows that complementary interventions focused on positive mental health can be helpful for people with severe mental illnesses, such as psychosis (Schrank et al., 2014) and in suicide prevention (Stecz et al., 2020). There are several psychological interventions and approaches that focus on strengthening or promoting positive mental health, such as acceptance and commitment therapy (ACT) (Fledderus et al., 2010) or well-being therapy (Fava et al., 1998).

However, research is increasingly examining the relationship between positive mental health and physical health (Pressman & Cohen, 2005). The influence of mental health improves recovery and survival rates in physically ill patients (Lamers et al., 2012) and reduce distress in clinical samples with psychiatric or somatic disorders (Chakhssi et al., 2018), including cancer patients. Sin (2016) hypothesizes three potential pathways in which positive well-being influences cardiovascular disease, first by improving adaptive physiological functions, second by reinforcing better health behaviors, and third by buffering the deleterious effects of stress on health. It is conceivable that this relationship may also be observed in other physical diseases, including cancer.

In psycho-oncology, however, the study of patient-reported outcomes and interventions has focused primarily on psychological distress and quality of life, as has

psychology research in general (Faller et al., 2013). However, interventions for cancer patients like meaning-based interventions are based on approaches of positive psychology (Mehnert et al., 2011). For example, results from mindfulness-based approaches show reductions in anxiety and depression in cancer patients and survivors (Piet et al., 2012) and several positive mental health interventions enhanced quality of life, well-being, hope, benefit finding, or optimism in breast cancer patients (Casellas-Grau et al., 2014). Intervention studies for cancer survivors increasingly include positive mental health outcomes (Holtmaat et al., 2019). In psycho-oncology research, there are several efforts to identify moderator and mediator variables that influence the development of psychological distress and to identify the benefits of interventions. The positive aspects of mental health are also studied from this point of view. However, contradictory results emerge here. While some authors cannot find clear evidence for moderating effects of optimism/hope (Gustavsson-Lilius et al., 2007), authors in a systematic review suggest a similar pathway as in cardiovascular disease, in the sense that individuals with greater optimism and greater hopefulness try to engage in healthier behaviors regardless of their clinical status, and that this contributes to the management of chronic disease (Schiavon et al., 2017). Nonetheless, the exploration of positive facets of mental health is increasing in research and, consequently, so are the clinical implications in terms of treatment options. However, more research is needed in this regard to provide more understanding and information. The following third study in the present thesis follows the trend of investigating positive mental health using cancer patients and examines the psychometric property of a short, unidimensional assessment instrument on PMH using the Rasch model.

2.7   Study 3 - Psychometric Evaluation of the Positive Mental Health (PMH) Scale using Item Analysis according to the Rasch Model

Psychometric Evaluation of the Positive Mental Health (PMH) Scale using Item Analysis according to the Rasch Model

Vaganian, L., Boecker, M., Bussmann, S., Kusch, M., Labouvie, H., Margraf, J., Gerlach, A. L., & Cwik, J. C. (in revision). Psychometric Evaluation of the Positive Mental Health (PMH) Scale using Item Analysis according to the Rasch Model. *BMC Psychology.*

The following article is the manuscript version.

Abstract

Background: The investigation of patient-reported outcomes and psycho-oncological interventions mainly focuses on psychological distress or psychopathology. However, the recognition of the equal importance of positive mental health (PMH) has increased lately. The PMH-scale is a brief questionnaire allowing to assess well-being in individuals in the general population and in patients. Previous studies evaluated the psychometric properties of the PMH-scale using classical test theory (CTT). This study is the first to investigate the PMH-scale in patients with cancer using item analysis according to the Rasch model.

Methods: In total, $N = 357$ cancer patients participated in the study. A Rasch analysis of the PMH-scale was conducted including testing of unidimensionality, local independence, homogeneity and differential item functioning (DIF) with regard to age, gender, type of cancer, the presence of metastases, psycho-oncological support, and duration of disease. Additionally, the ordering of the item thresholds as well as the targeting of the scale were investigated.

Results: After excluding one misfitting item and accounting for local dependence by forming superitems, a satisfactory overall fit to the Rasch model was established ($\chi^2 = 30.34$, $p = 0.21$). The new PMH-8 scale proved to be unidimensional, and homogeneity of the scale could be inferred. All items showed ordered thresholds, there was no further item misfit. DIF was found for age, but as the impact of DIF was not substantial, no adjustment related to the age-DIF had to be made. The Person Separation Index (PSI = 0.89) was excellent, indicating excellent discriminatory power between different levels of positive mental health. Overall, the targeting of the PMH-8 was good for the majority of the present sample. However, at both ends of the scale item thresholds are missing as indicated by a slight floor effect (1.4%) and a considerable ceiling effect (9.8%).

Conclusion: Overall, the results of the analysis according to the Rasch model support the use of the revised PMH-scale in a psycho-oncological context.

Keywords: mental health, Cancer, Rasch analysis, Psychometric properties, well-being

Background

Mental health research has predominantly concentrated on psychopathology and symptoms (Seligman & Csikszentmihalyi, 2000). In recent years, the focus from this deficit-centered approach started to change, taking into account the findings of positive psychology research and the recognition that mental health is not merely the absence of disease but rather a state of well-being that positively affects the whole range of life factors (e.g., coping with daily stressors and functioning in work and community) (Keyes, 2005; World Health Organization [WHO], 2004). Accordingly, facets of well-being, respectively positive mental health (PMH), and mental health problems, may be present simultaneously (Lukat et al., 2016). Attempts to conceptualize mental health assume that there are several PMH facets, which can be divided into eudaemonic well-being, i.e., positive psychological and social functioning in life, and hedonic well-being, i.e., positive emotions toward one's life (Keyes, 2007).

In psycho-oncology, the investigation of patient-reported outcomes and interventions likewise has mainly focused on psychological distress and quality of life (Faller et al., 2013). Indeed, cancer regularly is associated with physical and mental distress. This distress depletes patients' quality of life and negatively influences disease progression and survival rates (Chan et al., 2015; Karakas & Okanli, 2014; Linden et al., 2012). However, research on well-being's influence on mental health also shows effects and improves recovery and survival rates in physically ill patients (Lamers et al., 2012). Several psychological interventions like Acceptance and Commitment Therapy (ACT) (Fledderus et al., 2010) or well-being therapy (Fava et al., 1998) aim at enhancing well-being. Similarly, interventions for cancer patients like meaning-based interventions is rooted in positive psychology (Mehnert et al., 2011). Importantly, positive mental health can help to protect cancer survivors against distress and demoralization (Vehling et al., 2011).

This increased interest in positive mental health motivated the development of several assessment instruments (Miret et al., 2015). Valid and reliable instruments are needed in order to be able to evaluate clinical interventions, to ensure sound clinical decision-making, and to select the most appropriate interventions for individual patients. To this end, a scale has been developed that combines the hedonic and the eudaemonic aspect of mental health (Keyes, 2007) and aims to assess positive mental health with a brief, person-centered and unidimensional questionnaire (Lukat et al., 2016).

Unidimensionality means that a scale exclusively assesses one underlying construct. This is crucial because it ensures that the interpretation of the instruments' scores is representative of the measured construct (Ziegler & Hagemann, 2015).

The PMH-scale is a self-rating questionnaire constructed to assess the positive dimension of the dual-factor model of mental health, i.e., integrating positive and negative mental health factors (Suldo & Shaffer, 2008). The scale is available in 12 languages and validated in a study sample, the general population, and a patient sample (Lukat et al., 2016). Usage is continuously increasing, for example, in research for predicting adaptive and maladaptive responses to the Coronavirus (COVID-19) (Brailovskaia & Margraf, 2020), in studies looking at cross-cultural differences (Bieda et al., 2017), and suicide ideation (Brailovskaia et al., 2020).

Several psychometric studies based on classical test theory (CTT) have been conducted using the PMH-scale. They generally demonstrated high internal consistency, good retest- reliability, good discriminant and convergent validity, and supported unidimensionality within samples of students, patients, and the general population (e.g., Bibi et al., 2020; Bieda et al., 2017; Lukat et al., 2017; Lukat et al., 2016). However, in CCT based analyses, scores are calculated by summing up the responses on items and these test scores are assumed to be on interval scale level which is normally not the case (Streiner, 2010). An alternative to CCT is item response theory (IRT), which is a group of measurement models that explain the relationship between the responses to items and the person location of an underlying latent trait (Hays et al., 2000). One of these modern approaches is the item analysis according to the Rasch model (Cappelleri et al., 2014). Since the measurement model is characterized by its simplicity, it occupies a special position among IRT models (Christensen et al., 2019). In case that person responses to scale items fit the Rasch model the ordinal score can be converted into an interval-level person parameter. There are numerous potential advantages of IRT models, including Rasch analysis, over CTT in assessing self-reported health outcomes. For example, it allows testing for unidimensionality, bias across different subgroups, and the systematical investigation of local dependency (LD) which might inflate the reliability of a scale. Additionally, it enables the examination of targeting and how the response options of items are used by the assessed persons. Focusing on individual items and how persons respond to those items allows for a more sophisticated analysis of the

psychometric properties of the questionnaire under study (Fischer, 1987; Gustafsson, 1980; Hays et al., 2000; Streiner, 2010). However, to the best of the authors' knowledge, Rasch analysis has not yet been applied to the PMH-scale.

Since psycho-oncological interventions and cancer patients may benefit from positive effects of PMH improvement with respect to recovery and survival rates and as a protective factor, it is important to consider the PMH-scale application in the oncological context as well. However, research studying the psychometric properties of the PMH-scale in an oncological context does not yet exist. Against this background, we examined the psychometric properties of the PMH-scale in oncological context among various types of cancer patients using Rasch-analysis, especially to investigate the assumptions of unidimensionality, invariance across different exogenous variables, local independence of items. Additionally, a special focus was placed upon the investigation of targeting. A scale is well targeted to a sample if the majority of the sample is assessed with good measurement precision (Christensen et al., 2013).

Method

Participants and procedure

Using SoSciSurvey (Leiner, 2014), participants were invited to participate in the study as an online survey consisting of various questionnaires. Participants were asked about their cancer diagnosis and selected applicable types of cancer from a list. This question was designed as a multiple-choice task with several answer options as well as an open, descriptive category "other", so that several cancer diagnoses could be named at the same time. Social media platforms, a forum for cancer patients, and mailing lists from self-help groups were used to advertise the study as part of another validation study (Cwik et al., 2021). All participants gave their informed consent online, after being informed about study content and aims, procedures, and planned publications. Inclusion criteria were: age ≥ 18 years and at least one current or in the past cancer diagnosis. No exclusion criteria were defined. In total, $N = 357$ cancer patients ($n = 288$ women (80.7%), $n = 68$ men (19.0%), $n = 1$ gender diverse (0.3%)) completed the PMH-scale.

The study was approved by the Ethics Commission of the University's Faculty of Medicine (reference number 18-098). All procedures contributing to this work comply

with the relevant national and institutional committees' ethical standards on human experimentation and the Helsinki Declaration of 1975, as revised in 2008.

Assessment instrument

PMH. The German version of the PMH-scale (Lukat et al., 2016) was used, a self-report instrument consisting of nine items rated on a four-point Likert type scale ranging from 0 ("do not agree") to 3 ("agree"). It assesses the emotional, psychological, and social aspects of positive mental health. Higher scores reflect greater positive mental health. In a series of six studies that included samples of students, patients, and the general population, the scale showed good psychometric properties e.g., high internal consistency (Cronbach's alpha = . 93), satisfactory retest reliability (r = .74 - .81), and convergent validity was confirmed, e.g., with Satisfaction With Life Scale (Diener et al., 1985) (r = .75), Subjective Happiness Scale (Lyubomirsky & Lepper, 1999) (r = .81) (Lukat et al., 2016), and demonstrated strong cross-cultural measurement invariance in student samples from Germany, Russia, and China (Bieda et al., 2017).

Statistical analyses

Data were analyzed using SPSS version 26.0 (IBM Corporation, Released 2019) and RUMM2030 software (Andrich et al., 2009).

To assess the psychometric properties of the PMH-scale in an oncological context, item analysis according to the Rasch model was used. IRT models, including the Rasch model, can be used to analyze the psychometric properties of an instrument in detail because they focus on individual items and how people respond to those items. The probability of an item response is a function of the difference between person parameters and item difficulty parameters on the latent trait, which in this case is PMH (Lukat et al., 2016). Performing a Rasch analysis involves examining how well the data meet the expectations of the measurement model and whether certain requirements are met. This is a primacy of Rasch models, that the data must fit the model, not the other way around (Petrillo et al., 2015). As with other IRT models, the requirements relate to unidimensionality, local independence, and absence of differential item functioning (DIF). Specific to Rasch analyses is the requirement of homogeneity. The analysis of the Rasch model can be understood as an iterative process in which the model assumptions are checked and potential deviations found are resolved, if possible. In case model fit is found, the transformation of ordinal scores into interval-

level parameters is possible. The Rasch model uses a logistic transformation to convert ordinal scores into linear measures expressed in "logits" (i.e., log-odds units) (Christensen et al., 2013).

Due to the polytomous PMH-items, the Partial Credit Model (PCM) (Masters, 1982) was used. Overall model fit was evaluated using the chi-square item-trait interaction statistic. A good level of overall fit is characterized by a non-significant chi-square probability $p > 0.01$ (Christensen et al., 2013; Pallant & Tennant, 2007; Siegert et al., 2010). To conclude a good fit, the mean values of the residuals should be around 0 and have a standard deviation of 1. Besides the overall fit, the fit of the individual items (item fit) and persons (person fit) can be evaluated and are expressed as residuals. The fit residuals are expected to be within a range of ± 2.5 (Andrich et al., 2004; Christensen et al., 2013). The second fit-statistic is a chi-square statistic and the chi-square probability should be non-significant.

One fundamental requirement of the Rasch model is unidimensionality, i.e., the items of a scale should capture only one underlying construct, which was tested with principal component analysis (PCA) of the residuals (Christensen et al., 2013; Pallant & Tennant, 2007). The idea is to use the items with the highest negative/positive loadings on the first component to create two subsets of items. The separate person estimates of these two subsets are used to identify significant differences with independent t-tests. The proportion of significant t-tests should not exceed 5% to reject multidimensionality and infer unidimensionality (Smith, 2002).

Another assumption is that of local independence. This assumption implies that there should be no residual correlations between items when extracting the trait factor (Tennant & Conaghan, 2007). LD can occur when items are linked such that the response to one item determines the response to another item (Pallant & Tennant, 2007; Tennant & Conaghan, 2007). Because LD can lead to overestimation of reliability, bias in parameter estimation, and corrupt construct validity (Christensen et al., 2017) adequate handling of it is critical. Local independence was investigated using a residual correlation matrix of the items. Items with a residual correlation of 0.2 above the average were considered as locally dependent (Christensen et al., 2017; Marais, 2013). One strategy to deal with LD if one does not want to delete scale items is to combine the locally dependent items into 'superitems' by adding them together. Using the 'superitem'-strategy results in a bi-factor equivalent solution. The proportion of

explained common variance (ECV) (Andrich, 2016; Pomeroy et al., 2020; Rodriguez et al., 2016) of the general factor, should be >0.9 to consider the scale as unidimensional (Pomeroy et al., 2020).

A specific assumption of the Rasch model is that the items are assumed to be homogeneous in the sense that the ranking of the item parameters should be the same for all respondents, regardless of their expression of the latent trait. This requirement is reflected in tests of item-trait interaction based on group residuals, i.e., differences between observed and expected scores in groups matched by their total person-parameters scores (Andrich et al., 2004; Tennant & Conaghan, 2007; Vindbjerg et al., 2021).

Another assumption is the absence of DIF. If DIF is given, the difficulty of an item is different for different groups (e.g., men and women). In other words, the corresponding item indicates the latent trait in different ways in different groups (Christensen et al., 2013; Tennant & Conaghan, 2007). DIF analyses were examined using analysis of variance (ANOVA). Uniform DIF is shown by a significant main effect for person factor indicating that the different groups show a consistent difference in their responses to an item across the whole range of the assessed dimension. The presence of non-uniform DIF is shown by a significant interaction effect (person factor x class interval) indicating that the differences between groups vary across the levels of the assessed dimension. In this study, we tested the items for DIF in relation to gender (woman, man), age (median split of the sample: below and above 54), type of cancer (breast, other forms of cancer, multiple cancers), presence of metastases (yes, no, unknown), psycho-oncological support (yes, no) and duration of disease (median split of the sample: below and above 3.9 years). To avoid too small subgroups in the ANOVA, we had to exclude the one gender diverse person, and the metastasis category 'unknown' from the DIF analysis and combine the remaining cancer diagnoses with lower frequencies into one category 'other forms of cancer' for the cancer type DIF analysis. In the case of DIF, several strategies to deal with can be used. One possibility is to remove or reformulate items with DIFs or to split the item with regard to the respective DIF-variable. We used the latter strategy and split the item in case DIF was found and subsequently evaluated the impact of DIF by computing equated scores (Christensen et al., 2019). Following this method, the item for which DIF was found, is split for the respective DIF-variable

(e.g., for gender). For each DIF-subgroup (e.g., males vs. females) a score-to-measure transformation is performed and for each person parameter the equated scores of e.g., males and females can be compared and the size of score differences can be evaluated. (Cameron et al., 2014; Hagquist & Andrich, 2017).

Moreover, to assess the category functioning of each item, the threshold ordering was analyzed using the category probability curves. Item thresholds are the transition points between two adjacent respond categories. Disordered thresholds can affect the interpretation and validity of scale scores (Andrich, 2013). There may be several causes of threshold disorder, such as respondents having difficulty to consistently differentiate among response options or LD causing the disorder. If the disorder is due to problems with category differentiation, one option is to collapse the disordered response categories together.

The reliability of the scale was estimated using the Person Separation Index (PSI). The PSI indicates the discriminatory power of how well a set of items can distinguish between the individuals being measured. PSI values of .7 are considered appropriate for group and .85 appropriate for individual applications (Andrich et al., 2004; Christensen et al., 2013; Pallant & Tennant, 2007; Tennant & Conaghan, 2007).

Targeting describes the extent to which a scale is appropriate for a given sample in terms of scale difficulty. Targeting was assessed graphically using the person-item threshold distribution graph. Person-item maps show how person parameters and item thresholds are distributed along the measured dimension (Christensen et al., 2013). They indicate whether the item thresholds are located in the same range as the person parameters. If a scale is poorly targeted for a sample, the measurement precision is low in those ranges of the assessed dimension in which the persons are located. In case of the PMH-scale the scale would be poorly targeted if respondents either report less well-being than the scale assesses or have a higher level of well-being. Additionally, the extent of floor and ceiling effects and a mean person parameter deviating substantially from zero (which usually is the mean value of the item difficulty) can be indicators of poor targeting. (Christensen et al., 2013; Tennant & Conaghan, 2007).

Results

Sample

The mean age of the N = 357 participants was 52.40 years (SD = 14.01). All participants completed the PMH questionnaire. A selection of descriptive statistics and an overview of cancer diagnoses among the participants are presented in table 8. The cancer diagnosis question was a multiple choice question, and some respondents (n = 46, 12.9%) reported more than one cancer diagnosis.

Table 8 Characteristics of cancer patients (N = 357) - third study

Gender	
Male	68 (19.0)
Female	288 (80.7)
Divers	1 (0.3)
Age (years)	52.40±14.01 (20-83)
Job situation	
Active	147 (41.2)
Certified sick	58 (16.2)
Different form	152 (42.6)
Types of cancer*	
Breast	162 (45.4)
Urological	31 (8.7)
Prostate, testicular	25 (7.0)
Gynecological	19 (5.3)
Hematological	22 (6.2)
Intestinal, rectal	12 (3.4)
Skin	3 (0.8)
Lungs, Bronchia	5 (1.4)
Ear, Nose, Throat	7 (2.0)
Gastric, esophageal, pancreatic	4 (1.1)
Parts of central nervous system	2 (0.6)
Soft tissue	1 (0.3)
Residual category (including different forms of cancer)	18 (5.0)
Multiple cancer forms	46 (12.9)
Metastases	
No	267 (74.8)
Yes	78 (21.8)
Unknown	12 (3.4)

Current psycho-oncological, psychological, psychotherapeutical support	
No	256 (71.7)
Yes	101 (28.3)
HADS-T Score - Distress (HADS T ≥15)	158 (44.3)

Values are presented in frequency (%) or mean±standard deviation (range). HADS-T = Hospital Anxiety Depression Scale (Herrmann-Lingen et al., 2011) (To identify patients with an increased need for psycho-oncological care and especially for depression symptoms in cancer patients, a sum score of HADS-T ≥ 15 can be used as the cut-off value)(Jenniches et al., 2020); *Patients reporting more than one type of cancer diagnosis are reported in the multiple cancer forms category.

Analysis according to the Rasch model.

In the initial analysis, we reviewed the model assumptions and determined how well the data met the expectations of the measurement model. The results of this initial analysis of all nine PMH items showed an unsatisfactory overall model fit ($\chi^2 = 72.75$, $p = 0.005$). The items statistics displayed misfit with a residual mean of -0.32 ($SD = 2.57$). For persons the residual mean was -0.37 ($SD = 1.23$), indicating no serious misfit. Several pairs of items displayed LD, and DIF was found for item 5 ('I manage well to fulfill my needs.') in relation to age. Three items showed item-misfit when using the fit residual as criterion, but using the chi-square statistic and Bonferroni correction no significant item misfit was found. However, the p-value of item 9 ('I am a calm, balanced human being.'; $p = 0.007$) was only slightly above the Bonferroni corrected significance level ($p = 0.005$) and had a very high fit residual (4.95), reflecting potential multidimensionality. We decided to exclude item 9 from further analyses. The other two misfitting items based on item fit residuals were items 3 (item residual = -2.79) and 8 (item residual = -2.54), which had too high negative item residuals indicating possible LD. All items showed ordered thresholds. Person misfit was negligible with only four patients (1.12%) showing fit residuals higher than 2.5. The test statistics of the nine PMH items of the first observation analysis are shown in table 9, which shows the item location (difficulty), the corresponding standard errors (SE), the item residuals indicating the item fit, and chi-square statistics.

Table 9 Initial analysis test statistic of the nine items of PMH-Scale (items ordered by location) - third study

PHM-Scale items	Loc	SE	Res	χ2 (df)	*p*
8) Much of what I do brings me joy.	-0.96	0.10	-2.54	10.80 (5)	0.055
2) I enjoy my life.	-0.65	0.10	-1.31	2.65 (5)	0.754
4) In general, I am confident.	-0.57	0.10	-1.47	9.10 (5)	0.105
3) All in all, I am satisfied with my life.	-0.10	0.09	-2.79	7.88 (5)	0.163
5) I manage well to fulfill my needs.	0.06	0.09	-1.21	9.36 (5)	0.096
7) I feel that I am actually well equipped to deal with life and its difficulties.	0.12	0.09	-1.93	6.53 (5)	0.258
1) I am often carefree and in good spirits.	0.44	0.09	1.40	3.28 (5)	0.656
9) I am a calm, balanced human being.	0.67	0.09	4.95	15.81 (5)	0.007
6) I am in good physical and emotional condition.	1.02	0.09	1.99	7.36 (5)	0.195

Loc (Location) = difficulty; SE = Standard error; Res (Residual) = item fit; χ2 = Chi-square, df = degrees of freedom; Bonferroni adjusted significance level: 0.005556.

As LD seemed to be the major problem, we focused on accounting for it after excluding item 9 because of misfit. Starting with the highest residual correlation and adjusting successively for the following higher correlation and always checking model fit, item pairs 1&2, 3&4, and 6&7 were combined into 'superitems'. After applying these strategies, there was no further evidence of LD nor of item or person misfit. The assumption of unidimensionality could be derived (significant *t*-tests: 4.10%). The ECV was 0.98, indicating a high explained common variance and also suggesting the scale's unidimensionality as well. All items still showed ordered thresholds. The test statistics of the eight PMH items in the final analysis are shown in table 10, which again shows the item location (difficulty), the corresponding standard errors (SE), the item residuals indicating the item fit, and chi-square statistics.

Table 10 Final analysis test statistic of the eight items of PMH-Scale (items/ 'superitems' ordered by location) - third study

Items/ Super items	PHM-Scale items	Loc	SE	Res	χ2 (df)	*p*
i5	5) I manage well to fulfill my needs.	-0.79	0.10	-0.96	12.22 (5)	0.032
s3_4	3) All in all, I am satisfied with my life. 4) In general, I am confident.	-0.23	0.07	-0.69	2.23 (5)	0.817
s1_2	1) I am often carefree and in good spirits. 2) I enjoy my life.	0.07	0.07	1.10	1.41 (5)	0.922
i8	8) Much of what I do brings me joy.	0.23	0.10	0.16	11.07 (5)	0.050
s6_7	6) I am in good physical and emotional condition. 7) I feel that I am actually well equipped to deal with life and its difficulties.	0.73	0.06	1.53	3.41 (5)	0.637

Loc (Location) = difficulty; SE = Standard error; Res (Residual) = item fit; χ2 = Chi-square, df = degrees of freedom, Bonferroni adjusted significance level: 0.01000.

However, in the final analysis, uniform DIF related to age was found for 'superitem' 1&2 ($p = 0.001$) and item 5 ($p < 0.001$) (see table 11). The DIF found initially suggests that elderly individuals seem to find it easier to meet their needs than younger individuals with the same level of well-being (item 5), and younger individuals seem to find it easier to enjoy life than older individuals with the same level of well-being ('superitem' 1&2).

Table 11 DIF summary (Age) of the eight items of PMH-Scale - third study

Items/ super items	Uniform DIF for Age				Non-Uniform DIF for Age			
	M	F	DF	Prob	M	F	DF	Prob
s1_2	9.46	10.74	1	*0.001*	0.14	0.16	5	0.978
s3_4	0.09	0,12	1	0.730	2.19	3.06	5	0.010
s6_7	1.44	1.58	1	0.210	1.66	1.82	5	0.108
i5	9.80	12.87	1	*0.000*	0.18	0.23	5	0.948

| i8 | 0.30 | 0.43 | 1 | 0.512 | 0.56 | 0.80 | 5 | 0.547 |

Bonferroni adjusted significance level: 0.003333. Items showing significant p-values marked in italics

We investigated the impact of the found DIF with the before mentioned methods. After splitting item 5 for age-DIF, there was no more evidence of age-related DIF for 'superitem' 1&2 ($p = 0.840$) indicating that the latter was probably artificial DIF (Andrich & Hagquist, 2015). To evaluate the magnitude of the found age-related DIF in item 5, equated scores were computed. The difference in the equated scores between the younger and older patients was only minor, with a maximum score difference of about 0.5 points in the lower range of the PMH dimension (between -4 and -3). However, in the other parts of the dimension, the difference was even more negligible. The equated scores are presented in table 12. Thus, as the age-DIF was considered as being not substantial, we decided not to split this item for age in the final solution.

Table 12 Equated scores showing the minor impact of age-DIF - third study

Age < 54	Age ≥ 54	Age < 54	Age ≥ 54
1	1.50	13	13.37
2	2.47	14	14.36
3	3.43	15	15.33
4	4.38	16	16.29
5	5.33	17	17.22
6	6.29	18	18.14
7	7.26	19	19.06
8	8.25	20	19.97
9	9.26	21	20.88
10	10.29	22	21.79
11	11.32	23	22.69
12	12.35	24	23.41

The final solution's overall model fit with eight items was satisfactory ($\chi^2 = 30.34$, $p = 0.21$) with excellent reliability PSI = 0.89. After these adjustments, no patient showed fit residual scores higher than 2.5. The summary test statistics of the initial and final analyses are presented in table 13 with the number of items, overall model fit, unidimensionality test, reliability, item and person fit (residuals), and item misfit.

Table 13 Overall summary of test statistic - third study

Analysis	# items	Item-Trait interaction (Overall fit)			Uni-dimensionality t-test	Reliability	Item-Residual		Person-Residual		Item misfit
		χ^2	df	p-value	test (%)	PSI	M	SD	M	SD	Item number
Initial	9	72.75	45	0.01	3.31	0.90	-0.32	2.57	-0.37	1.23	3, 8, 9
Final	8	30.34	25	0.21	4.10	0.89	0.23	1.09	-0.32	1.01	None

df = degrees of freedom; # items = number of items; Item-/ Person Residual = differences between observed and expected responses; M= Mean; PSI = person separation index; SD = standard deviation; $\chi2$ = Chi-square

Figure 2 shows the targeting of the scale. Overall, the item threshold distribution shows that the scale measures a wide range of positive mental health, except for very low levels and very high well-being levels. The majority of the patients of the present sample were located within the same range as the item threshold parameters. The mean person location value was $M = 1.19$ ($SD = 2.15$). This value means that the patients had a slightly higher level of well-being than the scale's center (which is 0). Thus, the person distribution demonstrates slight mistargeting, with more people showing higher levels of well-being and 9.8% of people having the highest possible score (ceiling effect). There were also a few persons with the lowest possible score (1.4%) (floor effect).
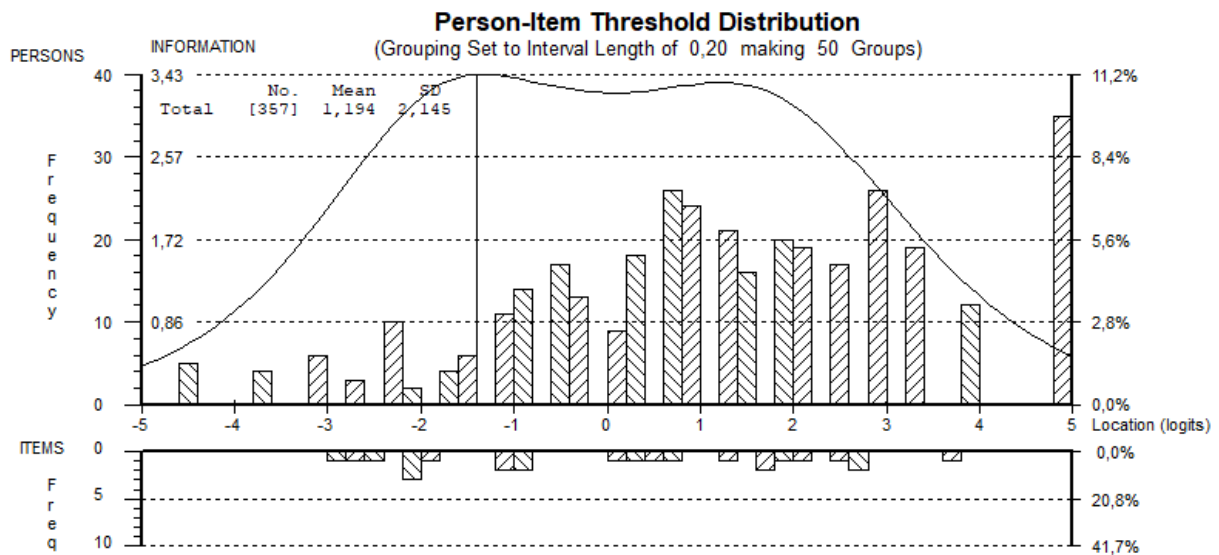

Figure 2 Person-Item threshold distribution (final analysis) - third study

Note. Person-item threshold distribution of the PMH responses. Higher values indicate a higher level of well-being (top of the half) and higher item difficulty (bottom half). At the left side the frequency and at the right side the percentage of persons respectively items are displayed.

A conversion table of the PMH-8 sum scores to interval-level person parameters and the corresponding interval-scale latent estimates transformed to a 0-100 interval scale is provided in table 14.

Table 14 Conversion table of the PHM-8 scale - third study

PMH ordinal scale score	Interval-scaled person estimate	Transformed interval scale 0–100
0	-4.54	0
1	-3.70	9
2	-3.10	15

3	-2.67	20
4	-2.32	24
5	-2.02	27
6	-1.73	30
7	-1.45	33
8	-1.17	36
9	-0.88	39
10	-0.58	42
11	-0.27	46
12	0.04	49
13	0.36	52
14	0.67	56
15	0.97	59
16	1.26	62
17	1.55	65
18	1.84	68
19	2.15	71
20	2.48	75
21	2.85	79
22	3.31	84
23	3.95	91
24	4.83	100

Discussion

This study is the first to provide information on the psychometric properties of the PMH scale within a sample of cancer patients and the first to use a modern psychometric analysis, i.e., Rasch analysis with its many potential advantages over CTT in assessing self-reported health outcomes. The use of relevant and cancer-specific DIF variables in this study should be highlighted. Adequate interval level measurement is of great importance when evaluating clinical interventions, ensuring sound clinical decision-making, and monitoring changes across the course of treatment.

Especially interventions for improving PMH for cancer patients like ACT or meaning-based interventions in psycho-oncology can reduce mental health problems and have positive effects on recovery and survival rates (Lamers et al., 2012; Vehling et al., 2011). Assessing the current status of the PMH of patients can be a starting point for selecting appropriate interventions for patients.

Overall, the PMH-scale showed a good model fit and excellent reliability after making some modifications due to LD and excluding one item. The excluded item was item 9, which displayed an item residual of 4.95. In contrast to our study, this item showed adequate factor loading in CTT studies (Bieda et al., 2017; Lukat et al., 2016), even though it had by far the smallest loadings. Compared in context to the other items of

the scale, it appears that item 9 ('I am a calm, balanced human being') assesses two different aspects. One can be hectic but still be balanced (i. e., exhibit positive mental health). Moreover, it seems to reflect trait character to a higher degree than the other items. According to the Rasch model, this trait character could be the reason why item 9 was misfitting in our analysis. Further research in other samples is needed to further investigate the fit of this item.

Furthermore, the scale contained several pairs of locally dependent items. After combining the locally dependent item pairs successively into 'superitems', no more LD was observed. In terms of content, the observed LD within the scale makes sense since items 1&2 are facets of enjoying life, items 3&4 assess satisfaction in the present and future, and items 6&7 are concerned with mastering daily life. In a study with a cross-cultural sample, a similar dependence was found between Item 1 and Item 2, and the same conclusion was drawn that these items relate to facets of enjoyment of life (Bieda et al., 2017). Since there are no other studies using Rasch analysis, future studies should also focus on investigating LD in the PMH-scale, given the influence of LD on parameter estimation and reliability.

DIF was tested in relation to gender, age, type of cancer, the presence of metastases, psycho-oncological support, and duration of disease. For most of these external variables, no DIF was found. However, uniform age-DIF was found for 'superitem' 1&2 and item 5. As the DIF for 'superitem' 1&2 was no longer present after splitting item 5 for DIF related to age, this might indicate that this DIF was artificial (Andrich & Hagquist, 2015). To evaluate the impact of the age-related DIF found for item 5, equated scores were computed. We only found a relatively small inconsiderably difference in the equated scores between the younger and older patients, with a maximum score difference of about 0.5 points in the lower range of the person location. This result shows an indication that patients with the same level of well-being responded differently to the managing to fulfill their needs item depending on their age. Specifically, elderly individuals seemed to have more ease in this field than younger persons with the same well-being level. However, this difference becomes visible only in the lower range. In contrast, patients with either a high or middle level of well-being responded comparable in the areas of high or middle level of well-being, irrespective of their age. Given the minor impact of DIF and given that it was only found in a tiny part of the assessed dimension, we decided not to adjust for DIF. Note that our sample

is relatively young, with a mean age of 52.40 years. In a sample with more elderly patients, a more relevant age-DIF might be found.

The conclusion on unidimensionality is consistent with other CTT analyses of the PMH-scale (Bibi et al., 2020; Bieda et al., 2017; Lukat et al., 2016). Overall, the targeting of the PMH-8 scale was good for the present sample of cancer patients. The PMH showed a widespread distribution of item thresholds that ensured good measurement accuracy across a large portion of the PMH dimension. However, for low and high PMH levels, the targeting was not as good as item thresholds were missing in these areas of the dimension. The PMH-scale was initially developed to provide a unidimensional assessment of PMH in the general population. Our results indicate that the differentiation in the higher segment of well-being is not equally good – an area where probably most of the people of a healthy population would be located. However, the differentiation within a healthy population or persons with a high, respectively a very high level of PMH may not be so relevant for the assessment of oncology patients with regard to clinical decision-making in psycho-oncology. Easier items are also missing, making it also hard to precisely assess PMH at a low level of well-being. It might be attractive in future research to either include some more items or to develop a better targeted scale for patients with low levels of well-being (e.g., with items related to other facets of mental health like life affirmation or meaning of life). This potential revision could be used, for example, to have a first starting point for resource-activation work with patients in psycho-oncological interventions. However, given the heterogeneity of individuals and their variability in perceiving the benefits of an intervention and their response to it, it is critical to identify individual variation in clinical significance of change in health care. Therefore, concepts of clinical significance of change are increasingly being used to improve change measurement and clinical decision making. Future studies should consider the clinical significance of the scale by also examining its use in the clinical setting based on individual significance.

Besides some strengths, the present study also has some limitations. The sample consisted of a relatively high percentage of breast cancer patients. The residual cancer types had to be combined into one category, 'other forms of cancer' for the DIF analysis due to small subgroup sizes. Accordingly, the results may only be generalized to other cancer patients with caution. Future studies with larger samples and higher proportions of different cancer types should be investigated, especially with regard to gender-

specific cancer diagnoses and thus a possible gender DIF. However, in our analysis, we found no evidence of a gender-DIF. Future research is also needed regarding the influence of different cancer types, especially those with a more severe disease progress. Nevertheless, the presence of metastases or the disease duration could also be used as an indication of severity. We examined both in our study, and both of them showed no DIF. Furthermore, the DIF analysis with cancer types was included because, in addition to breast cancer, reporting multiple cancer types could be an indicator of more severe disease. Additionally, the sample's psychological distress (HADS-T) is roughly equally distributed across the cancer forms. Therefore, one can assume that the type of cancer does not unduly influence the response behavior. Furthermore, the recruited sample is relatively young, with a mean age of 52.4 years. This may be the result of the recruitment procedure. The sample was recruited from social media platforms and from online cancer support groups. The scale assesses a wide range of well-being, but for the present sample it shows a slight mistargeting and an off-center distribution of persons with a relatively high frequency of persons with a high PMH level, which may indicate a bias in this sample. Also, a high percentage (41.2%) of the cancer patients had an active job situation, indicating a relative fit sample. Concerning this and the small age-DIF we found in our study, future research should examine a sample with a lower level of mental health and perhaps include some additional items suited for assessing lower and higher levels of PMH.

Conclusion

The present study provides basic information about the psychometric properties of the PMH-scale in the oncological context. The Rasch analysis showed that this scale can be used well in this context; in particular, it adequately captures individuals with intermediate PMH scores. However, the scale should be further investigated for its targeting, and better targeted items may need to be added to capture the full range of the PMH dimension. Given that PMH can predict mental health problems and positively impact recovery and survival rates, these findings are useful, especially for selecting appropriate interventions for patients. The instrument is non-biased with respect to gender, type of cancer, the presence of metastases, psycho-oncological support, and duration of disease. However, with regard to age, especially in elderly people, a critical consideration might be necessary.

3. Summary discussion and outlook

The aim of this thesis is to address the discrepancy between common methodological recommendations and empirical practice and to critically reflect three examples from the research field of psycho-oncology, illustrating which methodological principles are proposed in science and what is actually applied. Since the beginnings of experimental psychology, methodological principles and approaches have been developed and applied to examine non-visible, latent features of people's experience and behavior. As the quote from Wilhelm Wundt at the beginning of the thesis points out, methodology is a central component of science.

As a result of changing social understanding and acceptance of psychological mechanisms in western societies, a variety of clinical approaches have evolved, new fields of research have emerged, and existing methods for assessing human experience and behavior have been refined and expanded. This change of times and the accompanying innovations can be seen very well in the studies presented in this thesis.

Firstly, scientific advances and increasing public awareness of psychological processes and illnesses allowed the establishment of a new specialty, psycho-oncology, which addresses the experience and behavior associated with cancer, its treatment, and accompanying problems. Individualized psycho-oncology has increasingly become the focus of clinical health care research, and the importance for individualized approaches is enormous, depending on disease severity and special needs of the patients and how they individually respond to treatment. Measuring the effectiveness of interventions and associated changes may also require an individualized approach to measurement of changes, namely the inclusion of clinical significance.

The first study contrasted the most common methods of clinical significance and critically considered which methods should be used in the psycho-oncological context. The aim was not the evaluation of psycho-oncological treatment, but rather a comparison of two commonly used methods for determination of clinical significance. The careful selection of an optimal concept for clinical significance is essential for research and clinical practice. These findings have important potential implications since results may differ substantially when studying the efficacy and effectiveness of

psycho-oncological care on an individual level. Moreover, whenever the welfare of individual patients is considered, it is also essential to be able to make the right clinical decisions, ensuring the best possible care for everyone. In research, there have been some studies comparing different methods of clinical significance measures (e.g., Atkins et al., 2005; Hsu, 1999). For example, Bauer et al. (2004) compare five different approaches to clinical significance and found differences between methods in estimates of meaningful change. The authors also recommend the use of RCI in outcome studies and in research on clinically significant change and call for further research. The first article in the thesis joined this scientific effort and fills another gap by comparing two common methods of measuring clinical significance. This study highlights the advantages of using the RCI over the MID, which tends to overestimate effects and is therefore associated with bias in clinical decision making.

It is important to have valid assessment instruments to guide selection for and effects of clinical interventions. Selecting the most appropriate measurement is challenging given the growing number of assessment instruments. However, the success of effectiveness research and studies about decision-making depends on the selection of valid and reliable assessment instruments (Meyer et al., 2014). Therefore, in the second part of the thesis, a relatively novel methodological approach is investigated, namely the evaluation of instruments using methods of IRT.

The psychometric properties of assessment instruments can be evaluated with both CTT and IRT models. Advantages of IRT testing are that it is possible to obtain detailed information about the scale or items and about the person's ability. With IRT-based analysis, it is possible to make statements about whether the items represent the latent trait and form a unidimensional scale (Streiner, 2010). Furthermore, it is possible to evaluate the response categories and whether the corresponding item indicates the latent characteristic in different ways in different groups (Tennant & Conaghan, 2007). Due to these prevailing advantages, the use of IRT-based models is called for. However, since IRT approaches are much more complex and not easy to implement, they are not yet widely used in practice. The second study presented here aimed to investigate the psychometric properties of a widely used instrument using the IRT-based approach, i.e., Rasch-analysis among patients afflicted by various types of cancer. The WHODAS 2.0 (Üstün et al., 2010) is a recommended tool for measuring health and disability. Its usage is continuously increasing, but research on the

psychometric properties in oncological context is rare, i.e., so far, no research with modern psychometric analysis (Rasch analysis) has been conducted within a sample of cancer patients. Cancer patients have to cope with their diagnosis and master disease-associated tasks and challenges and, in addition, they can also suffer from disability. A valid and reliable assessment of self-reported disability is essential for the planning of an individual psycho-oncological treatment and the selection of adequate therapeutic interventions. Furthermore, a valid and reliable assessment instrument allows evaluation of effectiveness of these interventions. Overall, the study provides interesting detailed information using modern approach, i.e., Rasch analysis, about the WHODAS 2.0 12-item version and proved to be a suitable measure of disability in cancer patients.

Another important approach that has recently been gaining interest is the positive mental health approach. Positive psychological interventions aim at enhancing well-being. Improvements in mental health are associated with the prediction of decrease in mental illness (Keyes et al., 2010). Such positive effects have also been shown among cancer patients (Schiavon et al., 2017).

The increased interest in positive mental health motivated the development of several assessment instruments. One of them combines two approaches, i.e., the hedonic and eudaemonic aspects of mental health and is a brief self-rating scale, namely the PMH scale. The PMH scale is increasingly being used, but research on the psychometric properties in oncological context does not exist, especially no research with modern psychometric analysis, i.e., Rasch analysis. Since psycho-oncological interventions and cancer patients may benefit from positive effects of positive mental health improvement with respect to recovery and survival rates and as a protective factor (Lamers et al., 2012; Schiavon et al., 2017), it is important to consider the PMH-scale application in the oncological context as well. Therefore, the third study of the present work aimed to investigate the psychometric properties of the 9-item PMH scale in patients affected by different types of cancer using Rasch analysis. It turns out that the 8-item solution fits the model better, in contrast to most CTT studies (Bieda et al., 2017; Lukat et al., 2016). To the authors' knowledge, this study is the first to examine the psychometric properties of the PMH scale using the Rasch model. Further studies are needed to compare the results and the potential item solution.

Overall, this present work is about the critical reflection of methods based on three examples from the research field of psycho-oncology. In psychology in general, however, the same demands are made on methods and procedures. But the application of the requirements in research practice is another matter. Therefore, it is possible to apply the methodological critical reflection to other fields of psychology. It must be said that the presented and reflected methods, i.e., clinical significance, evaluation of psychometric properties with Rasch analysis, are only a selection of essential, frequently used methods relevant for both research and clinical application.

Further studies on advanced methodology are needed and the application of the requirements in practice should continue to be investigated and applied. In addition, the methods discussed in this work are interesting to study in other samples. In particular, with regard to the PMH scale, since item reduction is recommended based on the third study. Studies in other samples on this are already planned. All in all, any research and thus any clinical implication can only be as good as the method used. Therefore, we should not forget the roots of experimental psychology and dedicate ourselves to the given requirements and use them as an extension of many existing and well-developed methods.

4. References

Ahmed, S., Berzon, R. A., Revicki, D. A., Lenderking, W. R., Moinpour, C. M., Basch, E., . . . International Society for Quality of Life Research. (2012). The Use of Patient-reported Outcomes (PRO) Within Comparative Effectiveness Research: Implications for Clinical Practice and Health Care Policy. *Medical Care*, *50*(12), 1060-1070. https://doi.org/10.1097/MLR.0b013e318268aaff

Akechi, T., Okuyama, T., Onishi, J., Morita, T., & Furukawa, T. A. (2008). Psychotherapy for depression among incurable cancer patients. *Cochrane Database of Systematic Reviews*, *2008*(2), Cd005537. https://doi.org/10.1002/14651858.CD005537.pub2

American Psychiatric Association [APA]. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed* [doi:10.1176/appi.books.9780890425596]. American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*, *11*(5), 571-585. https://doi.org/10.1586/erp.11.59

Andrich, D. (2013). An Expanded Derivation of the Threshold Structure of the Polytomous Rasch Model That Dispels Any "Threshold Disorder Controversy". *Educational and Psychological Measurement*, *73*(1), 78-124. https://doi.org/10.1177/0013164412450877

Andrich, D. (2016). Components of Variance of Scales With a Bifactor Subscale Structure From Two Calculations of α. *Educational Measurement: Issues and Practice*, *35*(4), 25-30. https://doi.org/10.1111/emip.12132

Andrich, D., & Hagquist, C. (2015). Real and Artificial Differential Item Functioning in Polytomous Items. *Educational and Psychological Measurement*, *75*(2), 185-207. https://doi.org/10.1177/0013164414534258

Andrich, D., Sheridan, B., & Luo, G. (2004). *Interpreting RUMM2030*. RUMM Laboratory.

Andrich, D., Sheridan, B., & Luo, G. (2009). RUMM 2030. In. Perth, WA: RUMM Laboratory.

Arem, H., & Loftfield, E. (2017). Cancer Epidemiology: A Survey of Modifiable Risk Factors for Prevention and Survivorship. *American journal of lifestyle medicine*, *12*(3), 200-210. https://doi.org/10.1177/1559827617700600

Atkins, D. C., Bedics, J. D., McGlinchey, J. B., & Beauchaine, T. P. (2005). Assessing clinical significance: does it matter which method we use? *Journal of Consulting*

and Clinical Psychology, *73*(5), 982-989. https://doi.org/10.1037/0022-006x.73.5.982

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: a comparison of statistical techniques. *Journal of Personality Assessment*, *82*(1), 60-70. https://doi.org/10.1207/s15327752jpa8201_11

Bedard, G., Zeng, L., Zhang, L., Lauzon, N., Holden, L., Tsao, M., . . . Chow, E. (2013). Minimal Clinically Important Differences in the Edmonton Symptom Assessment System in Patients With Advanced Cancer. *Journal of Pain and Symptom Management*, *46*(2), 192-200. https://doi.org/10.1016/j.jpainsymman.2012.07.022

Bertolote, J. (2008). The roots of the concept of mental health. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *7*(2), 113-116. https://doi.org/10.1002/j.2051-5545.2008.tb00172.x

Bibi, A., Lin, M., & Margraf, J. (2020). Salutogenic constructs across Pakistan and Germany: A cross sectional study. *International Journal of Clinical and Health Psychology*, *20*(1), 1-9. https://doi.org/10.1016/j.ijchp.2019.10.001

Bieda, A., Hirschfeld, G., Schönfeld, P., Brailovskaia, J., Zhang, X. C., & Margraf, J. (2017). Universal happiness? Cross-cultural measurement invariance of scales assessing positive mental health. *Psychological Assessment*, *29*(4), 408-421. https://doi.org/10.1037/pas0000353

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69-77. https://doi.org/10.1016/S0022-3999(01)00296-3

Bjorner, J. B. (2019). State of the psychometric methods: comments on the ISOQOL SIG psychometric papers. *Journal of Patient-Reported Outcomes*, *3*(1), 49. https://doi.org/10.1186/s41687-019-0134-1

Blanchin, M., Hardouin, J. B., Le Neel, T., Kubis, G., Blanchard, C., Mirallié, E., & Sébille, V. (2011). Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine*, *30*(8), 825-838. https://doi.org/10.1002/sim.4153

Bothe, A. K., & Richardson, J. D. (2011). Statistical, Practical, Clinical, and Personal Significance: Definitions and Applications in Speech-Language Pathology. *American Journal of Speech-Language Pathology*, *20*(3), 233-242. https://doi.org/10.1044/1058-0360(2011/10-0034)

Boyes, A. W., Girgis, A., D'Este, C. A., Zucca, A. C., Lecathelinais, C., & Carey, M. L. (2013). Prevalence and predictors of the short-term trajectory of anxiety and

depression in the first year after a cancer diagnosis: a population-based longitudinal study. *J Clin Oncol*, *31*(21), 2724-2729. https://doi.org/10.1200/jco.2012.44.7540

Brailovskaia, J., & Margraf, J. (2020). Predicting adaptive and maladaptive responses to the Coronavirus (COVID-19) outbreak: A prospective longitudinal study. *International Journal of Clinical and Health Psychology*, *20*(3), 183-191. https://doi.org/10.1016/j.ijchp.2020.06.002

Brailovskaia, J., Teismann, T., & Margraf, J. (2020). Positive Mental Health, Stressful Life Events, and Suicide Ideation. *Crisis*, *41*(5), 383-388. https://doi.org/10.1027/0227-5910/a000652

Breitbart, W., Rosenfeld, B., Pessin, H., Applebaum, A., Kulikowski, J., & Lichtenthal, W. G. (2015). Meaning-centered group psychotherapy: an effective intervention for improving psychological well-being in patients with advanced cancer. *Journal of clinical oncology*, *33*(7), 749-754. https://doi.org/10.1200/JCO.2014.57.2198

Breitbart, W. S., & Alici, Y. (2009). Psycho-Oncology. *Harvard Review of Psychiatry*, *17*(6), 361-376. https://doi.org/10.3109/10673220903465700

Bultz, B. D. (2016). Cancer Care and the Role of Psychosocial Oncology: Where are We and Where are We Going? *Asia-Pacific journal of oncology nursing*, *3*(2), 118-120. https://doi.org/10.4103/2347-5625.182929

Bundesministerium für Gesundheit. (2017). Nationaler Krebsplan: Handlungsfelder, Ziele, Umsetzungsempfehlungen und Ergebnisse. In. Bonn: Bundesministerium für Gesundheit.

Bussmann, S., Vaganian, L., Kusch, M., Labouvie, H., Gerlach, A. L., & Cwik, J. C. (2018). Angst und Emotionsregulation bei Krebspatienten. *Psychotherapeut*, *63*(3), 194-203. https://doi.org/10.1007/s00278-018-0283-3

Cameron, I. M., Scott, N. W., Adler, M., & Reid, I. C. (2014). A comparison of three methods of assessing differential item functioning (DIF) in the Hospital Anxiety Depression Scale: ordinal logistic regression, Rasch analysis and the Mantel chi-square procedure. *Quality of Life Research*, *23*(10), 2883-2888. https://doi.org/10.1007/s11136-014-0719-3

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). 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. https://doi.org/10.1016/j.clinthera.2014.04.006

Casellas-Grau, A., Font, A., & Vives, J. (2014). Positive psychology interventions in breast cancer. A systematic review. *Psychooncology*, *23*(1), 9-19. https://doi.org/10.1002/pon.3353

Chakhssi, F., Kraiss, J. T., Sommers-Spijkerman, M., & Bohlmeijer, E. T. (2018). The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatry*, *18*(1), 211. https://doi.org/10.1186/s12888-018-1739-2

Chan, C. M. H., Ahmad, W. A. W., Yusof, M. M., Ho, G. F., & Krupat, E. (2015). Effects of depression and anxiety on mortality in a mixed cancer group: a longitudinal approach using standardised diagnostic interviews. *Psycho-Oncology*, *24*(6), 718-725. https://doi.org/doi:10.1002/pon.3714

Chen, C. M., Cano, S. J., Klassen, A. F., King, T., McCarthy, C., Cordeiro, P. G., . . . Pusic, A. L. (2010). Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast Journal*, *16*(6), 587-597. https://doi.org/10.1111/j.1524-4741.2010.00983.x

Christensen, K. B., Kreiner, S., & Mesbah, M. (2013). *Rasch Models in Health*. Wiley.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, *41*(3), 178-194. https://doi.org/10.1177/0146621616677520

Christensen, K. B., Thorborg, K., Hölmich, P., & Clausen, M. B. (2019). Rasch validation of the Danish version of the shoulder pain and disability index (SPADI) in patients with rotator cuff-related disorders. *Quality of Life Research*, *28*(3), 795-800. https://doi.org/10.1007/s11136-018-2052-8

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Earlbaum Associates.

Cook, S. A., Salmon, P., Hayes, G., Byrne, A., & Fisher, P. L. (2018). Predictors of emotional distress a year or more after diagnosis of cancer: A systematic review of the literature. *Psycho-Oncology*, *27*(3), 791-801. https://doi.org/doi:10.1002/pon.4601

Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., & Schuler, T. C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*, *7*(5), 541-546. https://doi.org/https://doi.org/10.1016/j.spinee.2007.01.008

Crocker, L. M., & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*. Cengage Learning.

Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, *56*. https://doi.org/10.1016/s0895-4356(03)00044-1

Cwik, J. C., Vaganian, L., Bussmann, S., Labouvie, H., Houwaart, S., Gerlach, A. L., & Kusch, M. (2021). Assessment of coping with cancer-related burdens: psychometric properties of the Cognitive-Emotional Coping with Cancer scale and the German Mini-mental Adjustment to Cancer scale. *Journal of Psychosocial Oncology Research and Practice*, *3*(1), e046. https://doi.org/10.1097/or9.0000000000000046

da Rocha, N. S., Chachamovich, E., de Almeida Fleck, M. P., & Tennant, A. (2013). An introduction to Rasch analysis for Psychiatric practice and research. *Journal of Psychiatric Research*, *47*(2), 141-148. https://doi.org/https://doi.org/10.1016/j.jpsychires.2012.09.014

de Beurs, E., Carlier, I. V. E., & van Hemert, A. M. (2019). Approaches to denote treatment outcome: Clinical significance and clinical global impression compared. *International Journal of Methods in Psychiatric Research*, *28*(4), e1797. https://doi.org/https://doi.org/10.1002/mpr.1797

Deutsche Krebsgesellschaft, Deutsche Krebshilfe, & AWMF. (2014). *S3-Leitlinie Psychoonkologische Diagnostik Beratung und Behandlung von erwachsenen Krebspatienten.* Retrieved 06.06.2021 from https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Psychoonkologieleitlinie_1.1/LL_PSO_Langversion_1.1.pdf

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *J Pers Assess*, *49*(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health*, *1*(3), 70-72. https://doi.org/10.1136/ebmh.1.3.70

Faller, H., Schuler, M., Richard, M., Heckl, U., Weis, J., & Küffner, R. (2013). Effects of Psycho-Oncologic Interventions on Emotional Distress and Quality of Life in Adult Patients With Cancer: Systematic Review and Meta-Analysis. *Journal of Clinical Oncology*, *31*(6), 782-793. https://doi.org/10.1200/jco.2011.40.8922

Fann, J. R., Ell, K., & Sharpe, M. (2012). Integrating Psychosocial Care Into Cancer Services. *Journal of Clinical Oncology*, *30*(11), 1178-1186. https://doi.org/10.1200/jco.2011.39.7398

Fava, G. A., Rafanelli, C., Cazzaro, M., Conti, S., & Grandi, S. (1998). Well-being therapy. A novel psychotherapeutic approach for residual symptoms of affective

disorders. *Psychological Medicine*, *28*(2), 475-480. https://doi.org/10.1017/s0033291797006363

Federici, S., Bracalenti, M., Meloni, F., & Luciano, J. V. (2017). World Health Organization disability assessment schedule 2.0: An international systematic review. *Disability and Rehabilitation*, *39*(23), 2347-2380. https://doi.org/10.1080/09638288.2016.1223177

Ferguson, R. J., Robinson, A. B., & Splaine, M. (2002). Use of the Reliable Change Index to Evaluate Clinical Significance in SF-36 Outcomes. *Quality of Life Research*, *11*(6), 509-516.

Fischer, G. H. (1987). Applying the principles of specific objectivity and of generalizability to the measurement of change. *Psychometrika*, *52*(4), 565-587. https://doi.org/10.1007/bf02294820

Fledderus, M., Bohlmeijer, E. T., Smit, F., & Westerhof, G. J. (2010). Mental health promotion as a new goal in public mental health care: a randomized controlled trial of an intervention enhancing psychological flexibility. *American Journal of Public Health*, *100*(12), 2372. https://doi.org/10.2105/ajph.2010.196196

Forsythe, L. P., Rowland, J. H., Padgett, L., Blaseg, K., Siegel, S. D., Dingman, C. M., & Gillis, T. A. (2013). The cancer psychosocial care matrix: a community-derived evaluative tool for designing quality psychosocial cancer care delivery. *Psycho-Oncology*, *22*(9), 1953-1962. https://doi.org/doi:10.1002/pon.3254

Garssen, B. (2004). Psychological factors and cancer development: evidence after 30 years of research. *Clinical Psychology Review*, *24*(3), 315-338. https://doi.org/10.1016/j.cpr.2004.01.002

Glasgow, R. E., Kaplan, R. M., Ockene, J. K., Fisher, E. B., & Emmons, K. M. (2012). Patient-Reported Measures Of Psychosocial Issues And Health Behavior Should Be Added To Electronic Health Records. *Health Affairs*, *31*(3), 497-504. https://doi.org/10.1377/hlthaff.2010.1295

Gold, L. H. (2014). DSM-5 and the Assessment of Functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online*, *42*(2), 173-181. http://jaapl.org/content/jaapl/42/2/173.full.pdf

Grassi, L., Spiegel, D., & Riba, M. (2017). Advancing psychosocial care in cancer patients. *F1000Research*, *6*, 2083. https://doi.org/10.12688/f1000research.11902.1

Gustafsson, J.-E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *33*(2), 205-233. https://doi.org/10.1111/j.2044-8317.1980.tb00609.x

Gustavsson-Lilius, M., Julkunen, J., & Hietanen, P. (2007). Quality of Life in Cancer Patients: The Role of Optimism, Hopelessness, and Partner Support. *Quality of Life Research*, *16*(1), 75-87. http://www.jstor.org/stable/27641256

Guyatt, G., Osoba, D., Wu, A. W., Wyrwich, K. W., & Norman, G. R. (2002). Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings*, *77*. https://doi.org/10.4065/77.4.371

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and quality of life outcomes*, *15*(1), 181-181. https://doi.org/10.1186/s12955-017-0755-0

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Hays, R. D., Brodsky, M., Johnston, M. F., Spritzer, K. L., & Hui, K.-K. (2005). Evaluating the Statistical Significance of Health-Related Quality-Of-Life Change in Individual Patients. *Evaluation & the Health Professions*, *28*(2), 160-171. https://doi.org/10.1177/0163278705275339

Hays, R. D., Farivar, S. S., & Liu, H. (2005). Approaches and Recommendations for Estimating Minimally Important Differences for Health-Related Quality of Life Measures. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *2*(1), 63-67. https://doi.org/10.1081/COPD-200050663

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical care*, *38*, II28-42. https://doi.org/10.1097/00005650-200009002-00007

Hays, R. D., Spritzer, K. L., Sherbourne, C. D., Ryan, G. W., & Coulter, I. D. (2018). Group and Individual-level Change on Health-related Quality of Life in Chiropractic Patients With Chronic Low Back or Neck Pain. *Spine*, *44*(9), 647-651. https://doi.org/10.1097/brs.0000000000002902

Herrmann-Lingen, C., Buss, U., & Snaith, P. R. (2011). *Hospital Anxiety and Depression Scale-Deutsche Version (HADS-D)*. Huber.

Herschbach, P., Marten-Mittag, B., & Henrich, G. (2003). Revision und psychometrische Prüfung des Fragebogen zur Belastung von Krebskranken (FBK-R23). *Zeitschrift für Medizinische Psychologie*, *12*, 69-76.

Hinz, A., & Brähler, E. (2011). Normative values for the Hospital Anxiety and Depression Scale (HADS) in the general German population. *Journal of Psychosomatic Research*, *71*, 74-78. https://doi.org/10.1016/j.jpsychores.2011.01.005

Holland, J. C. (2002). History of psycho-oncology: overcoming attitudinal and conceptual barriers. *Psychosomatic Medicine*, *64*(2), 206-221. https://doi.org/10.1097/00006842-200203000-00004

Holland, J. C., Breitbart, W., Butow, P., Jacobsen, P., Loscalzo, M., & McCorkle, R. (2015). *Psycho-oncology* (3 ed.). Oxford University Press. https://books.google.de/books?id=oy9mBgAAQBAJ

Holtmaat, K., van der Spek, N., Lissenberg-Witte, B. I., Cuijpers, P., & Verdonck-de Leeuw, I. M. (2019). Positive mental health among cancer survivors: overlap in psychological well-being, personal meaning, and posttraumatic growth. *Supportive Care in Cancer*, *27*(2), 443-450. https://doi.org/10.1007/s00520-018-4325-8

Hsu, L. M. (1999). A comparison of three methods of identifying reliable and clinically significant client changes: commentary on Hageman and Arrindell. *Behaviour Research and Therapy*, *37*(12), 1195-1202; discussion 1219-1133. https://doi.org/10.1016/s0005-7967(99)00033-9

IBM Corporation. (Released 2019). IBM SPSS Statistics for Windows, Version 26.0. In. Armonk, NY: IBM Corp.

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied psychological measurement*, *40*(8), 559-572. https://doi.org/10.1177/0146621616664046

Jacobsen, P. B. (2007). Screening for Psychological Distress in Cancer Patients: Challenges and Opportunities. *Journal of Clinical Oncology*, *25*(29), 4526-4527. https://doi.org/10.1200/jco.2007.13.1367

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*(3), 300-307. https://doi.org/10.1037//0022-006x.67.3.300

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, *10*(4), 407-415. https://doi.org/https://doi.org/10.1016/0197-2456(89)90005-6

Jenniches, I., Lemmen, C., Cwik, J. C., Kusch, M., Labouvie, H., Scholten, N., . . . Dresen, A. (2020). Evaluation of a complex integrated, cross-sectoral psycho-oncological care program (isPO): a mixed-methods study protocol. *BMJ open*, *10*(3), e034141-e034141. https://doi.org/10.1136/bmjopen-2019-034141

Jensen, S. A., & Corralejo, S. M. (2017). Measurement Issues: Large effect sizes do not mean most people get better – clinical significance and the importance of individual results. *Child and Adolescent Mental Health*, *22*(3), 163-166. https://doi.org/https://doi.org/10.1111/camh.12203

Karakas, S. A., & Okanli, A. (2014). The relationship between meaning of illness, anxiety, depression, and quality of life for cancer patients. *Collegium antropologicum*, *38*(3), 939-944. http://www.ncbi.nlm.nih.gov/pubmed/25420377

Kash, K. M., Mago, R., Duffany, S., & Kunkel, E. J. (2006). Psycho-oncology: review and update. *Current Psychiatry Reports*, *8*(3), 246-252. https://doi.org/10.1007/s11920-006-0031-1

Kendall, P. C., Comer, J. S., & Chow, C. (2013). The Randomized Controlled Trial: Basic and Beyond. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford Handbook of Research Strategies for Clinical Psychology* (pp. 40-61). Oxford University Press.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, *73*(3), 539-548. https://doi.org/10.1037/0022-006x.73.3.539

Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *American Psychologist*, *62*(2), 95-108. https://doi.org/10.1037/0003-066x.62.2.95

Keyes, C. L. M., Dhingra, S. S., & Simoes, E. J. (2010). Change in level of positive mental health as a predictor of future risk of mental illness. *American journal of public health*, *100*(12), 2366-2371. https://doi.org/10.2105/AJPH.2010.192245

King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research*, *11*(2), 171-184. https://doi.org/10.1586/erp.11.9

Kirchberger, I., Braitmayer, K., Coenen, M., Oberhauser, C., & Meisinger, C. (2014). Feasibility and psychometric properties of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0) in a population-based sample of patients with myocardial infarction from the MONICA/KORA myocardial infarction registry. *Population Health Metrics*, *12*(1), 27. https://doi.org/10.1186/s12963-014-0027-8

Kost, C., Labouvie, H., Kosfelder, J., Höhl, U., Herrmann-Lingen, C., & Kusch, M. (2009). Veränderung der psychischen Belastung von Krebspatienten im Rahmen einer strukturierten psychoonkologischen Versorgung. [Change in psychological distress of cancer in-patients attending a structured psychooncological care program.]. *Psychotherapie Psychosomatik Medizinische Psychologie, 59*(12), 432-439. https://doi.org/10.1055/s-2008-1067575

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of Clinical Significance. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(12), 1524-1529. https://doi.org/10.1097/00004583-200312000-00022

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Küçükdeveci, A. A., Kutlay, Ş., Yıldızlar, D., Öztuna, D., Elhan, A. H., & Tennant, A. (2013). The reliability and validity of the World Health Organization Disability Assessment Schedule (WHODAS-II) in stroke. *Disabil Rehabil, 35*(3), 214-220. https://doi.org/10.3109/09638288.2012.690817

Kusch, M., Labouvie, H., Gerlach, A. L., Hellmich, M., & Hallek, M. (2016). Kommt die personalisierte Psychoonkologie? *Tumor Diagnostik und Therapie, 37*(10), 560-564. https://doi.org/10.1055/s-0042-118685

Kusch, M., Labouvie, H., & Hein-Nau, B. (2013). *Klinische Psychoonkologie*. Springer-Verlag.

Kusch, M., Labouvie, H., Hein-Nau, B., Schwarzkamp, U., Wolf, J., & Hallek, M. (2014). Integrierte Psychoonkologie: Implementierung eines psychoonkologischen Versorgungsprogramms am Centrum für Integrierte Onkologie Köln Bonn, Standort Köln [Integrated psychooncology: Implementation of psychooncological health care at the Center for Integrated Oncology Cologne – Bonn]. *Deutsche medizinische Wochenschrift, 139*(46), 2357-2360. https://doi.org/10.1055/s-0034-1387384

Kutlay, S., Küçükdeveci, A. A., Elhan, A. H., Oztuna, D., Koç, N., & Tennant, A. (2011). Validation of the World Health Organization disability assessment schedule II (WHODAS-II) in patients with osteoarthritis. *Rheumatol Int, 31*(3), 339-346. https://doi.org/10.1007/s00296-009-1306-8

Lambert, M. J., & Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: The need for a common procedure and validity data. *Psychotherapy Research, 19*(4-5), 493-501. https://doi.org/10.1080/10503300902849483

Lamers, S. M. A., Bolier, L., Westerhof, G. J., Smit, F., & Bohlmeijer, E. T. (2012). The impact of emotional well-being on long-term recovery and survival in physical illness: a meta-analysis. *Journal of behavioral medicine*, *35*(5), 538-547. https://doi.org/10.1007/s10865-011-9379-8

Leiner, D. J. (2014). *SoSci Survey (Version 2.4.00-i) [Computer Software]. Available at https://www.soscisurvey.de*. In

Lenz, A. S. (2021). Clinical Significance in Counseling Outcome Research and Program Evaluation. *Counseling Outcome Research and Evaluation*, *12*(1), 1-3. https://doi.org/10.1080/21501378.2021.1877097

Levenson, J. L., & Bemis, C. (1991). The Role of Psychological Factors in Cancer Onset and Progression. *Psychosomatics*, *32*(2), 124-132. https://doi.org/https://doi.org/10.1016/S0033-3182(91)72083-5

Linden, W., Vodermaier, A., Mackenzie, R., & Greig, D. (2012). Anxiety and depression after cancer diagnosis: prevalence rates by cancer type, gender, and age. *J Affect Disord*, *141*(2-3), 343-351. https://doi.org/10.1016/j.jad.2012.03.025

Lohr, K. N., & Steinwachs, D. M. (2002). Health Services Research: An Evolving Definition of the Field. *Health Services Research*, *37*(1), 15-17. https://doi.org/10.1111/1475-6773.01020

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care*, *46*(3), 266-274. https://doi.org/10.1097/MLR.0b013e318160d093

Luciano, J. V., Ayuso-Mateos, J. L., Aguado, J., Fernandez, A., Serrano-Blanco, A., Roca, M., & Haro, J. M. (2010). The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): a nonparametric item response analysis. *BMC Medical Research Methodology*, *10*(1), 45. https://doi.org/10.1186/1471-2288-10-45

Luciano, J. V., Ayuso-Mateos, J. L., Fernández, A., Serrano-Blanco, A., Roca, M., & Haro, J. M. (2010). Psychometric properties of the twelve item World Health Organization Disability Assessment Schedule II (WHO-DAS II) in Spanish primary care patients with a first major depressive episode. *Journal of Affective Disorders*, *121*(1), 52-58. https://doi.org/https://doi.org/10.1016/j.jad.2009.05.008

Lukat, J., Becker, E. S., Lavallee, K. L., van der Veld, W. M., & Margraf, J. (2017). Predictors of Incidence, Remission and Relapse of Axis I Mental Disorders in Young Women: A Transdiagnostic Approach. *Clin Psychol Psychother*, *24*(2), 322-331. https://doi.org/10.1002/cpp.2026

Lukat, J., Margraf, J., Lutz, R., van der Veld, W., & Becker, E. S. (2016). Psychometric properties of the positive mental health scale (PMH-scale). *BMC Psychology*, *4*(8), 1-14. https://doi.org/10.1186/s40359-016-0111-x

Lyubomirsky, S., & Lepper, H. S. (1999). A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research*, *46*(2), 137-155. https://doi.org/10.1023/A:1006824100041

Magistrale, G., Pisani, V., Argento, O., Incerti, C., Bozzali, M., Cadavid, D., . . . Nocentini, U. (2015). Validation of the World Health Organization Disability Assessment Schedule II (WHODAS-II) in patients with multiple sclerosis. *Multiple Sclerosis Journal*, *21*, 448-456. https://doi.org/10.1177/1352458514543732

Marais, I. (2013). Local Dependence. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch Models in Health* (pp. 111-130). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781118574454.ch7

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174. https://doi.org/10.1007/BF02296272

Mehnert-Theuerkauf, A., & Lehmann-Laue, A. (2019). Psychoonkologie [Psychoonkologie]. *Psychother Psychosom Med Psychol*, *69*(03/04), 141-156. https://doi.org/10.1055/a-0824-3622

Mehnert, A., Braack, K., & Vehling, S. (2011). Sinnorientierte Interventionen in der Psychoonkologie [journal article]. *Psychotherapeut*, *56*(5), 394 - 399. https://doi.org/10.1007/s00278-011-0857-9

Meyer, A. M., Wheeler, S. B., Weinberger, M., Chen, R. C., & Carpenter, W. R. (2014). An overview of methods for comparative effectiveness research. *Semin Radiat Oncol*, *24*(1), 5-13. https://doi.org/10.1016/j.semradonc.2013.09.002

Miret, M., Cabello, M., Marchena, C., Mellor-Marsá, B., Caballero, F. F., Obradors-Tarragó, C., . . . Ayuso-Mateos, J. L. (2015). The state of the art on European well-being research within the area of mental health. *International Journal of Clinical and Health Psychology*, *15*(2), 171-179. https://doi.org/https://doi.org/10.1016/j.ijchp.2015.02.001

Mitchell, A. J., Chan, M., Bhatti, H., Halton, M., Grassi, L., Johansen, C., & Meader, N. (2011). Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. *Lancet Oncol*, *12*(2), 160-174. https://doi.org/10.1016/s1470-2045(11)70002-x

Mitchell, A. J., Ferguson, D. W., Gill, J., Paul, J., & Symonds, P. (2013). Depression and anxiety in long-term cancer survivors compared with spouses and healthy

controls: a systematic review and meta-analysis. *The Lancet Oncology*, *14*(8), 721-732. https://doi.org/https://doi.org/10.1016/S1470-2045(13)70244-4

Mitchell, A. J., Meader, N., & Symonds, P. (2010). Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: A meta-analysis. *Journal of Affective Disorders*, *126*(3), 335-348. https://doi.org/https://doi.org/10.1016/j.jad.2010.01.067

Moreno, C., Wykes, T., Galderisi, S., Nordentoft, M., Crossley, N., Jones, N., . . . Arango, C. (2020). How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry*, *7*(9), 813-824. https://doi.org/10.1016/S2215-0366(20)30307-2

Mouelhi, Y., Jouve, E., Castelli, C., & Gentile, S. (2020). How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and Quality of Life Outcomes*, *18*(1), 136. https://doi.org/10.1186/s12955-020-01344-w

Nakaya, N. (2014). Effect of psychosocial factors on cancer risk and survival. *Journal of epidemiology*, *24*(1), 1-6. https://doi.org/10.2188/jea.je20130124

Niedzwiedz, C. L., Knifton, L., Robb, K. A., Katikireddi, S. V., & Smith, D. J. (2019). Depression and anxiety among people living with and beyond cancer: a growing clinical and research priority. *BMC Cancer*, *19*(1), 943. https://doi.org/10.1186/s12885-019-6181-4

Norouzi, H., Roohi, S., Shahhosseini, M., & Nouri Ghasemabady, R. (2020). Psychometric Properties of the World Health Organization Disability Assessment Scale 2.0 among Iranian Cancer Patients. *Middle East Journal of Cancer*, *11*(3), 333-342. https://doi.org/10.30476/mejc.2019.81280.0

Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, *21*(3), 421-446. https://doi.org/https://doi.org/10.1016/S0272-7358(99)00058-6

Osborn, R. L., Demoncada, A. C., & Feuerstein, M. (2006). Psychosocial interventions for depression, anxiety, and quality of life in cancer survivors: meta-analyses. *Int J Psychiatry Med*, *36*(1), 13-34. https://doi.org/10.2190/eufn-rv1k-y3tr-fk0l

Ousmen, A., Touraine, C., Deliu, N., Cottone, F., Bonnetain, F., Efficace, F., . . . Anota, A. (2018). Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health and Quality of Life Outcomes*, *16*(1), 228. https://doi.org/10.1186/s12955-018-1055-z

Page, P. (2014). Beyond statistical significance: clinical interpretation of rehabilitation research literature. *International Journal of Sports Physical Therapy*, *9*(5), 726-736. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4197528/

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*, *46*(Pt 1), 1-18. https://doi.org/10.1348/014466506x96931

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value in Health*, *18*(1), 25-34. https://doi.org/https://doi.org/10.1016/j.jval.2014.10.005

Pfaff, H. (2003). Versorgungsforschung - Begriffsbestimmung, Gegenstand und Aufgaben. In H. Pfaff, M. Schrappe, K. W. Lauterbach, U. Engelmann, & M. Halber (Eds.), *Gesundheitsversorgung und Disease Management - Grundlagen und Anwendungen der Versorgungsforschung* (pp. 13-23). Verlag Hans Huber.

Piet, J., Würtzen, H., & Zachariae, R. (2012). The effect of mindfulness-based therapy on symptoms of anxiety and depression in adult cancer patients and survivors: a systematic review and meta-analysis. *J Consult Clin Psychol*, *80*(6), 1007-1020. https://doi.org/10.1037/a0028329

Pitman, A., Suleman, S., Hyde, N., & Hodgkiss, A. (2018). Depression and anxiety in patients with cancer. *Bmj*, *361*, k1415. https://doi.org/10.1136/bmj.k1415

Pomeroy, I. M., Tennant, A., Mills, R. J., Young, C. A., & Group, T. O. S. (2020). The WHOQOL-BREF: a modern psychometric evaluation of its internal construct validity in people with multiple sclerosis. *Quality of Life Research*, *29*(7), 1961-1972. https://doi.org/10.1007/s11136-020-02463-z

Pösl, M., Cieza, A., & Stucki, G. (2007). Psychometric properties of the WHODASII in rehabilitation patients. *Qual Life Res*, *16*(9), 1521-1531. https://doi.org/10.1007/s11136-007-9259-4

Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health? *Psychol Bull*, *131*(6), 925-971. https://doi.org/10.1037/0033-2909.131.6.925

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health and quality of life outcomes*, *1*, 27-27. https://doi.org/10.1186/1477-7525-1-27

Puhan, M. A., Frey, M., Büchi, S., & Schünemann, H. J. (2008). The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. *Health and Quality of Life Outcomes*, *6*, 46-51. https://doi.org/10.1186/1477-7525-6-46

Rehse, B., & Pukrop, R. (2003). Effects of psychosocial interventions on quality of life in adult cancer patients: meta analysis of 37 published controlled outcome studies. *Patient Educ Couns*, *50*(2), 179-186. https://doi.org/10.1016/s0738-3991(02)00149-0

Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*(2), 102-109. https://doi.org/https://doi.org/10.1016/j.jclinepi.2007.03.012

Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes [journal article]. *Health and Quality of Life Outcomes*, *4*(1), 70-74. https://doi.org/10.1186/1477-7525-4-70

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *J Pers Assess*, *98*(3), 223-237. https://doi.org/10.1080/00223891.2015.1089249

Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion* (Vol. 2). Huber.

Roth, A. J., Kornblith, A. B., Batel-Copel, L., Peabody, E., Scher, H. I., & Holland, J. C. (1998). Rapid screening for psychologic distress in men with prostate carcinoma: a pilot study. *Cancer*, *82*(10), 1904-1908. https://doi.org/10.1002/(sici)1097-0142(19980515)82:10<1904::aid-cncr13>3.0.co;2-x

Rumpold, G., Augustin, M., Zschocke, I., Strittmatter, G., & Söllner, W. (2001). Die Validität des Hornheider Fragebogens zur psychosozialen Unterstützung bei Tumorpatienten [The Validity of the Hornheide Questionnaire for Psychosocial Support in Skin Tumour Patients: A Survey in an Austrian and a German Outpatient Fopulation with Melanoma]. *Psychother Psychosom Med Psychol*, *51*(01), 25-33.

Schiavon, C. C., Marchetti, E., Gurgel, L. G., Busnello, F. M., & Reppold, C. T. (2017). Optimism and Hope in Chronic Disease: A Systematic Review [Review]. *Frontiers in Psychology*, *7*(2022). https://doi.org/10.3389/fpsyg.2016.02022

Schrank, B., Brownell, T., Tylee, A., & Slade, M. (2014). Positive psychology: an approach to supporting recovery in mental illness. *East Asian Arch Psychiatry*, *24*(3), 95-103.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5-14. https://doi.org/10.1037/0003-066X.55.1.5

Siegert, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. *Disabil Rehabil*, *32*(1), 8-17. https://doi.org/10.3109/09638280902971398

Sin, N. L. (2016). The Protective Role of Positive Well-Being in Cardiovascular Disease: Review of Current Evidence, Mechanisms, and Clinical Implications. *Current cardiology reports*, *18*(11), 106-106. https://doi.org/10.1007/s11886-016-0792-z

Sinaiko, A. D., Landrum, M. B., Meyers, D. J., Alidina, S., Maeng, D. D., Friedberg, M. W., . . . Rosenthal, M. B. (2017). Synthesis Of Research On Patient-Centered Medical Homes Brings Systematic Differences Into Relief. *Health affairs (Project Hope)*, *36*(3), 500-508. https://doi.org/10.1377/hlthaff.2016.1235

Singer, S., Danker, H., Roick, J., Einenkel, J., Briest, S., Spieker, H., . . . Kersting, A. (2017). Effects of stepped psychooncological care on referral to psychosocial services and emotional well-being in cancer patients: A cluster-randomized phase III trial. *Psychooncology*, *26*(10), 1675-1683. https://doi.org/10.1002/pon.4492

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*, *3*(2), 205-231.

Snell, D. L., Siegert, R. J., & Silverberg, N. D. (2020). Rasch analysis of the World Health Organization Disability Assessment Schedule 2.0 in a mild traumatic brain injury sample. *Brain Inj*, *34*(5), 610-618. https://doi.org/10.1080/02699052.2020.1729417

Snyder, C. R., Lopez, S. J., Edwards, L. M., & Marques, S. C. (2021). *The Oxford Handbook of Positive Psychology* (3 ed.). Oxford University Press.

Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol Serv Saude*, *26*(3), 649-659. https://doi.org/10.5123/s1679-49742017000300022 (Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade.)

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*, *166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Stecz, P., Slezáčková, A., Millová, K., & Nowakowska-Domagała, K. (2020). The Predictive Role of Positive Mental Health for Attitudes Towards Suicide and Suicide Prevention: Is the Well-Being of Students of the Helping Professions a Worthwhile Goal for Suicide Prevention? *Journal of Happiness Studies*, *21*(6), 1965-1984. https://doi.org/10.1007/s10902-019-00163-1

Streiner, D. L. (2010). Measure for measure: new developments in measurement and item response theory. *Can J Psychiatry*, *55*(3), 180-186. https://doi.org/10.1177/070674371005500310

Strittmatter, G., Mawick, R., & Tilkorn, M. (2000). Entwicklung und klinischer Einsatz von Screening-Instrumenten zur Identifikation betreuungsbedürftiger Tumorpatienten. In M. Bullinger, J. Siegrist, & U. Ravens-Sieberer (Eds.), *Lebensqualitätsforschung aus medizin- psy chologischer und – soziologischer Perspektive. Jahrbuch der Medizinischen Psychologie* (Vol. 18, pp. 59-75). Hogrefe.

Strobl, C. (2015). *Das Rasch-Modell: eine verständliche Einführung für Studium und Praxis* (Vol. 3). Hampp.

Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review*, *37*(1), 52-68.

Tatrow, K., & Montgomery, G. H. (2006). Cognitive behavioral therapy techniques for distress and pain in breast cancer patients: a meta-analysis. *J Behav Med*, *29*(1), 17-27. https://doi.org/10.1007/s10865-005-9036-1

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*, *57*(8), 1358-1362. https://doi.org/10.1002/art.23108

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological assessment*, *31*(12), 1442-1455. https://doi.org/10.1037/pas0000597

Üstün, T. B., Kostanjesek, N., Chatterji, S., & Rehm, J. (2010). Measuring health and disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0) In. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

Vehling, S., Lehmann, C., Oechsle, K., Bokemeyer, C., Krüll, A., Koch, U., & Mehnert, A. (2011). Global meaning and meaning-related life attitudes: exploring their role in predicting depression, anxiety, and demoralization in cancer patients. *Supportive Care in Cancer*, *19*(4), 513-520. https://doi.org/10.1007/s00520-010-0845-6

Vindbjerg, E., Mortensen, E. L., Makransky, G., Nielsen, T., & Carlsson, J. (2021). A rasch-based validity study of the HSCL-25. *Journal of Affective Disorders Reports*, *4*, 100096. https://doi.org/https://doi.org/10.1016/j.jadr.2021.100096

Vodermaier, A., Linden, W., & Siu, C. (2009). Screening for emotional distress in cancer patients: a systematic review of assessment instruments. *Journal of the*

National Cancer Institute, *101*(21), 1464-1488. https://doi.org/10.1093/jnci/djp336

Vodermaier, A., & Millman, R. D. (2011). Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. *Supportive Care in Cancer*, *19*(12), 1899 - 1908. https://doi.org/10.1007/s00520-011-1251-4

Walters, S. J., & Brazier, J. E. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and quality of life outcomes*, *1*, 4. https://doi.org/10.1186/1477-7525-1-4

Weis, J., Schumacher, A., Blettner, G., Determann, M., Reinert, E., Rüffer, J. U., . . . Weyland, P. (2007). Psychoonkologie. *Der Onkologe*, *13*(2), 185-194. https://doi.org/10.1007/s00761-006-1165-9

World Health Organization [WHO]. (2001). *International Classification of Functioning, Disability and Health*. World Health Organization.

World Health Organization [WHO]. (2004). *Promoting mental health: concepts, emerging evidence, practice: summary report/a report from the World Health Organization* (Vol. 29). World Health Organization.

World Health Organization [WHO]. (2019). *ICD-10 Version: 2019 - Chapter II Neoplasms (C00-D48)* Retrieved 27.05.2021 from https://icd.who.int/browse10/2019/en#/II

World Health Organization [WHO]. (2021). *Cancer: Overview*. Retrieved 28.05.2021 from https://www.who.int/health-topics/cancer#tab=tab_1

World Health Organization [WHO], & The World Bank. (2011). *World Report on Disability*. World Health Organization. https://doi.org/10.13140/RG.2.1.4993.8644

Wundt, W. (1910). *Principles of Physiological Psychology* (E. B. Titchener, Trans.). Swan Sonnenschein & Co. Lim.

Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., Acaster, S., & The Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). (2013). Methods for interpreting change over time in patient-reported outcome measures [journal article]. *Quality of Life Research*, *22*(3), 475-483. https://doi.org/10.1007/s11136-012-0175-x

Xie, F., Pickard, A. S., Krabbe, P. F., Revicki, D., Viney, R., Devlin, N., & Feeny, D. (2015). A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). *Pharmacoeconomics*, *33*(8), 867-877. https://doi.org/10.1007/s40273-015-0292-9

Zentrum für Krebsregisterdaten im Robert Koch-Institut. (2016). Bericht zum Krebsgeschehen in Deutschland 2016. In. Berlin: Robert Koch-Institut.

Zentrum für Krebsregisterdaten im Robert Koch-Institut, & Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2020). *Krebs in Deutschland für 2015/2016*. R. Koch-Institut.

Zhao, H. P., Liu, Y., Li, H. L., Ma, L., Zhang, Y. J., & Wang, J. (2013). Activity limitation and participation restrictions of breast cancer patients receiving chemotherapy: psychometric properties and validation of the Chinese version of the WHODAS 2.0. *Quality of Life Research*, *22*(4), 897-906. https://doi.org/10.1007/s11136-012-0212-9

Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of Items. *European Journal of Psychological Assessment*, *31*(4), 231-237. https://doi.org/10.1027/1015-5759/a000309