

# BMJ Open Effectiveness and cost-effectiveness of a web-based routine assessment with integrated recommendations for action for depression and anxiety (RehaCAT+): protocol for a cluster randomised controlled trial for patients with elevated depressive symptoms in rehabilitation facilities

Johannes Knauer <sup>1</sup>, Yannik Terhorst <sup>1</sup>, Paula Philippi,<sup>1</sup> Selina Kallinger,<sup>1</sup> Sandro Eiler,<sup>1</sup> Reinhold Kilian,<sup>2</sup> Tamara Waldmann,<sup>2</sup> Morten Moshagen <sup>3</sup>, Martina Bader,<sup>3</sup> Harald Baumeister<sup>1</sup>

**To cite:** Knauer J, Terhorst Y, Philippi P, *et al.* Effectiveness and cost-effectiveness of a web-based routine assessment with integrated recommendations for action for depression and anxiety (RehaCAT+): protocol for a cluster randomised controlled trial for patients with elevated depressive symptoms in rehabilitation facilities. *BMJ Open* 2022;**12**:e061259. doi:10.1136/bmjopen-2022-061259

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-061259>).

Received 20 January 2022  
Accepted 06 June 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Johannes Knauer;  
[johannes.knauer@uni-ulm.de](mailto:johannes.knauer@uni-ulm.de)

## ABSTRACT

**Introduction** The integration of a web-based computer-adaptive patient-reported outcome test (CAT) platform with persuasive design optimised features including recommendations for action into routine healthcare could provide a promising way to translate reliable diagnostic results into action. This study aims to evaluate the effectiveness and cost-effectiveness of such a platform for depression and anxiety (RehaCAT+) compared with the standard diagnostic system (RehaCAT) in cardiological and orthopaedic health clinics in routine care.

**Methods and analysis** A two-arm, pragmatic, cluster-randomised controlled trial will be conducted. Twelve participating rehabilitation clinics in Germany will be randomly assigned to a control (RehaCAT) or experimental group (RehaCAT+) in a 1:1 design. A total sample of 1848 participants will be recruited across all clinics. The primary outcome, depression severity at 12 months follow-up (T3), will be assessed using the CAT Patient-Reported Outcome Measurement Information System Emotional Distress-Depression Item set. Secondary outcomes are depression at discharge (T1) and 6 months follow-up (T2) as well as anxiety, satisfaction with participation in social roles and activities, pain impairment, fatigue, sleep, health-related quality of life, self-efficacy, physical functioning, alcohol, personality and health economic-specific general quality of life and socioeconomic cost and benefits at T1-3. User behaviour, acceptance, facilitating and hindering factors will be assessed with semistructured qualitative interviews. Additionally, a smart sensing substudy will be conducted, with daily ecological momentary assessments and passive collection of smartphone usage variables. Data analysis will follow the intention-to-treat principle with additional per-protocol analyses. Cost-effectiveness

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Large pragmatic, cluster-randomised controlled trial conducted in orthopaedic and cardiologic rehabilitation care in Germany.
- ⇒ Comprehensive observer-masked effectiveness, cost-effectiveness and feasibility analyses of a web-based computer-adaptive patient-reported outcome measure platform with action recommendations regarding depression and anxiety.
- ⇒ Fine granular disease and treatment trajectories modelling using smart sensing data.
- ⇒ Cluster randomisation and implementation of intervention or control condition on clinic level without blinding of clinical personnel.
- ⇒ Limited generalisability to other healthcare settings and countries.

analyses will be conducted from a societal perspective and the perspective of the statutory pension insurance.

**Ethics and dissemination** The study will be conducted according to the Declaration of Helsinki. The Ethics Committee of Ulm University, has approved the study (on 24 February 2021 ref. 509/20). Written informed consent will be obtained for all participants. Results will be published via peer-reviewed journals.

**Trial registration number** DRKS00027447

## INTRODUCTION

Biopsychosocial healthcare in patients with somatic diseases such as musculoskeletal or cardiovascular diseases is faced with the challenge of designing the initiation and

implementation of medical and psychosocial measures in a needs-based manner.<sup>1-4</sup> Patient-reported outcome measures (PROMs) could become important means to achieve this goal in somatic healthcare.<sup>5-8</sup> For instance, they can serve as a basis for screening mental health comorbidities, quality of life and functioning in daily living. Based on these test results, treatment planning, process and outcome measurement and quality assurance can take place.<sup>5,7</sup>

To promote acceptance and to optimise the quality of psychodiagnostics, there is a demand for an economic, resource-saving assessment that minimises the burden on patients (measurement efficiency) and facilitates evaluation and interpretation of results (usefulness).<sup>1</sup> Besides, a reliable assessment of differences between patients (measurement precision) and changes in the condition of patients during and after treatment (change sensitivity of the assessment) is a necessary prerequisite.<sup>1</sup> Paper-and-pencil testing procedures are most commonly used, but have crucial limitations, such as the limited scope or the test load as well as difficulties in collecting these measures before, during and after the treatment process.<sup>1,9-13</sup>

A computer-based implementation of PROM assessments is considered as a possible solution for an equally precise and time-efficient, user-friendly and useful assessment.<sup>1,5,14</sup> Furthermore, a likewise web-based implementation of such an assessment removes time and location dependencies. Hence, assessments of patients before and after the end of treatment could be implemented more easily in routine care settings.

However, PROMs are often static and non-adaptive to the user's responses, resulting in limited accuracy, presentation of inappropriate items for the individual, and an overall long assessment duration.<sup>15-17</sup> Computer-adaptive testing (CAT), which is based on item response theory models, is a promising option in this context to substantially reduce the burden on patients (personalised testing) and healthcare institutions (eg, immediate test evaluations).<sup>10,18-23</sup> In CAT, the items providing the maximum information about the respective patient are selected and assessed during test administration based on the previous answers of a patient.<sup>24</sup> In this way, besides a considerable reduction in test duration, an estimation accuracy that is equally good or sometimes even better compared with non-adaptive procedures can be achieved.<sup>18,23,25-27</sup>

Whereas PROMs are often used in clinical research, the usefulness and effectiveness of routine PROM assessment in somatic healthcare is still controversially discussed.<sup>8,28-33</sup> On the one hand, routine PROMs allow for real-time monitoring. Moreover, PROMs have been shown to lead to improved communication between clinician and patient, decision-making and patient satisfaction with care, and improved health outcome and detection of symptoms and mental comorbidities. On the other hand, there are barriers such as limited ability of clinicians for score interpretation, unfamiliarity with PROM software usage and less time for controlling.<sup>8,23,31-42</sup> Additionally one of the central challenges is the implementation of

further evidence-based measures on the basis of the assessment results.<sup>43</sup> Action plans directly derivable from assessment results are regarded as a prerequisite to implement PROMs beneficially.<sup>31,33,44,45</sup> However, there is still an insufficient linkage between assessment results and implementation of existing evidence-based guidelines and recommendations for action in everyday clinical practice.<sup>33,46</sup>

One way to promote the desired probability of action following diagnostic results could be persuasive design components, such as reminder features that are automatically triggered depending on the test environment.<sup>47-49</sup> Persuasive designed technological approaches are defined as interactive systems that purposefully influence the user, aiming to change behaviour or attitudes.<sup>50</sup> The provision of computer-based databases and concrete decision-making aids and recommendations for action is seen as one way of reducing these existing barriers.<sup>51,52</sup> In this context, it could be useful to link the individual test results with therapy standards as well as recommendations for action and guideline knowledge. These have been formulated in particular for the areas of comorbid depression and anxiety in patient populations with somatic diseases.<sup>53-57</sup> Such a combination could offer the practitioner (1) background knowledge, (2) recommendations for action as well as (3) documentation aids. Ideally, such elements should be directly integrated into testing systems (eg, web based and CAT based) to provide a comprehensive platform from screening to action.

Hence, the aim of the present trial is to examine a persuasive design optimised CAT system (RehaCAT+) providing background knowledge, recommendations for action as well as documentation aids against a standard CAT system (RehaCAT). This will be exemplified with a focus on depression as the primary outcome and anxiety as major mental health comorbidities in cardiological and orthopaedic care. The following research questions will be addressed:

1. Does RehaCAT +improve rehabilitation patients' depression after 1 year (T3)?
2. Does RehaCAT +improve depression, anxiety, satisfaction with participation in social roles and activities, pain impairment, fatigue, sleep, health-related quality of life, self-efficacy, physical function, and alcohol use at discharge (T1) and 6 months follow-up (T2) as well as 1 year later regarding all the secondary outcomes (T3)?
3. Does RehaCAT +lead to improved documentation and improved follow-up and postrehabilitation recommendations?
4. Does RehaCAT +lead to improved utilisation of rehabilitation therapy standard and guideline compliant healthcare services during and after rehabilitation?
5. What is the cost-effectiveness of RehaCAT +compared with to RehaCAT?
6. What is the acceptance and feasibility of RehaCAT?
7. What are facilitators, hindering factors, mediators and potential risks associated with RehaCAT+?

A web-based CAT system provides a powerful way to assess PROM, however, it is still subject to limitations: (1) it requires active input of the patient—even if reduced through CAT, (2) the assessment is limited to fixed time points, which may lead to long unassessed time intervals in which significant symptom change may occur and (3) due to the nature of self-report the answers by patients may be biased (eg, social desirability or recall bias).<sup>58–61</sup>

One solution to this could be the addition of ecological momentary assessment and smart sensing to allow for digital phenotyping.<sup>62–63</sup> Digital phenotyping is defined as the moment-by-moment quantification of the individual health in situ through digital variables and data generated by personal devices (eg, smartphone or smartwatch).<sup>62–63</sup> First studies show promising results highlighting the potential of this method to complement PROM assessments for monitoring and predicting symptoms with minimal added patient burden.<sup>64–70</sup> In future the combination of high quality PROM at fixed timepoints combined with continuous monitoring through smart sensing and information from the clinical information system could become a promising data base, which could be used to (1) predict symptom trajectories, (2) build early-detection of adverse events systems (RED-flag) or (3) personalised treatment recommendation systems.<sup>71–74</sup>

Hence, this study additionally investigates the extent to which smart sensing is suitable for assessing mental health in a routine care setting. In the context of this exploratory study, we will focus on the following research questions:

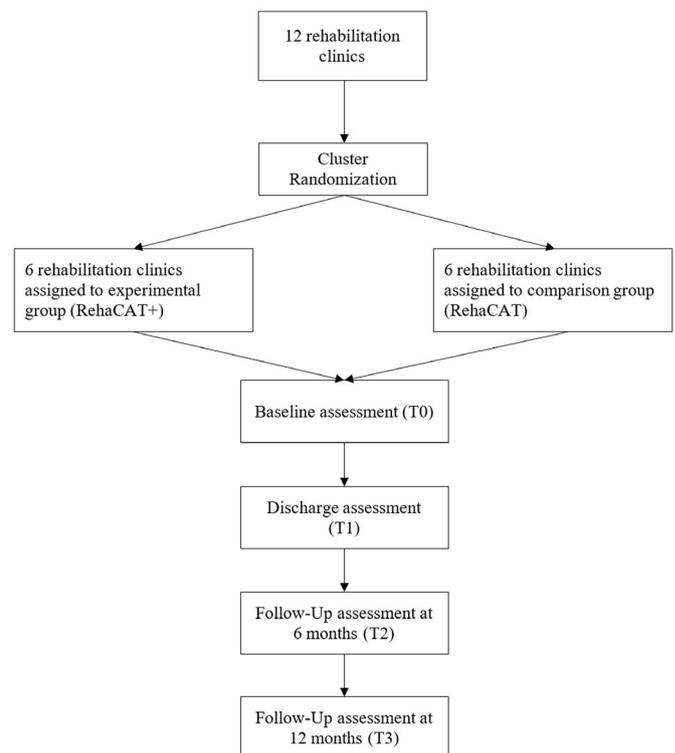
8. What are the associations between digital markers and health-related variables?
9. Are digital markers suitable for predicting health-related variables and disease or disorder status?
10. What is patient acceptance, adherence and perceived usefulness of smart sensing?

## METHODS AND ANALYSIS

### Study design

A two-arm, pragmatic, cluster-randomised controlled trial (cRCT) will be conducted, comparing the experimental group receiving an enhanced version of a PROM system called ‘RehaCAT+’ to the control group receiving the basic version of the PROM system called ‘RehaCAT’ in a 1:1 design (figure 1). See below for detailed description of the experimental and control group.

This cRCT has been approved by the ethics committee of Ulm University (509/20-FSt/Sta) and will be reported in accordance with the Consolidated Standards of Reporting Trials Statement 2010 and the extensions for reporting pragmatic trials and cluster randomised trials.<sup>75–77</sup> Cost-effectiveness analyses will be reported following the Consolidated Health Economic Evaluation Reporting Standards statement<sup>78</sup> and the guidelines from the International Society for Pharmacoeconomics and Outcomes Research.<sup>79</sup> This trial protocol was created according to Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines.<sup>80</sup> The



**Figure 1** Flow chart.

expected timeline for trial completion is September 2024 with first patient enrolment in July 2022.

### Procedure and recruitment

The testing systems RehaCAT (control group) and RehaCAT+ (experimental group) will be implemented in the clinic routine of 12 clinics offering cardiological or orthopaedic rehabilitation treatment in Germany. Included clinics pursue medical and occupational oriented stationary rehabilitation according to German ICF diagnosis-based rehabilitation guidelines.<sup>81</sup> With a psychosocial approach rehabilitation is focused on patients’ impairments (eg, body functions and structure), restoration of activities and removing restrictions of participation.<sup>82</sup> Accordingly, the treatment in clinics often contains diagnostics, pharmacotherapy, physiotherapy and psychotherapy. Standard stationary stay usually lasts for 3 weeks. The treatment as well as the duration of treatment is expected to vary across patients and between clinics. Treatment will be further described post hoc using the results from the cost-effectiveness questionnaires (see Assessments). Neither the control condition nor the experimental condition will interfere with clinical treatment (see Conditions).

For the study, one of two versions of a web-based computer-adaptive diagnostic platform will be implemented within the clinics (see Conditions). Clinic personnel will be trained in an on-site workshop during the implementation phase. The training will cover technical functions of the platform (eg, how new patients can be registered, how patients’ results can be received) as well as recommendations and guidelines for clinical practice

(eg, how results should be interpreted, information about national treatment guidelines for mental health). Lastly, clinicians will also be trained in the communication with patients and procedures for patients. After the training, written manuals providing a summary of the workshop will be available in the system for the clinic personnel. Qualification level of clinic personnel operating the system will vary across clinics (eg, nurses, medical doctors, clinical psychologists). This will be monitored and reported (see Usage behavior, acceptance, facilitating and hindering factors). Furthermore, the technical administrator has direct contact options (eg, email) to the research team. The platform is designed so patients can go through the respective version of the testing system to deliver patient-reported outcomes at various points in time. A subset of patients in routine care fulfilling the inclusion criteria (see Inclusion and exclusion criteria) will be included in this study. Study participants will receive all questionnaires from routine care and additional research questionnaires. Routine patients will go through the diagnostic measures at admission (T0) and before discharge (T1), as well as at 6 months (T2) follow-up as part of their clinical routine. Study participants will additionally be assessed at 12 months (T3) follow-up.

Data collection will be digital. Due to the web-based character of the platform, inpatient and outpatient assessments are possible. Clinics are free to implement the admission and discharge assessments as inpatient or outpatient assessments. Data for follow-up will be assessed solely in an outpatient setting. Assessment procedures (eg, inpatient or outpatient assessment at admission) are expected to vary across clinics and will be further described post-hoc. For an explanatory illustration of the assessment procedures, see [figure 2](#).

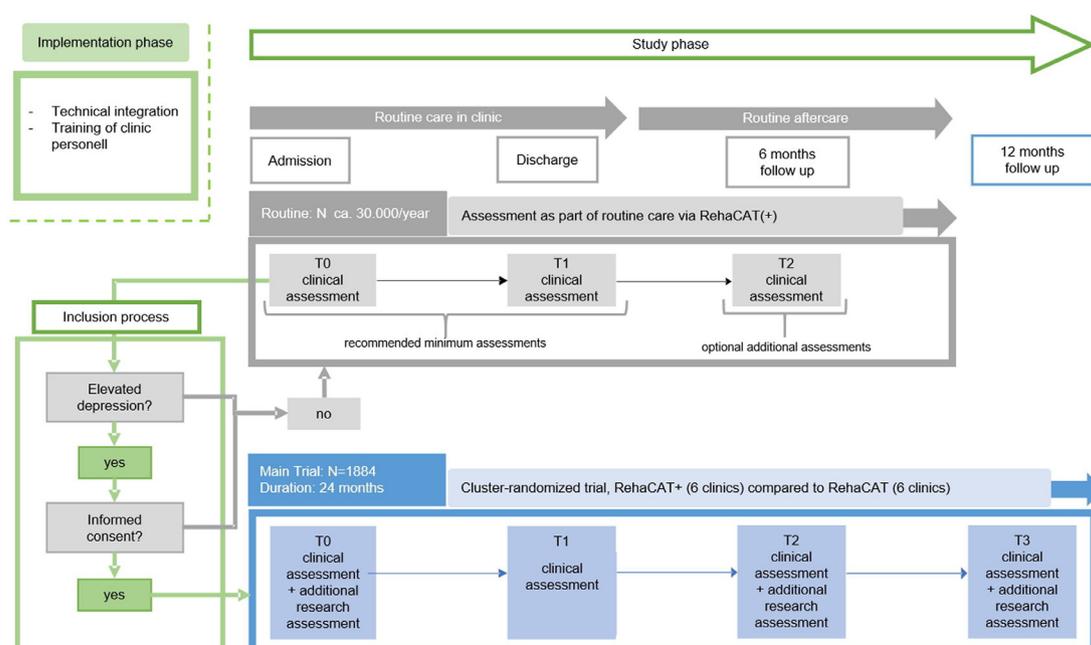
### Inclusion and exclusion criteria

Since this is a cRCT, inclusion and exclusion are differentiated at two levels a) cluster level (ie, eligibility of clinics) and b) patient level: To be eligible, rehabilitation clinics must be located in Germany, provide cardiological or orthopaedic rehabilitation and sign a cooperation agreement with Ulm University. There are no further exclusion criteria for clinics. Within each cluster (ie, rehabilitation clinic) patients who exhibit elevated depression scores (Patient-Reported Outcome Measurement Information System, PROMIS Emotional Distress Depression: T-value  $\geq 65.2$ )<sup>15</sup> at the initial assessment will be informed about the study and consecutively asked for their participation consent (online supplemental material: SPIRIT Supplement Informed Consent). To be eligible, patients with elevated depression scores must (1) be 18 years or older, (2) have sufficient German language skills, (3) provide an email address, (4) agree to the data privacy and processing procedures according to the European General Data Protection Regulation and (5) sign the informed consent. There are no further exclusion criteria for patients.

### Randomisation, allocation and masking

Randomisation and allocation regarding the control (RehaCAT) and experimental group (RehaCAT+) of the 12 participating rehabilitation clinics will be performed by an independent researcher to avoid selection bias. Randomisation will be done on cluster level.

Researchers responsible for randomisation will be obscured to the rehabilitation clinic names and agencies. Randomisation will be done using an automatically created randomisation list.



**Figure 2** Procedure.

For the outcome analyses, the conducting analyst will be obscured to group allocation. Patients will remain obscured to their study arm assignment. Neither the clinics (clinic personnel) nor the research team will be obscured to assigned study condition.

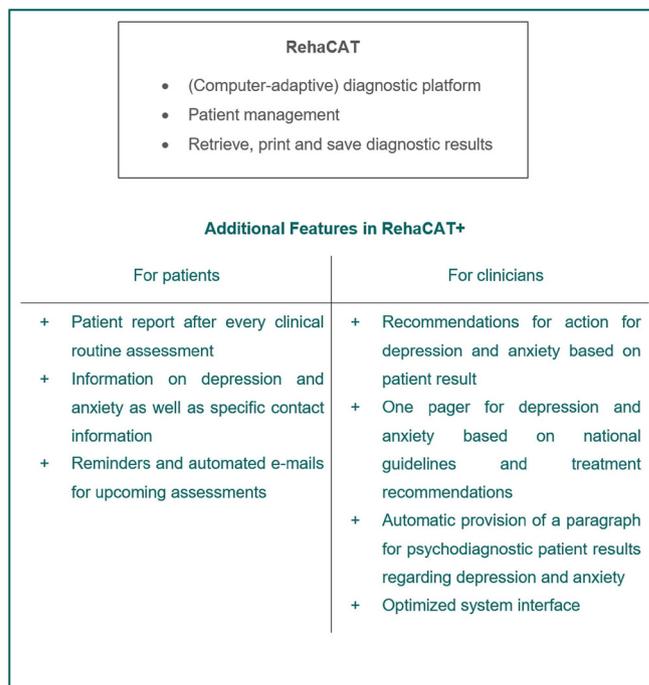
### Conditions

**RehaCAT-Control Group:** RehaCAT is a server-based and web-based, device-independent test system, which allows the use of classical test procedures as well as computer-adaptive procedures. RehaCAT does not interfere with the (clinic specific) standard treatment and patients have unrestricted access to treatment as usual. RehaCAT is divided into four user areas: (1) patient, (2) staff, (3) administrator, (4) researcher. The platform allows system administrators to upload and manage patients. Patients go through the diagnostic measures. Clinicians can view the test results of their patients immediately after completion of each assessment point (T0, T1 and T2). Test results consist of a traffic light feedback (green=normal severity, yellow=elevated severity based on clinical cut-off values, red=high severity 2.5 SDs above mean<sup>15 83</sup>), patients' test results expressed in T-values combined with clinical cut-off values, and a line graph visualising the results and change over assessment times. For a full overview of the assessment see 2.5.1 and 2.5.2.

**RehaCAT+Experimental Group:** In addition to the structure and features of RehaCAT, RehaCAT +follows a persuasive design optimised technology (eg, motivation, ability and automatic trigger considering test environment)<sup>47 48</sup> to increase the desired probability of action. RehaCAT +offers additional<sup>1</sup> system features (automated email reminders for patients),<sup>2</sup> clinician features (stored recommendations for action for depression and anxiety based on respective patient results, call to action plans, individualised documentation aids and supporting information material for depression and anxiety) and<sup>3</sup> patient features (individual symptomatic information at discharge and T2/3, possible points of contact/help).

The urgency of the recommendation for action (ie, need for in-depth psychodiagnostics) varies depending on screening severity. Additionally, material on handling of psychological burden can be accessed. The material is based on: (1) the rehabilitation therapy standards and framework concepts,<sup>84–87</sup> (2) the practice recommendations for orthopaedic and cardiological rehabilitation,<sup>54 55</sup> (3) the recommendations for psychodiagnostics in somatic rehabilitation<sup>53</sup> and (4) the national S3 guidelines for depression<sup>56</sup> and anxiety.<sup>57</sup> A summary of the two conditions is provided in figure 3.

The clinics will be compensated with €100 per recruited patient for the resulting hospital expenses in the context of participant recruitment, data collection, study documentation as well as provision of the discharge reports. Study patients will receive an expense allowance of €20 each for their participation in the T2 and T3 measurements.



**Figure 3** Features.

RehaCAT(+) is developed as an open-source platform. It is currently in the certification process according to the medical device regulation (MDR). The platform is developed according to the requirements of the German Medical Devices Act and the MDR. Hence, the software development and validation process is taking the IEC 62304 (safety class B), the GAMP5 (category 4), the general principles of software validation of the Food and Drug Administration (FDA) as well as the Pharmaceutical Inspection Cooperation Scheme 11–3 into account. Furthermore, technical requirements and standards for the interoperability between different medical devices (eg, HL7 FHIR) are under development. The certification process of the platform is planned to be completed in 2022.

### Sample size and study power

The sample size calculation is based on the primary outcome, depression severity 12 months after the end of rehabilitation. It is assumed that the experimental group is superior to the control group in the way that patients will show fewer depressive symptoms at 12 months follow-up. In view of the additive study design testing for incremental benefit of RehaCAT +over RehaCAT and the distal outcome, a small additional effect of  $d=0.24$  compared with the standard condition is regarded as clinically significant following the recommendation of Cuijpers *et al.*<sup>88</sup> With  $2 \times 6$  cluster-randomised rehabilitation clinics, each clinic requires a sample of 110 ( $SD=25$ ) participating rehabilitants with elevated depression scores to achieve a test power of 80% given an alpha error (two-sided) of 0.05, an estimated ICC of 0.02,<sup>89 90</sup> and an assumed correlation with baseline depression scores of 0.50. With an estimated drop-out rate (rehabilitation

**Table 1** Assessments

Variable	Instrument	CAT	Time of measurement			
			T0	T1	T2	T3
Depression	PROMIS emotional distress—depression	✓	✓	✓	✓	✓
Anxiety	PROMIS emotional distress—anxiety	✓	✓	✓	✓	✓
Satisfaction with participation in social roles and activities	PROMIS satisfaction with social roles and activities	✓	✓	✓	✓	✓
Pain impairment	PROMIS pain interference	✓	✓	✓	✓	✓
Fatigue	PROMIS fatigue	✓	✓	✓	✓	✓
Sleep	PROMIS sleep disturbance	✓	✓	✓	✓	✓
Health-related quality of life	PROMIS global health		✓	✓	✓	✓
Self-efficacy	PROMIS self-efficacy general		✓	✓	✓	✓
Physical Function	PROMIS physical function	✓	✓	✓	✓	✓
Alcohol use	AUDIT-10		✓	✓	✓	✓
Personality	BFI-10		✓			
Generic quality of life	EQ5D-5L		✓		✓	✓
Health and social services use and costs	CSSRI		✓		✓	✓
Medical record data	Provided by clinicians (eg, discharge reports)			✓		

AUDIT-10, Alcohol Use Disorders Identification Test; BFI-10, 10 item Big Five Inventory; CAT, Computer-Adaptive Patient Reported Outcome Test; CSSRI, Client Sociodemographic and Service Receipt Inventory; EQ5D-5L, European Quality of Life 5 Dimension - 5 Level Questionnaire; PROMIS, Patient Reported Outcome Measurement Information System.;

start-end) of 20%<sup>91</sup> and the assumption of a doubling drop-out rate by T3, a total sample of N=1848 rehabilitants is required.

### Assessments

Quantitative outcome assessment will be performed at baseline/beginning of rehabilitation (T0), at discharge/end of rehabilitation (T1), and at 6 and 12 months (T2, T3). Primary outcome is the depression severity score twelve months after rehabilitation (T3). An overview is provided in [table 1](#).

### Primary outcome

Depression severity will be assessed with the computer-adaptive PROMIS Emotional Distress-Depression Item Set including an item bank with 28 items that capture negative mood, decrease in positive emotions, cognitive deficits, as well as negative self-image and negative social cognition.<sup>92</sup> All items are rated on a five-point response scale asking respondents to rate the frequency of their symptoms in the past 7 days (never, rarely, often, sometimes, always). A Cronbach's alpha of 0.99 was found for the internal consistency of the item set.<sup>93</sup>

### Secondary outcomes

Anxiety will be assessed with the computer-adaptive PROMIS Emotional Distress-Anxiety Item Set including 29 items that assess emotional distress caused by hypersensitivity, anxiety and stress, as well as associated somatic

symptoms.<sup>92</sup> All items are rated on a five-point response scale asking respondents to rate the frequency of their symptoms in the past 7 days (never, rarely, often, sometimes, always). The internal consistency of the item set was found to be good.<sup>92</sup>

The computer-adaptive PROMIS Satisfaction with Participation in Social Roles and Activities Item Set comprising 14 items will be used to assess the perceived ability to perform usual social roles and participate in social activities. All items are phrased in terms of perceived limitations and answered using a five-point response scale. Reliability was estimated to be  $\alpha > 0.90$ .<sup>94</sup>

Pain impairment will be assessed with the computer-adaptive PROMIS Pain Interference Item Set including 40 items that capture self-assessment of the consequences of pain in one's life. This includes the extent to which pain interferes with engagement in social, cognitive, emotional, physical, as well as leisure activities.<sup>95</sup> The items refer to the past 7 days and are rated on three different five-point Likert scales. The internal consistency of the item set was found to be good.<sup>95</sup>

Fatigue will be assessed with the computer-adaptive PROMIS Fatigue Scale comprising 95 items and measures both fatigue experience and the impact of fatigue on daily life and functionality. The intensity, frequency and duration of fatigue were graded on a five-point response scale. Reliability was estimated to be  $\alpha > 0.90$ .<sup>96</sup>

Sleep will be assessed with the computer-adaptive PROMIS Sleep Dimension that measures subjective sleep quality and quantity and sleep-related impairment in daily functioning. The scale comprises 27 items and is rated on a response scale from 1 (not at all or never) to 5 (very much or always). It is a validated instrument and has good psychometric properties with  $\alpha > 0.90$ .<sup>97 98</sup>

Health-related quality of life will be assessed with the PROMIS scale on global health aspects (Global Health) is used. The scale includes 10 items that capture global physical health (physical health, physical functioning, fatigue, pain), and global mental health (general quality of life, mental health, satisfaction with social activities and relationships, and emotional distress).<sup>99</sup> Nine items are scored on a response scale of 1–5, and the item assessing pain is scored from 0 to 10. Internal consistency was estimated to be good with  $\alpha > 0.82$ .<sup>99</sup>

Self-efficacy will be assessed with the short scale of the PROMIS General Self-Efficacy Scale which contains four items. It can be used to assess how much confidence one has in one's own abilities to master tasks and situations. Items are rated on a response scale from 1 (I am not at all confident) to 5 (I am very confident). Internal consistency was estimated to be high ( $\alpha = 0.96$ ).<sup>100 101</sup>

Physical function will be assessed using the computer-adaptive PROMIS Physical Function Item Set that measures the ability to perform daily life activities that require physical activity such as walking or lifting objects. It comprises 164 items that are assessed on a 5-point scale ranging from 1 (Without any difficulty) to 5 (Unable to do). Internal consistency was found to be very good with  $\alpha > 0.88$ .<sup>102</sup>

Alcohol use will be assessed with the Alcohol Use Disorders Identification Test which is a 10-item questionnaire developed by the WHO as a screening tool for hazardous and harmful alcohol use. Responses to each question are scored from 0 to 4, allowing a maximum of 40 points. Reliability has been investigated in some studies and is considered good with a median of  $\alpha = 0.80$ .<sup>103</sup>

Discharge reports will be analysed regarding (1) frequency of documented screening results, (2) therapy standard and guideline appropriate therapeutic services (documented services/therapy standard recommendations as a function of depression and anxiety results), (3) therapy standard and guideline appropriate follow-up and postrehabilitation recommendations (documented recommendations/therapy standard/guideline recommendations as a function of depression and anxiety results).

### Moderators

As potential moderators, sociodemographic data (age, gender, nationality) and personality will be recorded at admission (T0). Personality will be assessed with the BFI-10 (Big Five Inventory-10), a short version of the BFI that has good psychometric properties and a retest-reliability of  $\alpha = 0.73$ .<sup>104</sup> Additionally, data from medical records will be used as moderators (eg, indication area

orthopaedic or cardiologic, chronic conditions, rehabilitation duration).

### Health economics

Generic quality of life will be assessed with the European Quality of Life 5 Dimension-5 Level Questionnaire from the EuroQol foundation ([www.euroqol.org](http://www.euroqol.org)).<sup>105</sup> The five dimensions surveyed are mobility, self-care, general activities, pain/physical discomfort and anxiety/dejection. Each of the five areas is ranked on a five-point scale. Based on the answers, the respective health status is recorded.<sup>106</sup>

Health and social services use and costs will be assessed with the Client Sociodemographic and Service Receipt Inventory which is a standardised but adaptable inventory. Five domains are queried, including sociodemographic information, usual living situation, income and employment status, use of mental health services, and medication treatment.<sup>107</sup>

### Usage behaviour, acceptance, facilitating and hindering factors

Questions about usage behaviour, potential risks of the platforms, as well as barriers and facilitators to implementation, will be elicited based on qualitative semistructured interviews conducted with both patients and clinic staff centrally involved in the implementation of RehaCAT and RehaCAT+. The semistructured interviews will be conducted with the help of an interview guide based on existing instruments of previous studies.<sup>53 108</sup>

### Smart sensing substudy

Smart sensing data will be collected actively and passively via the mobile application AWARE for smartphones. The AWARE Framework is an open-source framework that allows to passively collect digital behavioural data via an app as well as to send ecological momentary assessments (eg, short questions: 'how are you feeling right now?') to the app user for answering.<sup>109</sup> The AWARE framework has been tested in previous studies<sup>64 68 109 110</sup> without technical, privacy or ethical issues. All collected data will be stored pseudonymised and personal data (eg, contact numbers) will be anonymised using the secure hash algorithm 256 (SHA-256).<sup>109</sup>

After completing the diagnostic measures at T0, T1 and T2, all patients will be informed in the RehaCAT(+) system about the optional mobile sensing substudy. If interested, they can provide an email address to receive further information on the study and a study invitation. This is independent from study participation in the cRCT. Therefore, both routine care patients and patients partaking in the cRCT will be able to participate. Participants who provide their informed consent will be instructed to instal the research application on their personal smartphones. After installation participants will be able to choose which data points will be collected over the next 6 months.

### Active assessment

Gender, age and personality with the BFI-10 will be assessed<sup>104</sup> once after installing the application.

Furthermore, acceptance of and satisfaction with smart sensing will be measured using the Unified Theory of Acceptance and Use of Technology questionnaire,<sup>111 112</sup> satisfaction with the research application will be measured with the User Version of the Mobile Application Rating Scale.<sup>113</sup> Both questionnaires will be assessed once after 6 months before deinstalling the application.

The following clinical questionnaires will be assessed every 2 weeks: Depression (dimensional and categorical) with the PHQ-8 (if PHQ-2 score >2)<sup>114–116</sup>; anxiety with the GAD-7<sup>117</sup>; stress with the PSS-10<sup>118</sup>; sleep with the ISI-7<sup>119</sup>; loneliness with the UCLA three-item version.<sup>120</sup>

Every morning, participants will be asked short questions about mood (valence), drive (arousal), control, unpredictability, stress and sleep, at midday about mood (valence), drive (arousal), control, unpredictability and stress, and in the evening, participants are again asked about mood (valence), drive (arousal), control, unpredictability, stress and activity during the day. This assessment is based on previous studies.<sup>65 68 121 122</sup>

### Passive outcomes

The research app allows to track a broad range of sensors (accelerometer, application usage, barometer, battery, Bluetooth, communication, gravity, gyroscope, light, locations, magnetometer, network, proximity, rotation, screen sensor). However, each user will be able to freely decide which sensors are activated and access permissions can always be activated and deactivated without giving reasons. In addition, sensible location data (eg, Global Positioning System (GPS) coordinates) will be obscured, so pseudonymisation can be upheld all the time.

The following digital markers can be collected (depending on permissions of user): frequency and duration of smartphone and individual app usage, frequency and duration/length of calls and text messages, randomly distorted GPS, and type of movement.

### Data management and data sharing plan

Data collection will be completed online using the server-based system RehaCAT(+) and the research application in pseudonymised form. Retrieved data will be stored encrypted by responsible employees. All data will be anonymised after completion of the trial. Furthermore, an independent data safety and monitoring board (DSMB) with long-standing experience in clinical trials has been established. The function of the DSMB is to monitor the course of the study and, if necessary, give recommendations to the steering committee for discontinuation, modification or continuation of the study.

Individual participant data will be made available on request after de-identification beginning 12 months following article publication of the effectiveness paper. Data will be made available to researchers who provide a methodologically sound proposal, not already covered by others. Proposals should be directed to HB. Data requestors will need to sign a data access agreement.

Provision of data is subject to data security regulations. Investigator support depends on available resources.

### Measures to reduce methodological sources of error

**Selection bias:** Randomisation and allocation regarding the group allocation (RehaCAT/ RehaCAT+) of the participating rehabilitation clinics will be done by an independent researcher. **Performance bias:** Rehabilitation staff centrally involved in the implementation will be trained along training materials, as well as continuously supervised regarding the training materials. RehaCAT(+) and its application will be described in detail in a test manual. Deviations from the test manual will be recorded and formatively reduced during the implementation process of RehaCAT(+) in the individual clinics. **Contamination bias:** Cluster randomisation is used to avoid study arm contamination. **Detection bias:** rating procedures (analysis of discharge reports) are performed by independent raters who are obscured to the study arm affiliation of the clinics and thus the patients. Patients also remain obscured to their study arm assignment, as the differentiation between RehaCAT and RehaCAT(+) is not obvious to them. **Obscuring** will not be realised for clinic personnel only. **Reporting bias:** A detailed definition of all methodological aspects of the present clinical study is provided in this study protocol, submitted for publication prior to randomisation start. **Evaluating representativeness:** To assess the representativeness of the results, quantitative and qualitative analyses will be performed regarding the reasons that may lead to a non-existent 100% exhaustion in the routine assessment.

### Statistical analyses

#### Clinical analyses

Study data will be centrally processed and analysed by an independent researcher. Missing values and missingness patterns will be explored and analysis will be adjusted accordingly using multiple imputation strategies (based on heteroscedastic two-level linear models considering the metric of outcome). The analysis will follow the intention-to-treat principle. In addition, per-protocol analyses will be conducted. The primary outcome as well as all other continuous outcomes will be analysed based on Hierarchical Linear Models considering cluster structure and baseline values. Binary outcomes will be analysed using mixed logistic regression models. Moderator and mediator analyses will be performed to determine differential effects with respect to key sociodemographic and medical variables.

The effect of study participation will also be measured (participation rate at T1-3) in order to be able to make statements about the transferability of the results from the present randomised study to routine care without research support.

#### Health economic evaluation

In the health economic evaluation, an incremental cost–utility analysis will be performed from the societal

perspective, as well as from the perspective of the German statutory pension insurance (SPI) according to the net benefit approach.<sup>78 123</sup> The necessary maximum willingness to pay (MWTP) for a clinical improvement of depressive symptoms by 50% (=response) and for the gain of a quality-adjusted life-year will be determined. The estimation of the stochastic uncertainty will be done by means of nonparametric bootstrapping, the interpretation of the results is based on cost-effectiveness acceptance curves.<sup>124</sup> These provide information on how high the MWTP must be to be judged cost-effective with a probability of 95%, or with what probability a pre-determined MWTP is judged to be cost-effective.<sup>123</sup> Following international guidelines, a value range of the MWTP between €0 and €1 250 000 is chosen.<sup>125 126</sup> The analysis of the health economic relevance of moderator and mediator variables will be performed by means of net benefit regression models for net benefit ratios between €0 and €1 250 000.<sup>127-130</sup> The analysis from the macroeconomic perspective will consider all direct and indirect disease costs,<sup>131</sup> the analysis from the perspective of the SPI will take the disease costs to be borne by the SPI (eg, for medical rehabilitation services) as well as the costs for the testing platform to be borne by the SPI into account.

#### Qualitative data analysis

Qualitative interviews of patients and clinic staff will be conducted and analysed. The analysis of qualitative data will be based on qualitative content analysis. An inductive-deductive approach will be applied along the theory-based interview guide. Reliability of results will be established (indicated by intercoder agreement) with two independent raters coding all transcripts on the basis of coding guide and rules. This coding guide will be developed in an iterative process with consensus finding.

#### Smart sensing

##### Identification of prognostic factors

Correlation analysis and multilevel regression models will be used to identify bivariate relationships between PROMs, digital markers, disease-free survival and postoperative complications. The presence of missing values and missingness patterns will be investigated and analysis will be adjusted accordingly (eg, multiple imputation).<sup>132</sup>

##### Prediction modelling

Significance test-based methods as well as machine learning models will be used. In case of nested data structure (eg, development of quality of life over time with repeated measures) multilevel regression models will also be used for prediction.<sup>133-135</sup> For continuous outcomes (eg, depression severity) linear models will be used, while logistic models will be applied for dichotomous outcomes (eg, depression state yes or no).

In addition to these significance test-based methods, machine learning models will be used. Machine learning is an iterative process and several modelling approaches will be tested (eg, K-Nearest Neighbour algorithm<sup>136</sup> or

gradient-boosted trees<sup>137 138</sup>). However, since the field is rapidly developing, we cannot a priori define the exact approaches that will be used. Hyperparameter optimisation will be conducted using grid-search.

#### Patient and public involvement

Patient and public involvement (PPI) representatives have provided input to this study in several stages. Results of previous projects including patient feedback, were used to further develop and optimise study design and procedures. PPI representatives (eg, as members of an advisory board) are included to improve usability, design and comprehensibility but have no influence on the outcomes, data analysis methods or study design.

#### ETHICS AND DISSEMINATION

Ethical approval has been obtained from Ulm University on 24 February 2021 (ref. 509/20). The study is conducted according to the Declaration of Helsinki. Informed consent will be obtained from all participants.

Results will be published in peer-reviewed journals. They will also be made known through local conferences and research seminars, national and international scientific congresses, and through direct and indirect contacts with clinicians, public health managers and other health-care professionals.

#### Author affiliations

<sup>1</sup>Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

<sup>2</sup>Department of Psychiatry and Psychotherapy II, Ulm University, Ulm, Germany

<sup>3</sup>Department of Psychological Research Methods, Ulm University, Ulm, Germany

**Acknowledgements** Authors would like to thank the project members Carolin von Gottberg, Minh Duc Duong, Erce Rodopman, Tobias Sterns, Aydin Spieler, Rebecca Knoll and Robin Kraft for their technical support and development of RehaCAT.

**Contributors** HB is principle investigator of RehaCAT+. HB, RK and MM obtained funding for this study. HB, JK, YT, PP, SK, MM, MB, RK and TW contributed to the study design. HB, SE, JK, YT, PP and SK developed the platform RehaCAT(+). MM and RK contributed to the design of the effectiveness and health economic evaluation. JK drafted the manuscript. All authors contributed to the article and approved the submitted version.

**Funding** This work was supported by the Federal Ministry of Education and Research (BMBF, 01GX1901).

**Disclaimer** BMBF had no role in study design, decision to publish or preparation of this manuscript. BMBF will not be involved in data collection, analyses, decision to publish or preparation of future papers regarding this study.

**Competing interests** Authors of the manuscript were partly involved in the development of RehaCAT(+). HB has been the beneficiary of study support (third party funding) from several public funding organisations in the context of research on computer-adaptive testing and patient-reported outcome systems.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Consent obtained directly from patient(s)

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content

includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Johannes Knauer <http://orcid.org/0000-0002-7679-7090>

Yannik Terhorst <http://orcid.org/0000-0003-4091-5048>

Morten Moshagen <http://orcid.org/0000-0002-2929-7288>

#### REFERENCES

- Baumeister H, Lin J, Ebert DD. [Internet- and mobile-based approaches: Psycho-social diagnostics and treatment in medical rehabilitation]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2017;60:436–44.
- Haywood KL. Patient-reported outcome I: measuring what matters in musculoskeletal care. *Musculoskeletal Care* 2006;4:94 doi:10.1002/msc.94
- Kelkar AA, Spertus J, Pang P, et al. Utility of patient-reported outcome instruments in heart failure. *JACC Heart Fail* 2016;4:165–75.
- Algurén B, Coenen M, Malm D, et al. A scoping review and mapping exercise comparing the content of patient-reported outcome measures (PROMs) across heart disease-specific scales. *J Patient Rep Outcomes* 2020;4:7.
- Calvert M, Kyte D, Price G, et al. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ* 2019;364:k5267. doi:10.1136/bmj.k5267
- Klag MJ, Mackenzie EJ, Carswell CI, et al. The role of the patient in promoting patient-centered outcomes research. *Patient* 2008;1:1–3. doi:10.2165/01312067-200801010-00001
- Nelson EC, Eftimovska E, Lind C, et al. Patient reported outcome measures in practice. *BMJ* 2015;350:g7818. doi:10.1136/bmj.g7818
- Field J, Holmes MM, Newell D. Proms data: can it be used to make decisions for individual patients? A narrative review. *Patient Relat Outcome Meas* 2019;10:233–41.
- Gibbons RD, Weiss DJ, Kupfer DJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv* 2008;59:361–8.
- Gamper EM, Martini C, Petersen MA. Do patients consider computer-adaptive measures more appropriate than static questionnaires? *J Patient-Reported Outcomes* 2019;3.
- Forkmann T, Boecker M, Wirtz M, et al. Validation of the Rasch-based depression screening in a large scale German general population sample. *Health Qual Life Outcomes* 2010;8:105–12.
- Abberger B, Haschke A, Tully PJ, et al. Development and validation of parallel short forms PaSA-cardio for the assessment of general anxiety in cardiovascular rehabilitation patients using Rasch analysis. *Clin Rehabil* 2017;31:104–14.
- Sharpe JP, Gilbert DG, et al. Effects of repeated administration of the Beck depression inventory and other measures of negative mood states. *Pers Individ Dif* 1998;24:457–63.
- Sehlen S, Ott M, Marten-Mittag B. Machbarkeit und Akzeptanz Computer-gestützter Indikationsdiagnostik (CgID) Zur Identifizierung psychosozial belasteter Patienten Im klinischen Alltag. *PPmP - Psychother Psychosom Medizinische Psychol* 2012;62:276–83.
- Wahl I, Löwe B, Bjorner JB, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol* 2014;67:19 doi:10.1016/j.jclinepi.2013.04.019
- Embretson S, Yang X. Item response theory. In: Green JL, Camilli G, Elmore PB, eds. *Handbook of complementary methods in education research*. Lawrence Erlbaum Associates Publishers, 2006: 385–409.
- Rose M, Wahl I, Löwe B. Computer adaptive tests in Der Medizin. *Psychother Psychosom* 2013;63:48–54.
- Abberger B, Haschke A, Wirtz M, et al. Development and evaluation of a computer adaptive test to assess anxiety in cardiovascular rehabilitation patients. *Arch Phys Med Rehabil* 2013;94:2433–9.
- Forkmann T, Kroehne U, Wirtz M, et al. Adaptive screening for depression — recalibration of an item bank for the assessment of depression in persons with mental and somatic diseases and evaluation in a simulated computer-adaptive test environment. *J Psychosom Res* 2013;75:437–43.
- Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res* 2014;23:1609–18.
- Papuga MO, Dasilva C, McIntyre A, et al. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Syst* 2018;7:1–12.
- Ling G, Attali Y, Finn B, et al. Is a computerized adaptive test more motivating than a fixed-item test? *Appl Psychol Meas* 2017;41:495–511.
- Cella D, Gershon R, Lai J-S, et al. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16:133–41.
- Gorin JS, Dodd BG, Fitzpatrick SJ. Computerized adaptive testing with the partial credit model. *Appl Psychol Meas* 2005;29:433–56.
- Forkmann T, Boecker M, Norra C, et al. Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. *Rehabil Psychol* 2009;54:186–97.
- Linacre JM. Computer-adaptive testing: a methodology whose time has come. By John Michael Linacre, Ph. D. MESA Psychometric Laboratory University of Chicago. *Test [online]* 2000;69 [https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000\\_CAT.pdf](https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000_CAT.pdf)
- Bjorner JB, Chang C-H, Thissen D, et al. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res* 2007;16:95–108.
- Vodicka E, Kim K, Devine EB, et al. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007–2013). *Contemp Clin Trials* 2015;43
- Rivera SC, Kyte DG, Aiyegbusi OL, et al. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. *Health Qual Life Outcomes* 2019;17:1–19.
- Dawson J, Doll H, Fitzpatrick R, et al. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010;340:c186. doi:10.1136/bmj.c186
- Howell D, Molloy S, Wilkinson K, et al. Patient-Reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Ann Oncol* 2015;26:1846–58.
- Turner GM, Litchfield I, Finnikin S, et al. General practitioners' views on use of patient reported outcome measures in primary care: a cross-sectional survey and qualitative study. *BMC Fam Pract* 2020;21:1–10.
- Porter I, Gonçalves-Bradley D, Ricci-Cabello I, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. *J Comp Eff Res* 2016;5:507–19.
- Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ* 2001;322:406–9. doi:10.1136/bmj.322.7283.406
- Härter M, Reuter K, Gross-Hardt K, et al. Screening for anxiety, depressive and somatoform disorders in rehabilitation—validity of HADS and GHQ-12 in patients with musculoskeletal disease. *Disabil Rehabil* 2001;23:737–44.
- Härter M, Woll S, Reuter K, et al. Recognition of psychiatric disorders in musculoskeletal and cardiovascular rehabilitation patients. *Arch Phys Med Rehabil* 2004;85:106 doi:10.1016/j.apmr.2003.08.106
- Härter M, Reuter K, Weisser B, et al. A descriptive study of psychiatric disorders and psychosocial burden in rehabilitation patients with musculoskeletal diseases. *Arch Phys Med Rehabil* 2002;83:461–8.
- Tirosh O, Tran P, Renouf J, et al. PROMsBase: web-based repository portal for patient-reported outcome measures in orthopaedics. *Health Informatics J* 2019;25:867–77.
- Santana M-J, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Qual Life Res* 2014;23:1505–13.
- Muñoz RF, Cuijpers P, Smit F, et al. Prevention of major depression. *Annu Rev Clin Psychol* 2010;6:181–212.
- Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 2017;318:7156 doi:10.1001/jama.2017.7156
- Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a

- randomized controlled trial. *J Clin Oncol* 2016;34:830:557–65. doi:10.1200/JCO.2015.63.0830
- 43 Siu AL, Bibbins-Domingo K, Grossman DC. Screening for depression in adults. *JAMA* 2016;315:380.
- 44 Mitchell AJ, Vahabzadeh A, Magruder K. Screening for distress and depression in cancer settings: 10 lessons from 40 years of primary-care research. *Psychooncology* 2011;20 doi:10.1002/pon.1943
- 45 Haverman L, van Oers HA, van Muilekom MM, et al. Options for the interpretation of and recommendations for acting on different PROMs in daily clinical practice using KLIK. *Med Care* 2019;57:S52–8.
- 46 Farin E, Glattacker M, Jäckel WH. Leitlinien und Leitlinienforschung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2011;54:429–35. doi:10.1007/s00103-011-1238-9
- 47 Marcus A. *Mobile persuasion design: changing behaviour by combining persuasion design with information design. Human-Computer interaction*. London: Springer-Verlag, 2015.
- 48 Fogg B. A behavior model for persuasive design. Proc 4th Int Conf Persuas Technol - Persuas '09 [online], 2009. Available: <http://portal.acm.org/citation.cfm?doid=1541948.1541999>
- 49 Baumeister H, Kraft R, Baumel A. Persuasive e-health design for behavior change. In: Baumeister H, Montag C, eds. *Mobile sensing and digital phenotyping: new developments in psychoinformatics*. Berlin: Springer, 2019.
- 50 Fogg B. Persuasive computers: perspectives and research directions. *Proc SIGCHI Conf Hum Factors Comput Syst - CHI'98* 1998:225–32.
- 51 Imison C, Castle-Clarke S, Watson R. *Delivering the benefits of digital health care*. London, 2016.
- 52 Sucher JF, Moore FA, Todd SR, et al. Computerized clinical decision support: a technology to implement and validate evidence based guidelines. *J Trauma* 2008;64:520–37. doi:10.1097/TA.0b013e3181601812
- 53 Baumeister H, Jahed J, Vogel B. *Diagnostik, Indikation und Behandlung von psychischen Störungen in Der medizinischen rehabilitation (DIBpS): ein Leitfaden Zur Implementierung eines psychodiagnostischen Stufenplans in Der medizinischen rehabilitation*. Berlin: DRV M4 - Citavi, 2011.
- 54 Reese C, Mittag O. *Psychologische Interventionen. Praxiseempfehlungen für psychologische Interventionen in Der rehabilitation: Chronische Rückenschmerzen und Koronare Herzerkrankung*. Berlin: Deutsche Rentenversicherung M4 - Citavi, 2013.
- 55 Reese C, Spieser A, Mittag O. Psychologische Interventionen in Der rehabilitation von Patienten MIT koronarer Herzerkrankung: Zusammenfassung Der Evidenz und Der Empfehlungen AUS systematischen Übersichtsarbeiten und Leitlinien. *Rehabilitation* 2012;51:405–14.
- 56 DGPPN, BÄK, KBV, AWMF, AkdÄ, BpTK DAGSHG, DEGAM, DGPM, DGPS, DGRW Ba. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare depression. *Langfassung* 2015.
- 57 Bandelow B, Wiltink J, Alpers GW. Deutsche S3-Leitlinie Behandlung von Angststörungen 2014.
- 58 Ben-Zeev D, Young MA. Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: an experience sampling study. *J Nerv Ment Dis* 2010;198:280–5.
- 59 Gorzelitz J, Peppard PE, Malecki K, et al. Predictors of discordance in self-report versus device-measured physical activity measurement. *Ann Epidemiol* 2018;28:427–31.
- 60 Paulhus DL. Socially desirable responding on self-reports. In: *Encyclopedia of personality and individual differences*. Springer International Publishing, 2017: 1–5.
- 61 Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. *Ann Behav Med* 2002;24:236–43.
- 62 Baumeister H, Montag C. Digital phenotyping and mobile sensing [Internet]. In: *Studies in neuroscience, psychology and behavioral economics*. Cham: Springer International Publishing, 2019. <http://link.springer.com/10.1007/978-3-030-31620-4>
- 63 Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016;41:1691–6. doi:10.1038/npp.2016.7
- 64 Low CA, Dey AK, Ferreira D, et al. Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. *J Med Internet Res* 2017;19:9046. doi:10.2196/jmir.9046
- 65 Messner E-M, Sariyska R, Mayer B. Insights: future implications of passive smartphone sensing in the therapeutic context. *Verhaltenstherapie [online]* 2019 <https://www.karger.com/Article/FullText/501735>
- 66 Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu Rev Clin Psychol* 2017;13:23–47. doi:10.1146/annurev-clinpsy-032816-044949
- 67 Montag C, Baumeister H, Kannen C, et al. Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology. *J* 2019;2:102–15.
- 68 Moshe I, Terhorst Y, Opoku Asare K, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Front Psychiatry* 2021;12:625247.
- 69 Rohani DA, Faurholt-Jepsen M, Kessing LV, et al. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR Mhealth Uhealth* 2018;6:9691 doi:10.2196/mhealth.9691
- 70 Umbricht D, Cheng W-Y, Lipsmeier F, et al. Deep learning-based human activity recognition for continuous activity and gesture monitoring for schizophrenia patients with negative symptoms. *Front Psychiatry* 2020;11:574375.
- 71 Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genomics* 2015;8:33.
- 72 Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med Overseas Ed* 2010;363:304:301–4. doi:10.1056/NEJMp1006304
- 73 Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216–9.
- 74 Wang F, Casalino LP, Khullar D. Deep learning in medicine-promise, progress, and challenges. *JAMA Intern Med* 2019;179:293.
- 75 Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661.
- 76 Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
- 77 Montgomery P, Grant S, Mayo-Wilson E, et al. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. *Trials* 2018;19:407. doi:10.1186/s13063-018-2733-1
- 78 Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (cheers) statement. *Value in Health* 2013;16:e1–5.
- 79 Ramsey SD, Willke RJ, Glick H. Cost-effectiveness analysis alongside clinical trials II—an ISPOR good research practices Task force report. *Value Heal [online]* 2015;18.
- 80 Chan A-W, Tetzlaff JM, Altman DG, et al. Spirit 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:583. doi:10.7326/0003-4819-158-3-201302050-00583
- 81 Rauch A, Cieza A, Stucki G. How to apply the International classification of functioning, disability and health (ICF) for rehabilitation management in clinical practice. *Eur J Phys Rehabil Med* 2008;44:329–42.
- 82 World Health Organization. Internationale Klassifikation Der Funktionsfähigkeit, Behinderung und Gesundheit (ICF). *Handb Der Neuro-und Biopsychologie* 2005:615–25.
- 83 van Muilekom MM, Luijten MAJ, van Oers HA, et al. From statistics to clinics: the visual feedback of PROMIS® cats. *J Patient Rep Outcomes* 2021;5:55. doi:10.1186/s41687-021-00324-y
- 84 Deutsche Rentenversicherung Bund. Reha-Therapiestandards Depressive Störungen - für die medizinische Rehabilitation der Rentenversicherung. *Geschäftsbereich Sozialmedizin und Rehabil* 2016.
- 85 Deutsche Rentenversicherung Bund. Reha-Therapiestandards Chronischer Rückenschmerz - für die medizinische Rehabilitation der Rentenversicherung. *Geschäftsbereich Sozialmedizin und Rehabil* 2016.
- 86 Deutsche Rentenversicherung Bund. Reha-Therapiestandards Koronare Herzkrankheit - für die medizinische Rehabilitation der Rentenversicherung. *Geschäftsbereich Sozialmedizin und Rehabil* 2016.
- 87 Deutsche Rentenversicherung Bund. Rahmenkonzept zur Nachsorge - für medizinische Rehabilitation nach §15 SGB VI. *Geschäftsbereich Sozialmedizin und Rehabil* 2015.
- 88 Cuijpers P, Turner EH, Koole SL, et al. What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depress Anxiety* 2014;31:374–8.
- 89 Bell ML, McKenzie JE. Designing psycho-oncology randomised trials and cluster randomised trials: variance components and intra-cluster correlation of commonly used psychosocial measures. *Psychooncology* 2013;22:1738–47.

- 90 Adams G, Gulliford MC, Ukoumunne OC, *et al.* Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785–94.
- 91 Baumeister H, Kallinger S, Scharm H. Sachbericht zum Zwischenverwendungsnachweis 2017 - Implementierung einer Computer-adaptiven Erst- und Verlaufsdiagnostik zur Erfassung der funktionalen Gesundheit in der orthopädischen und kardiologischen Rehabilitation (DRV-Bund gefördert) 2018.
- 92 Pilkonis PA, Choi SW, Reise SP, *et al.* Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS®): depression, anxiety, and anger. *Assessment* 2011;18:263–83.
- 93 Nolte S, Coon C, Hudgens S, *et al.* Psychometric evaluation of the PROMIS® depression item bank: an illustration of classical test theory methods. *J Patient Rep Outcomes* 2019;3:46.
- 94 Terwee CB, Crins MHP, Boers M, *et al.* Validation of two PROMIS item banks for measuring social participation in the Dutch general population. *Qual Life Res* 2019;28:211–20.
- 95 Crins MHP, Terwee CB, Ogreden O, *et al.* Differential item functioning of the PROMIS physical function, pain interference, and pain behavior item banks across patients with different musculoskeletal disorders and persons from the general population. *Qual Life Res* 2019;28:1231–43.
- 96 Cella D, Lai J-S, Jensen SE, *et al.* Promis fatigue item bank had clinical validity across diverse chronic conditions. *J Clin Epidemiol* 2016;73:128–34.
- 97 van Kooten JAMC, Terwee CB, Luijten MAJ, *et al.* Psychometric properties of the patient-reported outcomes measurement information system (PROMIS) sleep disturbance and sleep-related impairment item banks in adolescents. *J Sleep Res* 2021;30:1–12.
- 98 Wesselius HM, van den Ende ES, Alsmas J, *et al.* Quality and quantity of sleep and factors associated with sleep disturbance in hospitalized patients. *JAMA Intern Med* 2018;178:1201–71.
- 99 Katzan IL, Lapin B. Promis GH (patient-reported outcomes measurement information system global health) scale in stroke: a validation study. *Stroke* 2018;49:147–54.
- 100 Gruber-Baldini AL, Velozo C, Romero S, *et al.* Validation of the PROMIS® measures of self-efficacy for managing chronic conditions. *Qual Life Res* 2017;26:1915–24.
- 101 Salsman JM, Schalet BD, Merluzzi TV, *et al.* Calibration and initial validation of a general self-efficacy item bank and short form for the NIH PROMIS. *Qual Life Res* 2019;28:2513–23.
- 102 Liegl G, Rose M, Correia H, *et al.* An initial psychometric evaluation of the German PROMIS v1.2 physical function item bank in patients with a wide range of health conditions. *Clin Rehabil* 2018;32:84–93. doi:10.1177/0269215517714297
- 103 Reinert DF, Allen JP. The alcohol use disorders identification test (audit): a review of recent research. *Alcohol Clin Exp Res* 2002;26:272–9. doi:10.1111/j.1530-0277.2002.tb02534.x
- 104 Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German. *J Res Pers* 2007;41:203–12. doi:10.1016/j.jrp.2006.02.001
- 105 Herdman M, Gudex C, Lloyd A, *et al.* Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
- 106 Ludwig K, Graf von der Schulenburg J-M, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics* 2018;36:663–74. doi:10.1007/s40273-018-0615-8
- 107 Chisholm D, Knapp MRJ, Knudsen HC, *et al.* Client socio-demographic and service receipt inventory – European version: development of an instrument for international research. *Br J Psychiatry* 2000;177:s28–33. doi:10.1192/bjpp.177.39.s28
- 108 Baumeister H, Bengel J, Forkmann T. Implementierung einer Computer-adaptiven Erst- und Verlaufsdiagnostik Zur Erfassung Der funktionalen Gesundheit in Der orthopädischen und kardiologischen Rehabilitation. *Drittmittelprojekt, DRV-Bund* 2015.
- 109 Ferreira D, Kostakov V, Dey AK. Aware: mobile context instrumentation framework. *Front ICT* 2015;2. doi:10.3389/fict.2015.00006
- 110 Bae S, Chung T, Ferreira D, *et al.* Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. *Addict Behav* 2018;83:42–7. doi:10.1016/j.addbeh.2017.11.039
- 111 Philippi P, Baumeister H, Apolinário-Hagen J, *et al.* Acceptance towards digital health interventions - Model validation and further development of the Unified Theory of Acceptance and Use of Technology. *Internet Interv* 2021;26:100459. doi:10.1016/j.intvent.2021.100459
- 112 Venkatesh, Morris, Davis, *et al.* User acceptance of information technology: toward a unified view. *MIS Quarterly* 2003;27:425. doi:10.2307/30036540
- 113 Stoyanov SR, Hides L, Kavanagh DJ, *et al.* Development and validation of the user version of the mobile application rating scale (uMARS). *JMIR Mhealth Uhealth* 2016;4:e72. doi:10.2196/mhealth.5849
- 114 Kroenke K, Strine TW, Spitzer RL, *et al.* The PHQ-8 as a measure of current depression in the general population. *J Affect Disord* 2009;114:163–73.
- 115 Löwe B, Spitzer RL, Zipfel S. Komplettversion und Kurzform Autorisierte Deutsche version des. *Nervenarzt* 2002;2–11.
- 116 Löwe B, Kroenke K, Herzog W, *et al.* Measuring depression outcome with a brief self-report instrument: sensitivity to change of the patient health questionnaire (PHQ-9). *J Affect Disord* 2004;81:61–6.
- 117 Spitzer RL, Kroenke K, Williams JBW, *et al.* A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166:1092–7.
- 118 Klein EM, Brähler E, Dreier M, *et al.* The German version of the perceived stress scale – psychometric characteristics in a representative German community sample. *BMC Psychiatry* 2016;16. doi:10.1186/s12888-016-0875-9
- 119 Gerber M, Lang C, Lemola S, *et al.* Validation of the German version of the insomnia severity index in adolescents, young adults and adult workers: results from three cross-sectional studies. *BMC Psychiatry* 2016;16:174. doi:10.1186/s12888-016-0876-8
- 120 Hughes ME, Waite LJ, Hawkey LC, *et al.* A short scale for measuring loneliness in large surveys: results from two population-based studies. *Res Aging* 2004;26:655–72.
- 121 Rathner E-M, Terhorst Y, Cummins N. State of mind: classification through self-reported affect and word use in speech. In: Interspeech 2018 [online], 2018. Available: [https://www.isca-speech.org/archive/interspeech\\_2018/rathner18b\\_interspeech.html](https://www.isca-speech.org/archive/interspeech_2018/rathner18b_interspeech.html)
- 122 Fontaine JRJ, Scherer KR, Roesch EB, *et al.* The world of emotions is not two-dimensional. *Psychol Sci* 2007;18:1050–7.
- 123 Glick HA, Doshi JA, Sonnand SS. Economic evaluation in clinical trials. In: *Handbooks in health economic evaluation series*. Oxford: Oxford University Press, 2014.
- 124 Willan AR, Briggs AH. Statistical analysis of cost-effectiveness data. *Statistical analysis of cost-effectiveness data* 2006:1–196.
- 125 Marseille E, Larson B, Kazi DS, *et al.* Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bull World Health Organ* 2015;93:118–24.
- 126 Woods B, Revill P, Sculpher M, *et al.* Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value Health* 2016;19:929–35.
- 127 Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;13:843 doi:10.1002/hec.843
- 128 Kilian R, Becker T, Frasch K. Effectiveness and cost-effectiveness of home treatment compared with inpatient care for patients with acute mental disorders in a rural catchment area in Germany. *Neurol Psych Brain Res* 2016;22:81–6.
- 129 Kilian R, Frasch K, Steinert T, *et al.* Cost-Effectiveness of psychotropic polypharmacy in routine schizophrenia care. Results of the ELAN prospective observational trial. *Neurology, Psychiatry and Brain Research* 2018;30:47–55.
- 130 Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health economics and cost-effectiveness analysis. *Health Econ* 2002;11:415–30. doi:10.1002/hec.678
- 131 Salize HJ, Kilian R. *Gesundheitsökonomie in der Psychiatrie - Konzepte, Methoden und Analysen*. Stuttgart: Kohlhammer, 2010.
- 132 van Buuren S, Groothuis-Oudshoorn K. mice : multivariate imputation by chained equations in R. *J Stat Softw [online]* 2011;45.
- 133 Goldstein H. *Multi-Level statistical models*. New York: Halsted, 1995: 64–88.
- 134 MacCallum RC, Kim C, Malarkey WB, *et al.* Studying multivariate change using multilevel models and latent curve models. *Multivariate Behav Res* 1997;32:215–53. doi:10.1207/s15327906mbr3203\_1
- 135 Nezlek JB. Multilevel modeling for psychologists. *APA Handb Res methods Psychol Vol 3 Data Anal Res Publ* 2012;3:219–41.
- 136 Hastie T, Tibshirani R, Friedman J. Prototype methods and nearest-neighbors 2009.
- 137 The alternating decision tree learning algorithm. *Int Conf Mach learn* 1999.
- 138 Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. in: *lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 1995.